

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΝΟΣΗΛΕΥΤΙΚΗΣ

ΔΙΙΔΡΥΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΙΔΙΚΕΥΣΗ: ΠΛΗΡΟΦΟΡΙΚΗ ΤΗΣ ΥΓΕΙΑΣ

**ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΑΣΘΕΝΩΝ COVID-19 ΜΕ ΤΗ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

**ΚΟΥΡΜΠΑΝΗΣ ΝΙΚΟΛΑΟΣ
ΒΙΟΛΟΓΟΣ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΘΗΝΑ 2022

**ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΑΣΘΕΝΩΝ COVID-19 ΜΕ ΤΗ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΝΟΣΗΛΕΥΤΙΚΗΣ

ΔΙΙΔΡΥΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΙΔΙΚΕΥΣΗ: ΠΛΗΡΟΦΟΡΙΚΗ ΤΗΣ ΥΓΕΙΑΣ

ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΑΣΘΕΝΩΝ COVID-19 ΜΕ ΤΗ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

ΚΟΥΡΜΠΑΝΗΣ ΝΙΚΟΛΑΟΣ
ΒΙΟΛΟΓΟΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΘΗΝΑ 2022

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

ΙΩΑΝΝΗΣ ΜΑΝΤΑΣ, ΚΑΘΗΓΗΤΗΣ (ΕΠΙΒΛΕΠΩΝ)

ΙΩΣΗΦ ΛΙΑΣΚΟΣ, ΔΡ. ΜΕΛΟΣ Ε.ΔΙ.Π.

ΕΜΜΑΝΟΥΗΛ ΖΟΥΛΙΑΣ, ΔΡ. ΜΕΛΟΣ Ε.ΔΙ.Π.

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΝΟΣΗΛΕΥΤΙΚΗΣ

ΔΙΙΔΡΥΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΙΔΙΚΕΥΣΗ: ΠΛΗΡΟΦΟΡΙΚΗ ΤΗΣ ΥΓΕΙΑΣ

ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΑΣΘΕΝΩΝ COVID-19 ΜΕ ΤΗ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

ΚΟΥΡΜΠΑΝΗΣ ΝΙΚΟΛΑΟΣ
ΒΙΟΛΟΓΟΣ

ΠΕΡΙΛΗΨΗ

Στα τέλη του 2019 έκανε την εμφάνισή της η ασθένεια του νέου κορονοϊού 2019 (COVID-19), η οποία προκαλείται από τον ιό SARS-CoV-2. Ο ιός ανιχνεύτηκε για πρώτη φορά στην πόλη της Γουχάν της Κίνας το Δεκέμβριο του 2019 και έκτοτε έχει εξαπλωθεί παγκοσμίως ως η νέα πανδημία [1], η οποία συνεχίζεται μέχρι και σήμερα. Ο SARS-CoV-2 (Severe Acute Respiratory Syndrome Corona Virus 2) διασπείρεται αερογενώς μέσω αναπνευστικών σταγονιδίων ή αερολυμάτων που παράγονται από τον βήχα ή το φτέρνισμα, ενός μολυσμένου ατόμου. Ο SARS-CoV-2 ως RNA ιός εμφανίζει μεταλλάξεις λόγω του τρόπου αναπαραγωγής του με αποτέλεσμα τη δημιουργία αρκετών παραλλαγών του οι οποίες και κατατάσσονται σε παραλλαγές ανησυχίας (Variants of Concern-VOCs) και παραλλαγές ενδιαφέροντος (Variants of Interest-VOIs) από τον Παγκόσμιο Οργανισμό Υγείας (WHO). Βασικός παράγοντας που αυξάνει τον κίνδυνο επιπλοκών και τη σοβαρότητα νόσησης με COVID-19 είναι η ηλικία, καθώς οι ηλικιωμένοι είναι πιο πιθανό να αρρωστήσουν σοβαρά από COVID-19.

Τα μέτρα πρόληψης που λήφθηκαν παγκοσμίως για την ελαχιστοποίηση μετάδοσης του SAR-CoV-2 περιλαμβάνουν την κοινωνική αποστασιοποίηση, τον αερισμό εσωτερικών χώρων, την κάλυψη του προσώπου κατά το βήχα ή το φτέρνισμα, το πλύσιμο των χεριών και την χρήση μάσκας προσώπου σε εσωτερικούς χώρους και φυσικά τον εμβολιασμό.

Μέχρι και σήμερα τουλάχιστον δέκα εμβόλια έχουν λάβει έγκριση από τουλάχιστον μία εθνική ρυθμιστική αρχή για χρήση στο ευρύ κοινό.

Στην αντιμετώπιση αυτής της παγκόσμιας κρίσης έρχεται να συμβάλει η Τεχνητή Νοημοσύνη μέσω των δυνατότητων που προσφέρει στης δημιουργίας μοντέλων πρόβλεψης μέσω των αλγορίθμων της Μηχανικής Μάθησης. Οι αλγόριθμοι Μηχανικής Μάθησης επεξεργάζονται τη γνώση και την αναπαριστούν με πιο μαθηματικοποιημένους τρόπους και εφαρμόζονται με επιτυχία στην επίλυση πληθώρας προβλημάτων, για πολλούς επιστημονικούς τομείς, όπως η Εξόρυξη Δεδομένων, οι Πιθανότητες και η Στατιστική, η Νευροβιολογία, κ.α.

Στόχος της παρούσας εργασίας, είναι η σύγκριση διαφορετικών αλγορίθμικών μοντέλων, με σκοπό την εύρεση της καλύτερου τρόπου μοντέλου πρόβλεψης της θνησιμότητας των ασθενών με COVID-19, μέσα από **6** αλγορίθμους κατηγοριοποίησης, χρησιμοποιώντας δεδομένα τα κλινικά χαρακτηριστικά και το ιστορικό των ασθενών. Πιο συγκεκριμένα το σύνολο των δεδομένων αποτελείται από **12.425.179** άτομα ύποπτα για COVID-19, τα οποία προσήλθαν σε υγειονομικές μονάδες του Μεξικού, με τα **3.993.464** από αυτά να είναι θετικά στον SARS-CoV-2. Οι **6** αλγόριθμοι που χρησιμοποιήθηκαν, είναι η Λογιστική Παλινδρόμηση (LR), τα Δέντρα Απόφασης (DTs), το Τυχαίο Δάσος (RF), η Ακραία Ενίσχυση Κλίσης (XGB), τα TNΔ Πολυστρωματικών Αντιλήπτρων (MLPs) και οι Κ Εγγύτερων Γειτόνων (KNN).

Μετά από την ολοκλήρωση της Προεπεξεργασίας των δεδομένων, η οποία περιελάμβανε τον καθαρισμό των δεδομένων από δείγματα με ελλείπεις τιμές, την απαλοιφή κυρίως γεωγραφικών χαρακτηριστικών (columns), τα οποία δε σχετίζόντουσαν με τη θνησιμότητα από COVID-19 και τέλος την μετασχηματισμό των συνεχών δεδομένων με **6** διαφορετικούς τρόπους (No Scaling, Standard Scaling, Min-Max Scaling με εύρη 0-1, 0-10, 0-100, 0-1000), τα δεδομένα τροφοδοτήθηκαν στα διαφορετικά μοντέλα των αλγορίθμων. Για κάθε αλγόριθμο προέκυψαν **54** μοντέλα (**6** τρόποι προεπεξεργασίας **χ 3** σετ χαρακτηριστικών **χ 3** σετ υπερπαραμέτρων) με τον κάθε μοντέλο να εκτελείται **10** φόρες (iterations), με διαφορετικά υποσύνολα του συνόλου των δεδομένων, ώστε να ληφθεί η μέση τιμή (mean) των μετρικών του, φτάνοντας έτσι τις **540** εκτελέσεις για τον κάθε αλγόριθμο, με τελικό σύνολο τις **3.240** εκτελέσεις και για τους **6** αλγορίθμους.

Έπειτα προχωρήσαμε στην αξιολόγηση τους με βάση **5** μετρικά την Ακρίβεια (Precision), την Ευαισθησία-Ανάκληση (Recall), την Βαθμολογία F1 (F1 Score), την Περιοχή κάτω από την καμπύλη ROC (Area Under Curve ROC, AUC_ROC) και την Διάρκεια (Runtime). Η κατάταξη των αλγορίθμων, με βάση την απόδοση των μοντέλων τους, έφερε **1^α** τα μοντέλα της **Ακραίας Ενίσχυσης Κλίσης** (XGB), **2^α** αυτά του **Τυχαίο Δάσος** (RF), **3^α** εκείνα των **Τεχνητών Νευρωνικών Δικτύων Πολυστρωματικών Αντιλήπτρων** (MLPs), **4^α** αυτά των **Δέντρα Αποφάσεως** (DTs), **5^α** εκείνα των **Κ Εγγύτερων Γειτόνων** (KNN) και **6^α** αυτά της **Λογιστικής Παλινδρόμησης** (LR). Το βέλτιστο μοντέλο ήταν αυτό της **XGB** τα οποία χρησιμοποίησαν και τα 22 χαρακτηριστικά (columns), με τον Min-Max scaler με εύρος 0-100 και με τις optimal_01 τιμές υπερπαραμέτρων, με μέσο όρο (mean) για τα μετρικά της Precision 0,93764 (93,76%), της Recall 0,95472 (95,47%), του F1-score 0,9113, της AUC_ROC 0,97855 και της Runtime τα 6,67306 sec.

Λέξεις Κλειδιά: COVID-19, SARS-CoV-2, RNA ιοί, Κορονοϊοί, πρωτεΐνη Spike, υποδοχέας ACE2, Π.Ο.Υ. (WHO), Γουχάν, Παραλλαγές Ανησυχίας (Variants of Concern - VOCs), Παραλλαγές Ενδιαφέροντος (Variants of Interest-VOIs), Ανάλυση Αλυσιδωτής Αντίδρασης Πολυμεράσης μετά από Αντίστροφη Μεταγραφή (RT-PCR), Εμβόλια mRNA, Εμβόλια ιικών φορέων, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Υπολογιστική Νοημοσύνη (Computational Intelligence, CI), Ασαφής Λογική (Fuzzy Logic), Εξελικτικοί Αλγόριθμοι (Evolutionary Algorithms, EA), Λογιστική Παλινδρόμηση (Logistic Regression, LR), Δέντρα Απόφασης (Decision Trees, DT), DecisionTreeClassifier, Τυχαίο Δάσος (Random Forest, RF), RandomForestClassifier, Ακραία Ενίσχυση Κλίσης (eXtreme Gradient Boosting, XGB),

XGBClassifier, Τεχνητά Νευρωνικά Δίκτυα (TNΔ) Πολυστρωματικών Αντιλήπτρων (Multi-Layer Perceptrons, MLPs), MLPClassifier, K Εγγύτεροι Γείτονες (K-Nearest Neighbors, KNN), KNeighborsClassifier, Python, pandas, sklearn, GridSearchCV, Διασταυρούμενη Επικύρωσης 10 Τμημάτων (10-fold cross validation), Σημαντικότητα Χαρακτηριστικών (Feature Importance), Μετρικές Αξιολόγησης Αλγορίθμων, Πίνακας Σύγχυσης (Confusion Matrix), Ακρίβεια (Precision), Ευαισθησία-Ανάκληση (Recall), Βαθμολογία F1 (F1 Score), Διάρκεια (Runtime), Περιοχή κάτω από την καμπύλη ROC (Area Under Curve ROC, AUC_ROC)

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
FACULTY OF NURSING

**INTERUNIVERSITY POSTGRADUATE PROGRAM
IN HEALTH CARE MANAGEMENT AND HEALTH CARE INFORMATICS**

DISSERTATION

**CLASSIFICATION OF COVID-19 PATIENTS WITH THE USE OF MACHINE LEARNING
ALGORITHMS**

BY NIKOLAOS KOURMPANIS

SUMMARY

In late 2019, the novel coronavirus disease 2019 (COVID-19), which is caused by the SARS-CoV-2 virus, emerged. The virus was first detected in the city of Wuhan, China in December 2019 and has since spread worldwide as the new pandemic [1], which continues to this day. SARS-CoV-2 (Severe Acute Respiratory Syndrome Corona Virus 2) is spread airborne through respiratory droplets or aerosols produced by the cough or sneeze of an infected person. SARS-CoV-2, as an RNA virus, exhibits mutations due to its mode of reproduction, resulting in the creation of several variants which are classified as Variants of Concern (VOCs) and Variants of Interest (VOIs) by the World Health Organization (WHO). A key factor that increases the risk of complications and severity of illness with COVID-19 is age, as older people are more likely to get seriously ill from COVID-19.

Preventive measures taken worldwide to minimize transmission of SAR-CoV-2 included social distancing, indoor ventilation, covering the face when coughing or sneezing, washing hands, and wearing a face mask indoors and of course the vaccination. To date at least ten vaccines have been approved by at least one national regulatory authority for use in the general public.

In dealing with this global crisis, Artificial Intelligence comes to contribute through the possibilities it offers to create predictive models through Machine Learning algorithms. Machine Learning algorithms process knowledge and represent it in more mathematical ways and are successfully applied to solve a multitude of problems, in many scientific fields, such as Data Mining, Probability and Statistics, Neurobiology, etc.

The aim of this work is to compare different algorithmic models, with the aim of finding the best way to predict the mortality of patients with COVID-19, through **6** categorization algorithms, using data on the clinical characteristics and history of the patients. More

specifically, the data set consists of **12,425,179** people suspected of having COVID-19, who attended health facilities in Mexico, with **3,993,464** of them being positive for SARS-CoV-2. The **6** algorithms used are Logistic Regression (LR), Decision Trees (DTs), Random Forest (RF), Extreme Gradient Boosting (XGB), Multilayer Perceptrons TNDs (MLPs) and K Nearest Neighbors (KNN).

After the completion of the Preprocessing of the data, which included the cleaning of the data from samples with missing values, the elimination of mainly geographic features (columns), which were not related to mortality from COVID-19 and finally the transformation of the continuous data with **6** different ways (No Scaling, Standard Scaling, Min-Max Scaling for ranges 0-1, 0-10, 0-100, 0-1000), the data was fed to the different models of the algorithms. For each algorithm, **54** models were generated (**6** preprocessing modes x **3** feature sets x **3** hyperparameter sets) with each model running **10** iterations, with different subsets of the data set, to obtain the mean of the metrics of, thus reaching **540** executions for each algorithm, with a final total of **3,240** executions for all 6 algorithms.

Then we proceeded to evaluate them based on **5** metrics: Precision, Sensitivity-Recall, F1 Score, Area Under the ROC curve (AUC_ROC) and Runtime. The ranking of the algorithms, based on the performance of their models, brought **1st** the Extreme Gradient Boost (**XGB**) models, **2nd** the models of the Random Forest (**RF**), **3rd** the models of Multilayer Perceptrons (**MLPs**), **4th** the models of Trees Decision (**DTs**), **5th** the models of K Nearest Neighbors (**KNN**) and **6th** the models of Logistic Regression (**LR**). The optimal model was that of **XGB** which used all **22** features (columns), with the **Min-Max** scaler with a range of **0-100** and with the **optimal_01** set of hyperparameter values, with a mean for the Precision metrics of **0.93764** (93.76%), Recall **0.95472** (95.47%), F1-score **0.9113**, AUC_ROC **0.97855** and Runtime **6.67306** sec.

Keywords: COVID-19, SARS-CoV-2, RNA viruses, Coronaviruses, Spike protein, ACE2 receptor, W.H.O., Wuhan, Variants of Concern (VOCs), Variants of Interest (VOIs), RT-PCR, mRNA Vaccines, Viral Vector Vaccines, Artificial Intelligence, Machine Learning, Computational Intelligence (CI), Fuzzy Logic, Evolutionary Algorithms (EA), Logistic Regression (LR), Decision Trees (DT), DecisionTreeClassifier, Random Forest (RF), RandomForestClassifier, eXtreme Gradiend Boosting (XGB), XGBClassifier, Multi-Layer Perceptrons (MLPs), MLPClassifier, K-Nearest Neighbors, KNeighborsClassifier, Python, pandas, sklearn, GridSearchCV, 10-fold cross validation, Feature Importance, Algorithm Evaluation Metrics, Confusion Matrix, Precision, Recall, F1 Score, Runtime, Area Under the ROC Curve (AUC_ROC)

ΕΥΧΑΡΙΣΤΙΕΣ

Η διπλωματική εργασία πραγματοποιήθηκε στο πλαίσιο του Διεύρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Πληροφορική της Υγείας» του Τμήματος Νοσηλευτικής, Σχολής Επιστημών Υγείας, του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών.

Θα ήθελα να ευχαριστήσω

Αρχικά τον Καθηγητή Δρ. Ιωάννη Μαντά για τη συνεισφορά του για την πραγματοποίηση και την ολοκλήρωση της πτυχιακής, για την αμέριστη βοήθεια και τις πολύτιμες συμβουλές που μου έδωσε κατά τη διάρκεια της συγγραφής και τις γνώσεις που μου προσέφερε ώστε να φέρω εις πέρας τη διπλωματική μου εργασία.

Επιπλέον θα ήθελα να ευχαριστήσω τον Δρ. Ιωσήφ Λιάσκο και Δρ. Εμμανουήλ Ζούλια για την καθοδήγηση, τις εύστοχες παρατηρήσεις και σχόλια τους.

Σε αυτό το σημείο θέλω να ευχαριστήσω όλους όσους συνέβαλαν στην αποπεράτωση της πτυχιακής εργασίας.

Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου και τον αδερφό μου για την οικονομική και ηθική υποστήριξη που έδειξαν κατά τη διάρκεια των σπουδών μου.

Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου και τον αδερφό μου για την οικονομική και ηθική υποστήριξη που έδειξαν κατά τη διάρκεια των σπουδών μου.

Η συγκεκριμένη εργασία αποτελεί διπλωματική εργασία του Διεύρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Πληροφορική της Υγείας» του Τμήματος Νοσηλευτικής, Σχολής Επιστημών Υγείας, του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών και έχει ως αντικείμενο την κατηγοριοποίηση ασθενών COVID-19, με βάση τον κίνδυνο θανάτου από τη νόσο, με τη χρήση αλγορίθμων Μηχανικής Μάθησης.

copyright© Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό.

Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό εκπαιδευτικής ή ερευνητικής φύσης υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν.

Ερωτήματα που αφορούν τη χρήση της εργασίας με κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Περιεχόμενα

ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ	3
COVID-19	4
1.1 SARS-CoV-2	4
1.2 Προέλευση	5
1.3 Δομή και Γονιδιακή Οργάνωση	7
1.4 Φυλογενετική Ταξινόμηση του SARS-CoV-2	9
1.5 Μεταλλάξεις και Παραλλαγές του SARS-CoV-2	12
1.6 Μετάδοση και Μόλυνση	16
1.7 Συμπτώματα	17
1.8 Ανίχνευση και Διάγνωση του SARS-CoV-2	18
1.9 Παθογένεση και Ανοσοπαθολογία του ιού	20
1.10 Δείκτες Αυξημένου Κινδύνου Σοβαρής Νόσησης	23
1.11 Πρόληψη	25
1.12 Τρόποι Αντιμετώπισης-Θεραπείας της νόσου	30
1.13 Επιδημιολογικά Δεδομένα	33
Τεχνητή Νοημοσύνη	36
2.1 Ιστορική Αναδρομή	37
2.2 Έρευνα και Εξέλιξη	39
2.3 Μηχανική Μάθηση	41
2.4 Ταξινόμηση (Classification)	44
2.5 Παλινδρόμηση (Regression)	45
2.6 Συσταδοποίηση (Clustering)	46
2.7 Δένδρα Απόφασης (Decision Trees)	50
2.8 Υπολογιστική Νοημοσύνη	53
2.9 Ασαφής Λογική (Fuzzy Logic)	54
2.10 Εξελικτικοί Αλγόριθμοι (Evolutionary Algorithms)	57
2.11 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)	60
Χρήση τεχνικών της Μηχανικής Μάθησης στην καταπολέμηση της πανδημίας COVID-19	67
3.1 Διάγνωση της COVID-19	67
3.1.1 Με τη Χρήση κλινικών δεικτών	67
3.1.2 Με τη Χρήση Ακτινολογικών Δεδομένων	73
3.2 Πρόβλεψη της Σοβαρότητας της Νόσησης και της Θνησιμότητας ασθενών με COVID-19	78
3.3 Πρόβλεψη μεταλλάξεων και εντοπισμός των επιτόπων του SARS-CoV-2	85
ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ	93

Εργαλεία και Μέθοδοι	94
4.1 Σκοπός	94
4.2 Δεδομένα	94
4.3 Προεπεξεργασία	94
4.4 Μετασχηματισμός-Κωδικοποίηση	98
4.5 Μοντέλα και Αλγόριθμοι.....	100
4.5.1 Λογιστική Παλινδρόμηση	101
4.5.2 Δέντρα Απόφασης	105
4.5.3 Τυχαίο Δάσος	107
4.5.4 Ακραία Ενίσχυση Κλίσης	110
4.5.5 Τεχνητά Νευρωνικά Δίκτυα Πολυστρωματικών Αντιλήπτρων	113
4.5.6 Κ Εγγύτεροι Γείτονες	116
4.6 Μετρικές Αξιολόγησης Αλγορίθμων.....	119
Αποτελέσματα	123
5.1 Επεξήγηση Πινάκων	123
5.2 Πίνακες Αποτελεσμάτων.....	124
5.3 Σύγκριση Αποτελεσμάτων των αλγορίθμων	154
Συζήτηση.....	166
Συμπεράσματα	171
Βιβλιογραφία	173
Παραρτήματα	219
Παράρτημα Α: Πίνακας 1 - Κρούσματα και Θάνατοι ανά χώρα ή επικράτεια	219

ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

COVID-19

Η ασθένεια του νέου κορονοϊού 2019 (COVID-19), που είναι επίσης γνωστή και ως οξεία αναπνευστική νόσος 2019-nCoV, είναι η ασθένεια που προκαλείται από τον ιό SARS-CoV-2. Ο ιός ανιχνεύτηκε για πρώτη φορά στην πόλη της Γουχάν της Κίνας το Δεκέμβριο του 2019 και έκτοτε έχει εξαπλωθεί παγκοσμίως ως η νέα πανδημία [1].

Η Κίνα, είχε ονομάσει την νέα αυτή λοίμωξη ως «νέα πνευμονία κορονοϊού». Η ονομασία COVID-19, δόθηκε από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ), με βάσει τις συμφωνημένες συστάσεις ανάμεσα στον Παγκόσμιο Οργανισμό για την Υγεία των Ζώων, τον Οργανισμό Τροφίμων και Γεωργίας των Ηνωμένων Εθνών και τον ΠΟΥ, ώστε να μην γίνεται αναφορά σε γεωγραφική τοποθεσία, ζώο, άτομο ή ομάδα ατόμων αλλά να έχει σαφή σχέση με την ασθένεια [2]. Η ονομασία COVID-19, είναι ένα ακρωνύμιο των γραμμάτων, CO από τη λέξη corona (κορώνα), VI, από τη λέξη virus (ιός), D από τη λέξη disease (ασθένεια) και το 19 από το έτος 2019 όπου έγινε και η καταγραφή του πρώτου κρούσματος. Αρχικά στον ιό είχε δοθεί η ονομασία «Νέος Κορονοϊός 2019» ή «2019-nCov» [3].

Η ασθένεια COVID-19, που προκαλείται από τον κορονοϊό SARS-CoV-2, μεταδίδεται μεταξύ των ανθρώπων κατά κύριο λόγο όταν ένα μολυσμένο άτομο βρίσκεται σε στενή επαφή με ένα άλλο ή άλλους. Ο ιός διασπείρεται αερογενώς μέσω αναπνευστικών σταγονιδίων ή αερολυμάτων που παράγονται από τον βήχα ή το φτέρνισμα, ενός μολυσμένου ατόμου [4].

1.1 SARS-CoV-2

Ο νέος κορονοϊός που προκαλεί το σοβαρό οξύ αναπνευστικό σύνδρομο τύπου 2 [5], γνωστός με τη διεθνή ονομασία SARS-CoV-2 (Severe Acute Respiratory Syndrome Corona Virus 2) [6], είναι αυτός που προκαλεί την ασθένεια COVID-19 [7], υπεύθυνη για την σύγχρονη πανδημία που ξεκίνησε το 2019 και συνεχίζεται μέχρι και σήμερα. Στις 11 Φεβρουαρίου 2020, η Διεθνής Επιτροπή Ταξινόμησης των ιών υιοθέτησε την επίσημη ονομασία του ιού ως «κορονοϊός σοβαρού οξέος αναπνευστικού συνδρόμου τύπου 2», είχαν επίσης δοθεί προσωρινές ονομασίες όπως «ο νέος κορονοϊός του 2019» (2019-nCoV) και «ο ανθρώπινος κορονοϊός του 2019» (human coronavirus) [8-10].

Ο SARS-CoV-2 είναι γνωστός στο ευρύ κοινό ως «κορονοϊός» και η ασθένεια που προκαλεί «COVID-19», ενώ μερικοί ηγέτες κρατών χρησιμοποιούν τον όρο «κινέζικος ιός» σε κοινωνικά δίκτυα και συνεντεύξεις, καλλιεργώντας ξενοφοβικά αντανακλαστικά, αντίθετα με τις συστάσεις του ΠΟΥ όσον αφορά την σωστή διαδικασία ονοματοδοσίας ενός ιού [11,12].

Μετά το ξέσπασμα της λοίμωξης του SARS-CoV-2, στις 30 Ιανουαρίου 2020, ο ΠΟΥ αποφάσισε να χαρακτηρίσει την έξαρσή αυτή ως γεγονός διεθνούς ενδιαφέροντος, με τον χαρακτηρισμό της COVID-19 ως πανδημία, να ακολουθεί στις 11 Μαρτίου 2020 [13,14].

Ο SARS-CoV-2 είναι κορονοϊός με γονιδίωμα μονόκλωνο RNA θετικής πολικότητας [15]. Το Εθνικό Ινστιτούτο Υγείας (NIH) των ΗΠΑ τον θεωρεί ως τον διάδοχο του SARS-CoV-1 [10], που ήταν υπεύθυνος για την έξαρση της SARS τα έτη 2002 έως 2004.

Ο SARS-CoV-2 παρουσιάζει γενετικές ομοιότητες με τους κορονοϊούς SARS (SARS-CoV) σε ποσοστό 79,5%. Θεωρείται πως έχει προέλευση από τους ιούς που εμφανίζονται σε ζώα όπως οι νυχτερίδες, με τους οποίους εμφανίζει υψηλή συσχέτιση σε ποσοστό 96%. Ο ιός εμφανίζει μικρή γενετική ποικιλομορφία, υποδεικνύοντας ότι η μετάδοση του στους ανθρώπους πιθανότατα να είχε συμβεί στα τέλη του 2019.

Σύμφωνα με τις τελευταίες επιδημιολογικές μελέτες, κάθε μόλυνση από τον ιό έχει ως αποτέλεσμα 5,7 νέες μολύνσεις, σε συνθήκες που κανένα μέλος της κοινότητας δεν έχει ανοσία και που δεν λαμβάνονται μέτρα πρόληψης της διασποράς του [16]. Ο ιός εισέρχεται στα ανθρώπινα κύτταρα με τη βοήθεια του ενζύμου μετατροπής της αγγειοτενσίνης 2 (ACE2), πρωτεϊνικού υποδοχέα στο οποίο προσκολλάται αρχικά η ακίδα του, στη συνέχεια το ένζυμο μεμβράνης TMPRSS2 «κόβει» τη ακίδα, αποκαλύπτοντας έτσι τα πεπτίδια σύντηξης τα οποία και θα συγχωνευτούν με την κυτταρική μεμβράνη [17-19].

1.2 Προέλευση

Οι κορονοϊοί είναι γνωστό από καιρό ότι παρουσιάζουν υψηλό κίνδυνο πανδημίας. Ο SARS-CoV-2 είναι ο ένατος τεκμηριωμένος κορονοϊός που μολύνει ανθρώπους και ο έβδομος που εντοπίστηκε τα τελευταία 20 χρόνια [20,21]. Όλοι οι προηγούμενοι ανθρώπινοι κορονοϊοί έχουν ζωνοσογόνο προέλευση, όπως και η συντριπτική πλειοψηφία των ανθρώπινων ιών. Η εμφάνιση του SARS CoV-2 φέρει αρκετές υπογραφές αυτών των προηγούμενων ζωνοσογόνων συμβάντων. Εμφανίζει σαφείς ομοιότητες με τον SARS-CoV που μεταδόθηκε στους ανθρώπους της Φοσάν, στην επαρχία της Κουανγκτούνγκ, στην Κίνα τον Νοέμβριο του 2002, και ξανά στο Κουανγκτόσου, στην επαρχία Κουανγκτούνγκ το 2003 [22]. Τα δύο αυτά συμβάντα εμφάνισης του SARS-CoV σχετίζονταν με αγορές στη Γουχάν [24], που πωλούσαν ζωντανά ζώα και ειδικότερα είδη όπως η μοσχογαλή (*Civettictis civetta*) και οι σκύλοι ρακούν (*Nyctereutes procyonoides*) [23], τα οποία είναι γνωστό ότι είναι ευπαθή σε λοίμωξη SARS-CoV-2 [25]. Έμποροι ζώων που εργάζονταν το 2003, που δεν είχαν διαγνωστεί με SARS, βρέθηκαν να έχουν υψηλά επίπεδα ανοσοσφαιρίνης G (IgG) για τον SARS-CoV (13% συνολικά και >50% για εμπόρους που ειδικεύονται στα μοσχοκάρφι) [26]. Μεταγενέστερες έρευνες βρήκαν 3% πιοσσοτά θετικότητας στους κορονοϊούς που σχετίζονται με το SARS (SARSr-CoV) σε κατοίκους της επαρχίας Γιουνάν που ζουν κοντά σε σπήλαια νυχτερίδων [27], αποδεικνύοντας έτσι την υψηλή έκθεση των αγροτικών περιοχών στον ιό. Οι πλησιέστεροι γνωστοί συγγενείς τόσο του SARS-CoV όσο και του SARS-CoV-2 είναι ιοί από νυχτερίδες στη Γιουνάν, αν και στα ζώα από αυτήν την επαρχία έγινε κατά προτίμηση δειγματοληψία. Τόσο για το SARS-CoV όσο και για το SARS-CoV-2, υπάρχει ένα σημαντικό γεωγραφικό χάσμα μεταξύ της Γιουνάν και της τοποθεσίας των πρώτων κρουσμάτων σε ανθρώπους, υπογραμμίζοντας τη δυσκολία στον εντοπισμό της ακριβούς οδού εμφάνισης του ιού και τη σημαντικότητα της δειγματοληψίας πέρα από τη Γιουνάν.

Ο SARS-CoV-2 παρουσιάζει ομοιότητες με τους τέσσερις ενδημικούς ανθρώπινους κορονοϊούς, τον κορονοϊό-OC43 (HCoV-OC43), τον κορονοϊό-HKU1 (HCoV-HKU1), τον κορονοϊό-229E (HCoV-229E) και τον κορονοϊό NL63 (HCoV-NL63). Οι ιοί αυτοί έχουν ζωνοσογόνο προέλευση με τις συνθήκες εμφάνισής τους να είναι ασαφείς. Ομοίως με τον SARS-CoV-2, ο HCoV-HKU1, ο οποίος έκανε την εμφάνισή του στη Σενζέν τον χειμώνα του 2004, έχει άγνωστη ζωική προέλευση, περιέχει μια θέση διάσπασης φουρίνης στην πρωτεΐνη της ακίδας και εντοπίστηκε αρχικά σε μια περίπτωση ανθρώπινης πνευμονίας [28].

Με βάση επιδημιολογικά δεδομένα, η αγορά της Χουνάν της Γουχάν ήταν ένα πρώιμο και σημαντικό επίκεντρο μόλυνσης από τον SARS-CoV-2. Δύο από τα τρία πρώτα τεκμηριωμένα κρούσματα της νόσου του κορονοϊού 2019 (COVID-19) συνδέθηκαν άμεσα με αυτήν, όπως και το 28% όλων των περιπτώσεων που αναφέρθηκαν τον Δεκέμβριο του 2019 [29]. Συνολικά, το 55% των ατόμων που νόσησαν κατά τον Δεκέμβριο του 2019 είχαν περάσει ή διέμεναν κοντά στην Χουνάν είτε σε άλλες αγορές της Γουχάν, με αυτές τις περιπτώσεις να είναι πιο διαδεδομένες κατά το πρώτο μισό του Δεκεμβρίου [29]. Η μελέτη

των τοποθεσιών των πρώιμων κρουσμάτων δείχνει ότι τα περισσότερα συγκεντρώνονται γύρω από την αγορά Χουνάν, βόρεια του ποταμού Γιανγκτσέ αν και η αναφορά περιπτώσεων μπορεί να υπόκειται σε μεροληψία στη δειγματοληψία. Αυτές οι περιοχές ήταν επίσης οι πρώτες που εμφάνισαν υπερβολικούς θανάτους από πνευμονία τον Ιανουάριο του 2020, μια μέτρηση που είναι λιγότερο επιρρεπής σε πιθανή μεροληψία που σχετίζονται με την αναφορά περιπτώσεων. Δεν υπάρχει καμία επιδημιολογική σύνδεση με καμία άλλη τοποθεσία στη Γουχάν, συμπεριλαμβανομένου του Ινστιτούτου Ιολογίας της Γουχάν (WIV) που βρίσκεται νότια του Γιανγκτσέ και αποτελεί αντικείμενο εικασιών και θεωριών συνομιωσίας. Αν και ορισμένα πρώιμα κρούσματα δεν έχουν άμεση επιδημιολογική σύνδεση με μια μόνο αγορά [29], αυτό μπορεί να αιτιολογηθεί με βάσει τα υψηλά ποσοστά ασυμπτωματικής μετάδοσης καθώς και των μη τεκμηριωμένων δευτερογενών συμβάντων μετάδοσης, γεγονός που είχε παρατηρηθεί και σε πρώιμες περιπτώσεις SARS-CoV της Φοσάν [22].

Μετά το κλείσιμό της Χουνάν, ο SARS-CoV-2 εντοπίστηκε σε περιβαλλοντικά δείγματα της αγοράς, κυρίως στο δυτικό τμήμα που εμπορευόταν άγρια ζώα και κατοικίδια, ζωικά προϊόντα, καθώς και σε σχετικές περιοχές αποστράγγισης [29]. Αν και τα πτώματα ζώων ήταν αρνητικά για SARS-CoV-2, δεν ήταν αντιπροσωπευτικά των ειδών που πωλήθηκαν και συγκεκριμένα δεν περιλάμβαναν σκύλους ρακούν (Nyctereutes procyonoides) και άλλα ζώα που είναι γνωστό ότι είναι ευαίσθητα στον SARS-CoV-2 [24].

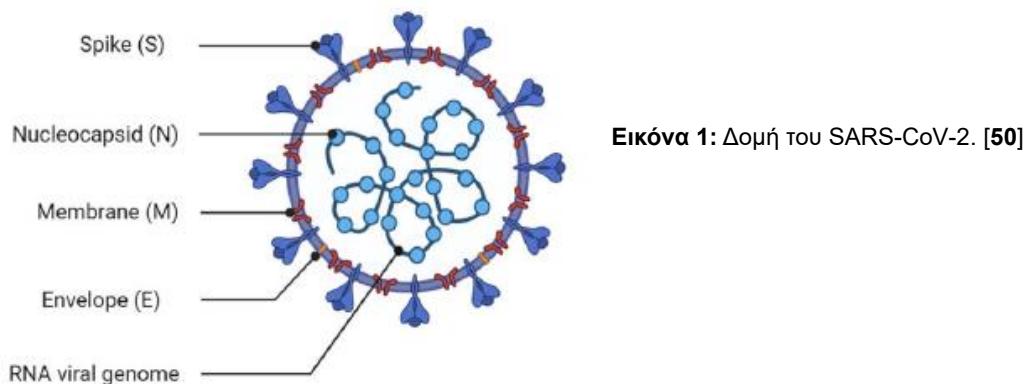
Μετά την πιο πρώιμη διάσπαση στη φυλογένεση του o SARS-CoV-2 διαχωρίστηκε σε δύο γενεαλογίες που ονομάστηκαν A και B [30] που πιθανότατα μεταδίδονταν ταυτόχρονα. Η γενεαλογία B, η οποία έγινε κυρίαρχη παγκοσμίως, παρατηρήθηκε σε πρώιμες περιπτώσεις που συνδέονταν με την Χουανάν και τα περιβαλλοντικά δείγματα που λήφθηκαν εκεί, ενώ η γενεαλογία A περιλαμβάνει τις περιπτώσεις έκθεσης σε άλλες αγορές καθώς και μεταγενέστερες περιπτώσεις της Γουχάν και άλλα μέρη της Κίνας [29]. Το φυλογενετικό μοτίβο αυτό συνάδει με την εμφάνιση του SARS-CoV-2 που περιλαμβάνει μία ή περισσότερες επαφές με μολυσμένα ζώα ή εμπόρους, συμπεριλαμβανομένων πολλαπλών συμβάντων μετάδοσης, καθώς και δυνητικά μολυσμένα ή ευαίσθητα ζώα τα οποία μεταφέρθηκαν εντός ή μεταξύ των αγορών της Γουχάν μέσω κοινών αλυσίδων προμήθειας για πώληση ή για ανθρώπινη κατανάλωση [24]. Η πιθανή πολυεστιακή εμφάνιση του SARS-CoV-2 σε πολλές αγορές, αντικατοπτρίζει και πάλι τον SARS-CoV ο οποίος εμφάνισε υψηλά επίπεδα μετάδοσης, οροεπιπολασμού και γενετικής ποικιλότητας στα ζώα τόσο στην αγορά Ντονγκμέν της Σενζέν [23,31] όσο και στην αγορά Χινιουάν της Κουανγκτσόου [32,33].

Ιοί σχετιζόμενοι με τον SARS-CoV-2 έχουν τεκμηριωθεί σε νυχτερίδες και παγκολίνους σε πολλαπλές τοποθεσίες στη Νοτιοανατολική Ασία, συμπεριλαμβανομένης της Κίνας, της Ταϊλάνδης, της Καμπότζης και της Ιαπωνίας [34,35], με ορολογικά δείγματα ιογενών λοιμώξεων σε παγκολίνους για περισσότερο από μια δεκαετία [36]. Ωστόσο, υπάρχει ένα σημαντικό εξελικτικό χάσμα μεταξύ του SARS-CoV-2 και των πλησιέστερων συγγενών ζωικών ιών, όπως για παράδειγμα, ο ίος της νυχτερίδας RaTG13 που συλλέχθηκε από το WIV έχει γενετική απόσταση ~4% (~1.150 μεταλλάξεις) από την αλληλουχία αναφοράς Wuhan-Hu-1 του SARS-CoV-2, που καταδεικνύει δεκαετίες εξελικτικής απόκλισης [37]. Αν και το RaTG13, που λήφθηκε από μια νυχτερίδα Rhinolophus affinis στο Γιουνάν [38], έχει την υψηλότερη μέση γενετική ομοιότητα με τον SARS-CoV-2, το ιστορικό ανασυνδυασμού δείχνει ότι τρεις άλλοι ιοί νυχτερίδας οι RmYN02, RpYN06 και PrC31 είναι πιο κοντά γενετικά στο γονιδίωμα του SARS-CoV-2 και επομένως πιθανόν μοιράζονται έναν πιο πρόσφατο κοινό πρόγονο με αυτόν [34,35]. Κανένας από αυτούς τους τρεις πιο κοντινούς ιούς δεν είχε συλλεχθεί από το WIV και όλοι αναλύθηκαν μετά την έναρξη της πανδημίας [35,39,40]. Συγκεντρωτικά, αυτά τα δεδομένα καταδεικνύουν χωρίς περιθώριο αμφιβολίας ότι ο RaTG13 δεν είναι ο πρόγονος του SARS-CoV-2, με ή χωρίς εργαστηριακό χειρισμό ή πειραματική μεταλλαξιγένεση.

Κανένας πληθυσμός νυχτερίδας ή ενδιάμεσος ζενιστής ζώων του SARS-CoV-2 δεν έχει εντοπιστεί μέχρι σήμερα. Αυτό πιθανόν να οφείλεται στο γεγονός πως δεν έχουν δειγματιστεί τα σωστά ζωικά είδη-πληθυσμοί ή πως οποιοσδήποτε προγονικός ιός μπορεί να έχει χαμηλό επιπολασμό. Τα αρχικά συμβάντα μετάδοσης μεταξύ ειδών είναι πολύ πιθανό να περάσουν απαρατήρητα. Οι περισσότερες μολύνσεις από SARS-CoV-2 πιθανότατα να μην οδηγήσαν σε περαιτέρω μετάδοση [41] και μόνο ένα πολύ μικρό κλάσμα διαχύσεων από ζώα σε ανθρώπους να προκαλέσει μεγάλα ξεσπάσματα. Επιπλέον η ζωική προέλευση πολλών γνωστών ανθρώπινων παθογόνων ιών, συμπεριλαμβανομένου του ιού Έμπολα, του ιού της ηπατίτιδας C, του ιού της πολιομυελίτιδας και των κορονοϊών HCoV-HKU1 και HCoV-NL63, δεν έχει ακόμη προσδιοριστεί, ενώ χρειάστηκε πάνω από μια δεκαετία για να ανακαλυφθούν ιοί νυχτερίδων με μεγαλύτερη από 95% ομοιότητα με τον SARS-CoV οι οποίοι χρησιμοποιούν το hACE-2 ως υποδοχέα [42].

1.3 Δομή και Γονιδιακή Οργάνωση

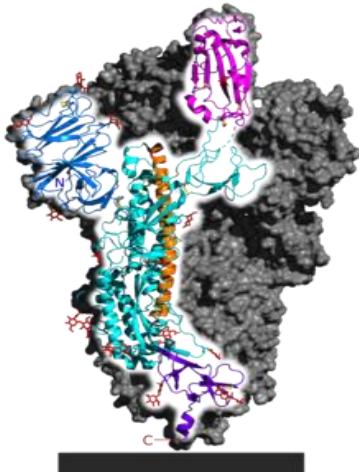
Ο SARS-CoV-2 (**Εικόνα 1**), όπως και άλλοι β-κορονοϊοί, έχει σφαιρικό σχήμα με γονιδίωμα που αποτελείται από μονόκλωνο RNA (29,0–30,2 kb) θετικής πολικότητας το οποίο περιβάλλεται από λιποπρωτεΐνικό έλυτρο[43]. Το όνομα κορονοϊός προέρχεται από την εμφάνισή που έχει ο ιός στις ηλεκτρονικές μικρογραφίες, με τις προεξοχές των γλυκοπρωτεΐνων ακίδας (**Εικόνα 2**) από τον ιικό φάκελο να μοιάζουν με κορώνα. Κάθε ισωμάτιο κυμαίνεται μεταξύ 50 και 200 nm σε διάμετρο [44]. Το γονιδίωμα του ιού εμπεριέχει τα γονίδια της ρεπλικάσης, της ακίδας (S), του φάκελου (E), της μεμβράνης (M) και του νουκλεοκαψίδιο (N) [45,46]. Σε αντίθεση με τους β-κορονοϊούς της γενεαλογίας A, ο SARS-CoV-2 στερείται του γονιδίου που κωδικοποιεί την εστεράση αιμοσυγκολλητίνης, τον ρόλο της οποίας έχει αναλάβει η γλυκοπρωτεΐνη S, διευκολύνοντας έτσι την είσοδο του ιού στο κύτταρο-ζενιστή [43]. Το γονίδιο της ρεπλικάσης περιέχει δύο μεγάλα ανοιχτά πλαίσια ανάγνωσης (ORFs), τα orf1ab και orf1a, που περικλείουν τα δύο τρίτα του γονιδίωματος, με το orf1ab ως το μεγαλύτερο γονίδιο. Τα Orf1ab και orf1a κωδικοποιούν την πολυπρωτεΐνη-1ab (pp1ab) και την πολυπρωτεΐνη-1a (pp1a), αντίστοιχα, οι οποίες διασπώνται συν-μεταφραστικά σε 16 μη δομικές πρωτεΐνες (nsps) που απαιτούνται για τη διευκόλυνση της ιικής αντιγραφής [47]. Αυτές οι διασπάσεις προκαλούνται από τις ιογενώς κωδικοποιούμενες πρωτεάσες τύπου παπαΐνης, τις πρωτεάσες τύπου χυμοθρυψίνης (3CLpro) ή από την κύρια πρωτεάση (Mpro), σε συγκεκριμένες θέσεις αναγνώρισης εντός των ORF[48,49].



Εικόνα 1: Δομή του SARS-CoV-2. [50]

Το επόμενο ORF κωδικοποιεί τη μονομερή πρωτεΐνη S (175 kDa), μια γλυκοπρωτεΐνη σύντηξης τύπου I που τριμερίζει και διευκολύνει τη σύνδεση του υποδοχέα [51-53]. Η περιοχή S1 βρίσκεται στο N-τερματικό-άκρο της πρωτεΐνης S, η οποία περιέχει τη

γλυκοζυλιωμένη N-Τερματική Περιοχή (NTD), μια Περιοχή Δέσμευσης Υποδοχέα (RBD) και ένα Μοτίβο Δέσμευσης Υποδοχέα (RBM) που αλληλεπιδρά άμεσα με τον υποδοχέα ACE2 του ξενιστή ώστε να γίνει η ίική προσκόλληση. Η περιοχή S2 δρα ως μηχανισμός σύντηξης που περιέχει δύο επαναλήψεις επτάδων (HR1 και HR2) που φέρνει κοντά τις μεμβράνες του ιού και του κυττάρου ξενιστή ώστε να διευκολύνει τη σύντηξη τους [54-57].



Εικόνα 2: Η ακίδα με σημασμένη την υπομονάδα S1, υπεύθυνη για την προσκόλληση στο ένζυμο ACE2. [58]

Η πρωτεΐνη E και η πρωτεΐνη M είναι γλυκοπρωτεΐνες 8-12 kDa και 25-30 kDa, αντίστοιχα, και βρίσκονται στον ίικό φάκελο. Τόσο η E όσο και η M είναι διαμεμβρανικές πρωτεΐνες που περιέχουν έναν C-τερματικό ενδοτομέα και έναν N-τερματικό εκτοτομέα [59]. Οι πιο σπάνιες πρωτεΐνες E έχουν δραστικότητα διαύλου ιόντων και εικάζεται ότι προάγουν την παθογένεση του ιού. Οι πρωτεΐνες M που είναι σε μεγαλύτερη αφθονία, αλληλεπιδρούν με άλλες δομικές πρωτεΐνες δίνοντας στο ιοσωμάτιο το σχήμα του [60,61].

Οι πρωτεΐνες N είναι βαριά φωσφορυλιωμένες πρωτεΐνες που υπάρχουν στο ελικοειδές νουκλεοκαψίδιο του ιού που προστατεύει το ίικό γονιδίωμα, βρίσκεται σε αφθονία σε μολυσμένα κύτταρα, διαθέτει μια C-τερματική περιοχή (CTD) και N-τερματική περιοχή (NTD) που μπορεί να συνδεθεί με μόρια RNA τόσο του ξενιστή όσο και σε άλλα ίικης προέλευσης [60]. Η πρωτεΐνη N αλληλεπιδρά με τις ρυθμιστικές θέσεις μεταγραφής (TSGs), που υπάρχουν στην αρχή κάθε δομικού και βοηθητικού γονιδίου για μεταγραφικό έλεγχο. Η C-τερματική περιοχή της περιέχει το σήμα γονιδιωματικής συσκευασίας(πακετάρισμα) που απαιτείται για τη συσκευασία του ίικού γονιδιώματος στο σύμπλεγμα ριβονουκλεοπρωτεϊνών, το καψίδιο [62-65]. Η πρωτεΐνη N στη συνέχεια αλληλεπιδρά με την πρωτεΐνη M και το nsr3, συστατικό του συμπλόκου ρεπλικάσης, ώστε να παράγει ιικά σωματίδια δίευκολύνοντας τη σύνδεση του ενθυλακωμένου ίικού γονιδιώματος στο σύμπλεγμα ρεπλικάσης-μεταγραφάσης [66,67].

1.4 Φυλογενετική Ταξινόμηση του SARS-CoV-2

Οι κορονοϊοί είναι σωματίδια που περιβάλλονται από ένα συμμετρικό εικοσαεδρικό πρωτεΐνικό περίβλημα, διαμέτρου περίπου 80-220 nm, τον φάκελο (Ε), περιέχουν ένα ενιαίο, μονόκλωνο, θετικής πολικότητας, γονιδίωμα RNA μεγέθους περίπου 26-32 kb [68]. Οι κορονοϊοί (CoVs) είναι μία από τις μεγαλύτερες ομάδες ιών που ανήκουν στην τάξη των Nidovirales (nido σημαίνει φωλιά και αναφέρεται στην ικανότητα των ιών αυτής της τάξης να δημιουργούν ένα εμφωλευμένο σύνολο υπογονιδιωματικού mRNA [69,70]), στην υποτάξη Cornidovirineae και στην οικογένεια Coronaviridae. Η οικογένεια Coronaviridae χωρίζεται σε δύο υποοικογένειες, την Letovirinae και την Orthocoronavirinae. Η υποοικογένεια Letovirinae περιλαμβάνει το γένος Alphaletoavirus, ενώ η υποοικογένεια Orthocoronaviridae χωρίζεται, με βάση τη φυλογενετική ανάλυση και τη δομή του γονιδιώματος, σε τέσσερα γένη, το γένος Alphacoronavirus (αCoV) που περιλαμβάνει 17 είδη, το γένος Betacoronavirus (βCoV) που περιλαμβάνει 14 είδη (Εικόνα 3), το γένος Gammacoronavirus (γCoV) που περιλαμβάνει 2 είδη και το γένος Deltacoronavirus (δCoV) που περιλαμβάνει 7 είδη.

Genus	Subgenus	Species	Abbreviations	Accession Numbers	Reservoir Host	Zoonotic
Betacoronavirus	Embecovirus	<i>Betacoronavirus 1</i>	HCoV-OC43	NC_006213.1	Human	No
			BCoV	NC_003045.1	Bovines	No
			ECoV	EF_446615.1	Equines	No
	<i>Embecovirus</i>	<i>China Rattus coronavirus HKU24</i>	RtCoV-HKU24	NC_026011.1	Rats	No
		<i>Human coronavirus HKU1</i>	HCoV-HKU1	NC_006577.2	Human	No
		<i>Murine coronavirus</i>	MHV	NC_001846.1	Mouse	No
		<i>Rabbit coronavirus HKU14</i>	RbCoV HKU14	JN_874559	Rabbits	No
	<i>Hibecovirus</i>	<i>Bat Hp-betacoronavirus Zhejiang2013</i>	BtHpCoV-ZJ13	NC_025217.1	Bats	No
		<i>Hedgehog coronavirus 1</i>	EriCoV-1	NC_039207.1	Hedgehog	No
	<i>Merbecovirus</i>	<i>Middle East respiratory syndrome-related coronavirus</i>	MERSr-CoV	NC_019843.3	Human, camels, and bats	Yes
		<i>Pipistrellus bat coronavirus HKU5</i>	BtPiCoV-HKU5	NC_009020.1	Bats	No
		<i>Tylonycteris bat coronavirus HKU4</i>	BtTyCoV-HKU4	NC_009019.1	Bats	No
<i>Nobecovirus</i>	<i>Nobecovirus</i>	<i>Roussetus bat coronavirus GCCDC1</i>	BtEoCoV-GCCDC1	NC_030886.1	Bats	No
		<i>Roussetus bat coronavirus HKU9</i>	BtRoCoV-HKU9	MG762674.1	Bats	No
	<i>Sarbecovirus</i>	<i>Severe acute respiratory syndrome-related coronavirus</i>	SARSr-CoV	NC_004718.3	Human, palm civets, and bats	Yes
<i>Unclassified Betacoronavirus</i>	<i>Betacoronavirus</i>	<i>Pangolin coronavirus</i>	Pangolin-CoV	NA_606875.1	Pangolins	No

Εικόνα 3: Η ταξινόμηση του γένους Betacoronavirus που ανήκει στο SARS-CoV-2. [71]

Οι κορονοϊοί μολύνουν ένα ευρύ φάσμα άγριων και κατοικίδιων ζώων. Οι αCoV και βCoV μολύνουν τα θηλαστικά, ενώ οι γCoV και δCoV μολύνουν κυρίως τα πουλιά. Ο πρώτος κορονοϊός που μόλυνε ανθρώπους (HCoV) απομονώθηκε για πρώτη φορά το 1960 από νοσηλευόμενους ασθενείς που υπέφεραν από συμπτώματα κοινού κρυολογήματος και ονομάστηκε B814 [72]. Μέχρι στιγμής, οι εππάντα διαφορετικοί HCoV που μολύνουν τους ανθρώπους είναι οι 229E, NL63, που ανήκουν στους αCoVs και οι HKU1, OC43, SARS, MERS και SARS-CoV-2, που ανήκουν στους βCoVs. Η πανδημία του 2002–2003 που ξεκίνησε από την Κίνα προκλήθηκε από τον SARS-CoV, που ανήκει στη γενεαλογία B του γένους βCoV [73]. Στη Μέση Ανατολή, ο MERS-CoV που ανήκει στη γενεαλογία C του γένους βCoV εμφανίστηκε το 2012 [74]. Το 2019 ο SARS-CoV-2, στενά συνδεδεμένος με

κορονοϊούς που σχετίζονται με το SARS των νυχτερίδων, ταξινομήθηκε στη γενεαλογία B του γένους των βCoV. Ο Chan και οι συνεργάτες του [75] έδειξε ότι ο SARS-CoV-2 αντιπροσωπεύει μια ξεχωριστή γενεαλογία στο υπογένος Sarbecovirus [76].

Επιπλέον υπάρχουν και άλλοι κορονοϊοί που έχουν προκαλέσει πανδημικές ασθένειες σε οικόσιτα και άγρια θηλαστικά και πτηνά, οδηγώντας σε υψηλά ποσοστά θνησιμότητας και οικονομικές απώλειες. Αυτοί οι ιοί περιλαμβάνουν τον IBV των κοτόπουλων [77], τον κορονοϊό της φάλαινας Beluga SW1 (BWCov-SW1) [78], τους κορονοϊούς της νυχτερίδας CDPHE15 και HKU10, τον ιό της επιδημικής διάρροιας των χοίρων (PEDV), τον TGEV και τον SADS-CoV (ιός του ξαφνικού οξείου διαρροϊκού συνδρόμου) [79].

Συγκριτική Φυλογενετική Ανάλυση

Οι μονόκλωνοι ιοί RNA εμφανίζουν ταχύτερο ρυθμό βιολογικής μετάλλαξης λόγω της έλλειψης δραστικότητας διόρθωσης των ιικών πολυμερασών RNA [80]. Ωστόσο, σε αντίθεση με άλλους επιρρεπείς σε μεταλλάξεις ιούς RNA, οι κορονοϊοί έχουν υποτοπώδεις δυνατότητες διόρθωσης, με την πρωτεΐνη nsp14 να είναι υπεύθυνη για το ενισχυμένο μέγεθος του γονιδιώματος των μελών της οικογένειας CoV [81].

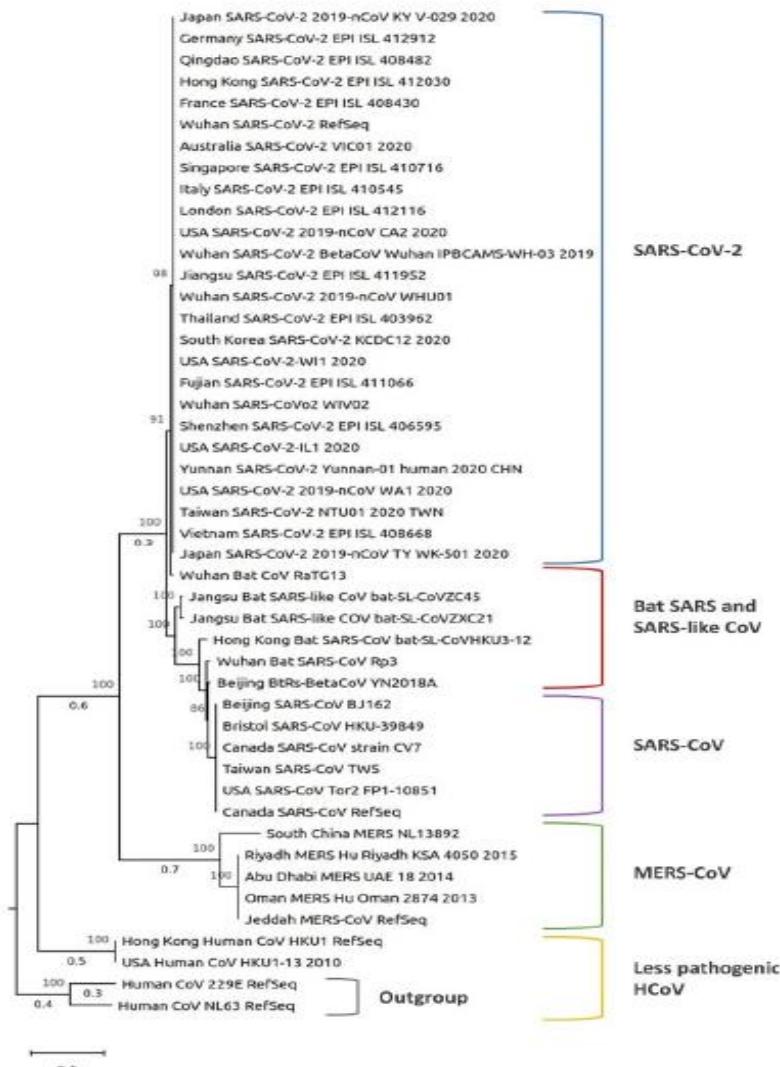
Ο ανασυνδυασμός είναι ένας ακόμα μηχανισμός εξέλιξης των κορονοϊών [82]. Αναφορές εμφάνισης ιικού ανασυνδυασμού υπάρχουν για τους MERS-CoV και SARS-CoV. Επτά πιθανές περιοχές ανασυνδυασμού ανιχνεύθηκαν στην πρωτεΐνη ORF1ab και S μεταξύ του SARS-CoV και έξι άλλων κορονοϊών με *in silico* ανάλυση των γονιδιωμάτων τους [83]. Ο ανασυνδυασμός στον SARS-CoV-2 δεν είναι ακόμη πλήρως κατανοητός. Οι αρχικές μελέτες πρότειναν ότι μπορεί να συνέβη κατά τη διάρκεια της εξέλιξης του [84], ενώ άλλοι ερευνητές απέκλεισαν την πιθανότητα ανασυνδυασμού με βάση μιας εξελικτικής ανάλυσης γονιδιώματος ελέγχου πιθανών συμβάντων ανασυνδυασμού [85].

Σε μελέτη φυλογενετικής ανάλυσης 45 αντιπροσωπευτικών κορονοϊών από 18 χώρες, συμπεριλαμβανομένων των SARS-CoV, SARS-CoV-2, HCoV, SARS CoV νυχτερίδας, CoV τύπου SARS νυχτερίδας και MERS- CoV, η Yosra A. Helmy και οι συνεργάτες της, κατασκεύασαν το φυλογενετικό δέντρο με βάση τις αλληλουχίες ολόκληρου του γονιδιώματος. Στη μελέτη τα ιικά γονιδιώματα ελήφθησαν από τις βάσεις δεδομένων GISAID και NCBI, η στοίχιση πολλαπλών αλληλουχιών έγινε με τη χρήση του λογισμικού kalign 3 [86], το δέντρο σχεδιάστηκε στο λογισμικό IQ-TREE χρησιμοποιώντας τη μέθοδο της μέγιστης πιθανότητας, την Υπερταχεία Προσέγγιση bootstrap και το λογισμικό ModelFinder [87,88]. Το δέντρο σχεδιάστηκε σε κλίμακα, με τα μήκη κλαδιών να μετρώνται στον αριθμό των νουκλεοτιδικών αντικαταστάσεων ανά τοποθεσία. Οι τιμές του bootstrap προσδιορίστηκαν από 1.000 επαναλήψεις. Το δέντρο απεικονίστηκε στο λογισμικό MEGA X [89] (**Εικόνα 4**).

Στην ανάλυση όλα τα δείγματα SARS-CoV-2 από τις 18 χώρες ομαδοποιήθηκαν και βρέθηκαν πως είναι φυλογενετικά κοντά στον ιό που προκαλεί SARS στις νυχτερίδες, σε κορονοϊούς τύπου SARS, με τον ιό CoV RaTG13 της νυχτερίδας Wuhan να είναι φυλογενετικά ο πλησιέστερος. Επιπλέον, ο MERS-CoV και ο ανθρώπινος CoV HKU1 απείχαν πολύ από τον SARS-CoV-2. Στα δείγματα MERS-CoV, ο ιός MERS-NL13892 της Νότιας Κίνας ομαδοποιήθηκε χωριστά από τους υπόλοιπους MERS-CoV. Οι δύο κορονοϊοί που μοιάζουν με τον SARS και μολύνουν νυχτερίδες (bat-SL-CoVZC45 και bat-SL-CoVZXC21) ήταν οι δεύτεροι πιο κοντινοί φυλογενετικά στον SARS-CoV-2 (**Εικόνα 4**). Οι SARS-CoV από Κίνα, Καναδά, Αγγλία και ΗΠΑ ομαδοποιήθηκαν σε ένα ενιαίο σύμπλεγμα.

Ο Zhou και οι συνεργάτες του διεξήγαγαν φυλογενετική ανάλυση του SARS-CoV-2 σε σχέση με τους υπόλοιπους ταυτοποιημένους κορονοϊούς με βάση τις αλληλουχίες ολόκληρου του γονιδιώματος, τα γονίδια των κύριων δομικών πρωτεϊνών και τα γονίδια των μη δομικών πρωτεϊνών. Η ομαδοποίηση του SARS-CoV-2 ήταν διέφερε ανάλογα με το αν

χρησιμοποιήθηκε ολόκληρο το γονιδίωμα ή κάποια συγκεκριμένα γονίδια. Για παράδειγμα, ο SARS-CoV-2 ομαδοποιήθηκε εμφανίζοντας μεγαλύτερη συγγένεια με τα μέλη του υπογένους Sarbecovirus, συμπεριλαμβανομένου του SARS-CoV (79,5% ταύτιση) και τους άλλους ιούς που μοιάζουν με SARS και μολύνουν νυχτερίδες (96% ταύτιση σε επίπεδο ολόκληρου του γονιδιώματος). Η θέση του SARS-CoV-2, εντός των Sarbecoviruses, άλλαξε όταν χρησιμοποιήθηκαν μεμονωμένα γονίδια (ORF1ab, S, E, M και N) για ομαδοποίηση [84,90].



Εικόνα 4: Φυλογενετικό δέντρο βασισμένο στις πλήρεις γονιδιακές αλληλουχίες 45 αντιπροσωπευτικών κορονοϊών από 18 χώρες. [91]

1.5 Μεταλλάξεις και Παραλλαγές του SARS-CoV-2

Οι μεταλλάξεις προκύπτουν ως φυσικό παραπτροϊόν της ιϊκής αντιγραφής [92]. Οι RNA ιοί έχουν αρκετά υψηλότερα ποσοστά μετάλλαξης από τους DNA ιούς. Ωστόσο οι κορονοϊοί κάνουν λιγότερες μεταλλάξεις από τους περισσότερους RNA ιούς, καθώς κωδικοποιούν ένα ένζυμο που διορθώνει ορισμένα από τα σφάλματα που έγιναν κατά την αντιγραφή. Στις περισσότερες περιπτώσεις, η πορεία που θα έχει μια νεοεμφανιζόμενη μετάλλαξη, ορίζεται από τη φυσική επιλογή. Αποτέλεσμα αυτού είναι η αύξηση της συχνότητας εμφάνισης μεταλλάξεων οι οποίες παρέχουν ανταγωνιστικό πλεονέκτημα σε σχέση με τον πολλαπλασιασμό του ιού, τη μετάδοση, ή τη διαφυγή από την ανοσία, ενώ αντίθετα μεταλλάξεις οι οποίες επηρεάζουν αρνητικά την ικανότητα πολλαπλασιασμού ή μετάδοσης του ιού τείνουν να εξαφανίζονται από τον γενικό πληθυσμό των ιών. Ωστόσο, η συχνότητα των μεταλλάξεων μπορεί να αυξηθεί ή να μειωθεί λόγω τυχαίων γεγονότων. Για παράδειγμα, ένα «ιδρυτικό αποτέλεσμα» εμφανίζεται όταν ένας μικρός αριθμός ιών δημιουργεί έναν νέο πληθυσμό κατά τη μετάδοση, με αποτέλεσμα οι μεταλλάξεις των γονιδιωμάτων της αρχικής ομάδας των ιών να κυριαρχήσουν τοπικά στον ιϊκό πληθυσμό ανεξάρτητα από τις επιπτώσεις τους στα χαρακτηριστικά του ιού. Μετά από την επίδραση της φυσικής επιλογής και της τύχης οι ιικοί πληθυσμοί με τις νέες μεταλλάξεις «αξιολογούνται» για το εάν θα κυριαρχήσουν, θα μείνουν στάσιμοι ή αν θα εξαλειφθούν έναντι των υπολοίπων.

Οι όροι μετάλλαξη, παραλλαγή και στέλεχος χρησιμοποιούνται συχνά εναλλακτικά για την περιγραφή της επιδημιολογίας του SARS-CoV-2, κάνοντας απαραίτητη την ανάγκη διευκρίνησης των μεταξύ τους σημειολογικών διαφορών. Με τον όρο μετάλλαξη αναφερόμαστε στην πραγματική αλλαγή στην αλληλουχία του γονιδιώματος, για παράδειγμα ή μετάλλαξη D614G του SARS-CoV-2 είναι μια σημειακή νουκλεοτιδική υποκατάσταση που έχει ως αποτέλεσμα την αλλαγή του αμινοξέος στη θέση 614 της ακίδας της γλυκοπρωτεΐνης από ασπαρτικό οξύ σε γλυκίνη. Τα γονιδιώματα που διαφέρουν ως προς την νουκλεοτιδική αλληλουχία ονομάζονται συχνά παραλλαγές. Αυτός ο όρος είναι κάπως λιγότερο ακριβής καθώς δύο παραλλαγές μπορεί να διαφέρουν κατά μία ή περισσότερες μεταλλάξεις, πιο συγκεκριμένα, μια παραλλαγή θεωρείται στέλεχος όταν έχει αποδειγμένα διαφορετικό φαινότυπο όπως για παράδειγμα διαφορά στην αντιγονικότητα, στη μεταδοτικότητα ή στη λοιμογόνο δύναμη.

Η πλειονότητα των παραλλαγών συνήθως δεν έχει καμία επίδραση στη λειτουργία του ιού. Ωστόσο, υπάρχουν ορισμένες παραλλαγές που η αλλαγή στην πρωτεΐνική σύσταση της ακίδας του SARS-CoV-2, έχει επίδραση στη δέσμευση του ιού στον υποδοχέα ACE2, στη διάσπαση της ακίδας του ιού από την TMPRSS2 διευκολύνοντας έτσι την είσοδο του στα κύτταρα του ξενιστή, στη διαφυγή από την ανοσία, στην αύξηση της μεταδοτικότητας, στη νοσηρότητα, στα κλινικά συμπτώματα στις επιπτώσεις ή στη μείωση της ανταπόκρισης σε πιθανή θεραπεία [93]. Αυτές οι παραλλαγές ονομάστηκαν παραλλαγές ανησυχίας (Variants of Concern - VOCs) και παραλλαγές ενδιαφέροντος (Variants of Interest - VOIs). Η ονομασία και η παρακολούθηση αυτών των παραλλαγών πραγματοποιήθηκαν από τον Παγκόσμιο Οργανισμό Υγείας (WHO; www.who.int/), τις βάσεις δεδομένων γονιδιώματος GISAID (Παγκόσμια Πρωτοβουλία για την Κοινή χρήση όλων των Δεδομένων Γρίπης) [94] και Nextstrain [95] καθώς και από το επιδημιολογικό εργαλείο για την Φυλογενετική Ανάθεση της Καταγωγής της Επώνυμων Παγκόσμιων Επιδημιών (PANGOLIN) [96]. Όλες οι παραπάνω βάσεις δεδομένων μοιράζονται ελεύθερα γονιδιωματικά δεδομένα και εξυπηρετούν την άμεση επιτήρηση όλων αυτών των νέων και σημαντικών παραλλαγών.

Μέχρι τις αρχές Αυγούστου 2021, έχουν ανιχνευθεί τέσσερις VOCs (Alpha έως Delta) και έξι VOIs (Epsilon έως Lamda). Στις 30 Αυγούστου 2021, μια νέα παραλλαγή του SARS-CoV-2 χαρακτηρίστηκε ως παραλλαγή Mu και ταξινομήθηκε επίσης ως VOI από τον ΠΟΥ [97]. Το Σεπτέμβριο του 2021, το σύστημα ταξινόμησης των παραλλαγών άλλαξε και σύμφωνα με τον ΠΟΥ, οι παραλλαγές διακρίθηκαν σε VOCs (Άλφα έως Δέλτα), VOIs (Lamda και Mu) και

παραλλαγές υπό παρακολούθηση (Variants under Monitoring) [97]. Η τελευταία ομάδα περιλαμβάνει όλες τις άλλες παραλλαγές, που προηγουμένως αναφέρθηκαν ως VOI, εξαιρουμένων των παραλλαγών Zeta και Theta[96].

Alpha VOC (B.1.1.7)

Τον Σεπτέμβριο του 2020, η πρώτη VOC του SARS-CoV-2 εμφανίστηκε στο Ηνωμένο Βασίλειο, στην Αγγλία. Αυτό το νέο στέλεχος SARS-CoV-2 ανήκει στη γενεαλογία B.1.1.7, προέρχεται από τον κλάδο 20I (V1) και ονομάστηκε Alpha VOC από τον ΠΟΥ.

Αυτή η VOC περιέχει εππά μεταλλάξεις, υποκατάστασης αμινοξέος σε όλο το γονίδιο της Spike και τρεις μεταλλάξεις απαλοιφής (p.H69del, p.V70del και p.Y144del) εντός της N-τερματικής περιοχής (NTD) της υπομονάδας S1 της Spike. Η πλήρης διαγραφή αμινοξέων στις θέσεις 69-70 της πρωτεΐνης ακίδας είναι το αποτέλεσμα μιας διαγραφής έξι νουκλεοτιδίων, η οποία πιθανώς επηρεάζει την αναγνώριση του ιού από τα αντισώματα εξουδετέρωσης (Nabs) [98]. Η μετάλλαξη απαλοιφής στη θέση 144 είναι το αποτέλεσμα μιας απαλοιφής τεσσάρων νουκλεοτιδίων που σχετίζεται επίσης με τη διαφυγή αναγνώρισης από τα αντισώματα [99]. Δύο μεταλλάξεις αυτής της VOC, p.N501Y και p.A570D, εντοπίζονται στην περιοχή δέσμευσης του Υποδοχέα (RBD) της υπομονάδας S1. Η μετάλλαξη p.N501Y είναι ίσως η πιο σημαντική καθώς ενισχύει την ιική σύνδεση με τον υποδοχέα ACE2 του ξενιστή, συμβάλλοντας στη μεταδοτικότητα αυτής της VOC [100-102].

Μελέτες *in vivo* έχουν δείξει ότι τα μολυσμένα χάμστερ με p.N501Y και p.H69del, p.V70del οδηγούν σε υψηλά επίπεδα ιού στις ρινικές εκκρίσεις και στα επιθηλιακά κύτταρα του ανθρώπινου αεραγωγού [103]. Η μετάλλαξη p.P681H έχει επίδραση στη διαφυγή της ανοσίας [99], η θέση της δίπλα στη θέση που διασπάται από τη φουρίνη, παίζει σημαντικό ρόλο στην σύντηξη του λιπιδικού φακέλου (Ε) του ιού με αυτήν του κυττάρου του ξενιστή. Η μετάλλαξη p.P681H είναι μοναδική για την Alpha VOC.

Η παραλλαγή Alpha μπορεί να είναι περίπου 50% πιο μεταδοτική από τον άγριο τύπο SARS CoV-2, πιθανώς λόγω της υψηλότερης συγκέντρωσης του ιού σε σχέση με προηγούμενα κυκλοφορούντα στέλέχη [104,105]. Είναι ενδιαφέρον ότι η πλειοψηφία των μολυσμένων ασθενών που νοσηλεύονται ήταν γυναίκες, οι οποίες έχουν υψηλότερο ποσοστό θνησιμότητας. Η παραλλαγή Alpha χαρακτηρίζεται από μια πιο χαμηλή αντίσταση στα μονοκλωνικά αντισώματα έναντι της RBD ή του N-τελικού άκρου της πρωτεΐνης ακίδας, που παράγονται από άτομα που νόσησαν, σε σχέση με εμβολιασμένους [106].

Beta VOC (B.1.351)

Τον Μάιο του 2020, μια δεύτερη VOC του ιού SARS-CoV-2 εμφανίστηκε στη Νότια Αφρική. Αυτή η παραλλαγή ανήκει στη σειρά B.1.351, προέρχεται από τον κλάδο 20H (V2) και ονομάστηκε Beta VOC από τον ΠΟΥ.

Αυτή η VOC περιλαμβάνει εππά μεταλλάξεις υποκατάστασης αμινοξέων καθώς και μία μετάλλαξη απαλοιφής τριών αμινοξέων, p.L241del, p.L242del και p.A243del. Η ακριβής θέση της διαγραφής είναι άγνωστη (241-243del ή 242 244del), αλλά η αλληλουχία της πρωτεΐνης Spike παραμένει αμετάβλητη [107]. Οι μεταλλάξεις p.N501Y και p.D614G είναι κοινές μεταξύ των Alpha και Beta. Η Beta φέρει δύο νέες μεταλλάξεις στην περιοχή RBD της πρωτεΐνης Spike, p.K417N και p.E484K.

Έρευνες του Cele και των συνεργατών του υποστηρίζουν επίσης ότι η Beta VOC, πιθανώς λόγω της μετάλλαξης p.E484K, μπορεί να είναι πιο ανθεκτική στα μονοκλωνικά αντισώματα έναντι των πρωτεΐνών του σε σχέση με την Alpha ακόμη και σε εμβολιασμένα ή άτομα που είχαν νοσήσει [108,109].

Επίσης μελέτες του Nelson και των συνεργατών έδειξαν ότι ο συνδυασμός των μεταλλάξεων p.E484K, p.K417N και p.N501Y, που αλλάζουν τη δομή της πρωτεΐνης ακίδας, ενισχύει τη συγγένεια του ιού με τον υποδοχέα ACE2, γεγονός που πιθανώς καθιστά αυτήν την παραλλαγή πιο μεταδοτική [110].

Gamma VOC (P.1)

Τον Νοέμβριο του 2020, μια τρίτη VOC του ιού SARS-CoV-2 εμφανίστηκε στο Μανάους, την πρωτεύουσα της πολιτείας Amazonas της Βραζιλίας [111,112]. Αυτή η παραλλαγή ανήκει στη γενεαλογία P.1, προέρχεται από τον κλάδο 20J (V3) και ονομάστηκε Gamma VOC από τον ΠΟΥ.

Αυτό το VOC περιλαμβάνει 12 μεταλλάξεις υποκατάσταση αμινοξέος στην πρωτεΐνη Spike. Οι Alpha, Beta και Gamma VOCs μοιράζονται τις μεταλλάξεις p.N501Y της περιοχής RBD και τη p.D614G της περιοχής πλησίον της θέσης διάσπασης της Spike από τη φουρίνη. Οι Beta και Gamma VOCs μοιράζονται επίσης τις μεταλλάξεις p.K417N και p.E484K της περιοχής RBD της πρωτεΐνης ακίδας.

Οι p.K417T, p.E484K και p.N501Y σχετίζεται με αυξημένη δέσμευση του ιού στον ανθρώπινο υποδοχέα ACE2 και δυναμικότητα της μεταδοτικότητας [113]. Τα αντισώματα που παράγονται από εμβόλια ή νόσηση είναι λιγότερο πιθανό να εξουδετερώσουν την Gamma VOC [114,115]. Κατά συνέπεια, η εμφάνιση της παραλλαγής Gamma εγείρει ανησυχίες σχετικά με τον αντίκτυπο της στην ανοσία και τις πιθανές στρατηγικές μόλυνσης και διασποράς της.

Delta VOC (B.1.617.2)

Τον Οκτώβριο του 2020, μια τέταρτη VOC του SARS-CoV-2 εμφανίστηκε στην Ινδία και μέχρι τώρα έχει εξαπλωθεί και στον υπόλοιπο κόσμο [116]. Αυτή η παραλλαγή ανήκει στη σειρά B.1.617.2, προέρχεται από τον κλάδο 21A και ονομάστηκε Delta VOC από τον ΠΟΥ.

Η Delta διαφέρει πολύ από τις προηγούμενες παραλλαγές καθώς η μόνη κοινή μετάλλαξη είναι η p.D614G. Έχει δύο νέες μεταλλάξεις, τις p.L452R και p.T478K στην περιοχή RBD της Spike και δύο μεταλλάξεις απαλοιφής δύο αμινοξέων εντός της υπομονάδας S1, των p.E156del και p.F157del.

Μελέτες *in vitro* έχουν δείξει ότι οι παραλλαγές του SARS-CoV-2 που φέρουν τη μετάλλαξη p.G142D, η οποία επηρεάζει την πρωτεϊνική σύσταση της περιοχής πλησίον της RBD, μπορούσαν να πολλαπλασιαστούν παρουσία μονοκλωνικών αντισωμάτων, επομένως αυτή η μετάλλαξη δύναται να καταστήσει τον ίο πιο ανθεκτικό απέναντι στη δράση του ανοσοποιητικού συστήματος [117].

Η μετάλλαξη p.L452R πιθανώς συμβάλλει στην ισχυρότερη σύνδεση του ιού με τον υποδοχέα ACE2, ενισχύοντας τη μολυσματικότητα αυτής της παραλλαγής. Επιπλέον η p.L452R μειώνει την αναγνώριση του ιού από τα μονοκλωνικά αντισώματα, αυξάνοντας έτσι την πιθανότατα νόσησης μετά τον εμβολιασμό [118-120]. Επιπλέον, η υπολογιστική ανάλυση, σχετικά με τη δυναμική μόλυνσης των παραλλαγών SARS-CoV-2, έδειξε ότι η μετάλλαξη p.L452R έχει παρόμοια λειτουργία με τη μετάλλαξη p.N501Y [121]. Αυτή η VOC χαρακτηρίζεται από υψηλά επίπεδα μετάδοσης [122].

Η νέα γενεαλογία AY (υπο-γενεαλογίες: AY.1, AY.2, AY.4 και AY.4.2.), η οποία προέρχεται από τη γενεαλογία B.1.617.2 και χαρακτηρίζεται από αυξημένη μεταδοτικότητα, ονομάστηκε «Delta Plus». Υπάρχουν αρκετές επιπλέον μεταλλάξεις της ακίδας στην παραλλαγή Delta Plus, όπως οι p.V70F, p.A222V, p.W258L και p.K417N [123,124]. Από τον Οκτώβριο του 2021, περισσότερες από 2000 αλληλουχίες του γονιδιώματος της παραλλαγής Delta Plus

έχουν εντοπιστεί σε τουλάχιστον 46 χώρες, συμπεριλαμβανομένων των ΗΠΑ, της Ινδίας και του Ηνωμένου Βασιλείου [125]. Σε σύγκριση με άλλες VOC, έχει αναφερθεί ότι η Delta Plus εμφανίζει αντίσταση στα μονοκλωνικά αντισώματα, όπως οι αγωγές μονοκλωνικών Casirivimab και Imdevimab, λόγω του ότι έχει αυξημένη συγγένεια με τον βλεννογόνο των πνευμόνων [126]. Στον παρακάτω πίνακα (Εικόνα 5) απεικονίζονται τα Επιδημιολογικά και γενετικά χαρακτηριστικά των 4 ανησυχητικών παραλλαγών SARS-CoV-2.

WHO nomenclature	PANGO lineage	Emergence		Spike mutations		Impact	Efficacy of vaccines
		Country	Time	S1 subunit (1–685aa)	S2 subunit (686–1273aa)		
Alpha	B.1.1.7	UK	September 2020	p.H69del, p.V70del, p.Y144del, p.N501Y, p.A570D, p.D614G, p.P681H	p.T71G, p.S982A, p.D1118H	↑Transmissibility Affects more [31] ↓resistant to monoclonal Abs [32]	90% for BNT162b2 [79] ↓Neutralizing activity with mRNA-1273 [84] 72% for Ad26.COV2.S [87] 74% for ChAdOx1 nCoV-19 [82] 89.3% for NVX-CoV2373 [93]
Beta	B.1.351	South Africa	May 2020	p.D80A, p.D215G, p.L241del, p.L242del, p.A243del, p.K417N, p.E484K, p.N501Y, p.D614G	p.A701V	↑Resistant to Abs than Alpha VOC [35, 36] ↑The affinity with the ACE2 receptor [37]	75% for BNT162b2 [79] ↓↓↓Neutralizing activity with mRNA-1273 [83] 64% for Ad26.COV2.S [87] 81.5% for ChAdOx1 nCoV-19 [90] 49.4% for NVX-CoV2373 [93] ↓↓↓Neutralizing activity with Gam-COVID-Vac [96]
Gamma	P.1	Brazil	November 2020	p.L18F, p.T20N, p.P26S, p.D138Y, p.R190S, p.K417T/N, p.E484K, p.N501Y, p.D614G, p.H655Y	p.T102I, p.V1176F	↑The affinity with the ACE2 receptor [40] ↑Transmissible [40] Resistant to neutralizing Abs [41, 42]	↓↓↓Neutralizing activity with mRNA-1273 [84] 50.4% for CoronaVac for symptomatic and 78% for mild SARS-CoV-2 infection [102]
Delta and Delta Plus	B.1.617.2	India	October 2020	p.T19R, (p.V70F)*, p.G142D, p.E156del, p.F157del, p.R158G, (p.A222V)*, (p.W258L)*, (p.K417N)*, p.L452R, p.T478K, p.D614G, p.P681R	p.D950N	↑Transmissible [49] ↑The affinity with the ACE2 receptor [45, 46] ↑Resistant to monoclonal Abs [47]	↓↓BNT162b2 compared to Alpha [81] 67% for ChAdOx1 nCoV-19 [82] ↓↓↓with Covaxin [104]

Εικόνα 5: Επιδημιολογικά και γενετικά χαρακτηριστικά των τεσσάρων ανησυχητικών παραλλαγών SARS-CoV-2 αα: αμινοξέα, Abs: αντισώματα, *Μεταλλάξεις της παραλλαγής Delta Plus, Ο αντίκτυπός τους στη φυσική μόλυνση και την αποτελεσματικότητα των εμβολίων: Εξουδετερωτική δραστηριότητα = ↓: μείωση 1–3 φορές, ↓↓: μείωση 3–5 φορές, ↓↓↓: >5πλάσια μείωση. [127]

Omicron VOC (B.1.1.529)

Η παραλλαγή Omicron εμφανίστηκε για πρώτη φορά στη Νότια Αφρική τον Νοέμβριο του 2021. Τις τελευταίες εβδομάδες, οι λοιμώξεις έχουν αυξηθεί κατακόρυφα, που συμπίπτει με την ανίχνευση της παραλλαγής B.1.1.529. Η πρώτη γνωστή επιβεβαιωμένη μόλυνση B.1.1.529 προήλθε από δείγμα που συλλέχθηκε στις 9 Νοεμβρίου 2021 [128].

Αυτή η παραλλαγή έχει μεγάλο αριθμό μεταλλάξεων, μερικές από τις οποίες είναι ανησυχητικές. Τα προκαταρκτικά στοιχεία υποδηλώνουν αυξημένο κίνδυνο επαναμόλυνσης με αυτήν την παραλλαγή, σε σύγκριση με άλλες VOCs. Ο αριθμός των περιπτώσεων αυτής της παραλλαγής φαίνεται να αυξάνεται σε όλες σχεδόν τις επαρχίες της Νότιας Αφρικής. Πολλά εργαστήρια έχουν υποδείξει ότι για μια ευρέως χρησιμοποιούμενη δοκιμή PCR, ένα από τα τρία γονίδια-στόχους δεν ανιχνεύεται (αποτυχία στόχου γονιδίου S) και επομένως αυτή η δοκιμή μπορεί να χρησιμοποιηθεί ως σημείο αναφοράς για αυτήν την παραλλαγή, έως ότου υπάρξει πλήρης καταγραφή της αλληλουχίας του. Με βάσει αυτήν την προσέγγιση, η αυτή παραλλαγή έχει αυξημένος ρυθμό εξάπλωσης σε σχέση με τις προηγούμενες, υποδηλώνοντας πως έχει πλεονέκτημα ανάπτυξης [128].

1.6 Μετάδοση και Μόλυνση

Όπως και οι άλλοι κορονοϊοί, η μετάδοση του SARS-CoV-2 μπορεί να συμβεί κυρίως μέσω μολυσμένων αναπνευστικών σταγονίδιων από την άμεση ή έμμεση επαφή με μολυσμένα άτομα [129]. Αυτά τα σταγονίδια έχουν διάμετρο >5-10μm και μπορούν να παραμείνουν αιωρούμενα σε ένα κλειστό χώρο για έως και 8-14 λεπτά ή να μεταδοθούν μέσω του αέρα από ένα μολυσμένο άτομο σε ένα υγιές άτομο [130-135]. Ο ίος είναι επίσης γνωστό ότι μεταδίδεται σε υγιή άτομα απευθείας με φιλί ή μέσω σεξουαλικής επαφής με ένα μολυσμένο άτομο [136,137]. Αρκετές μελέτες θεωρούν πως η έμμεση μετάδοση γίνεται μέσω των φωμιτών(λερωμένες επιφάνειες) και των κοπράνων-στοματικών οδών [136,138-140]. Σε μελέτες που διεξήχθησαν για την αξιολόγηση της πιθανότητας κάθετης μετάδοσης του ιού από τη μολυσμένη μητέρα στα νεογνά, παραμένει ασαφές εάν η μετάδοση του ιού θα μπορούσε να συμβεί κατά τη διάρκεια του φυσιολογικού τοκετού, καθώς οι τοκετοί με καισαρική τομή που πραγματοποιήθηκαν κατά τη διάρκεια της πανδημίας και τα νεογνά βρέθηκαν αρνητικοί στον SARS-CoV-2. Η μόνη εξαίρεση σε αυτό μέχρι στιγμής αναφέρθηκε από τον Alzamora και τους συνεργάτες του [141-143]. Επίσης τα παιδιά παρουσιάζουν παρόμοιο κίνδυνο μόλυνσης με αυτόν του γενικού πληθυσμού, αλλά έχουν χαμηλότερη πιθανότητα να αναπτύξουν σοβαρά συμπτώματα λόγω του χαμηλότερου ΙΙΚΟΥ φορτίου, των ηπιότερων ανοσολογικών αποκρίσεων και της μικρότερης ανοσολογικής βλάβης σε σύγκριση με σοβαρά προσβεβλημένους ενήλικες [144,145].

Ο ίος στοχεύει και μολύνει τα κύτταρα του ρινικού, του επιπεφυκότα ή του βλεννογόνου της στοματικής κοιλότητας, από τη στιγμή που θα εισπνευστούν ή θα εναποτεθούν μολυσμένα αναπνευστικά σωματίδια σε αυτές τις θέσεις. Οι υποδοχείς για το SARS-CoV-2, οι οποίοι βρίσκονται κυρίως στο επιθήλιο της ανθρώπινης αναπνευστικής οδού, διευκολύνουν την είσοδο και επιτρέπουν στον ίο να αναπαραχθεί και να εξαπλωθεί σε άλλα όργανα. Περίπου πέντε ημέρες μετά τη μόλυνση, ο ξενιστής μπορεί να μεταδώσει τον ίο εύκολα μέσω της εκπνοής του και να μολύνει δυνητικά και άλλους. Τα συμπτώματα που σχετίζονται με τη μόλυνση από τον ίο οφείλονται σε προσωρινή βλάβη σε κύτταρα διαφόρων οργάνων, όπως είναι οι πνεύμονες, το έντερο, τα νεφρά και τα κύτταρα των αιμοφόρων αγγείων [146].

Η μόλυνση από τον ίο πυροδοτεί μια αρχική φλεγμονώδη ανοσολογική απόκριση κατά την οποία λευκά αιμοσφαίρια που προέρχονται από τους λεμφαδένες, όπως τα βοηθητικά Τ-κύτταρα και τα κυτταροτοξικά Τ-κύτταρα, διεισδύουν στο σημείο της μόλυνσης για να εξαλείψουν τα μολυσμένα κύτταρα [147]. Επιπλέον προκαλείται πυρετός, αυξάνοντας έτσι τη θερμοκρασία του σώματος, ώστε να καταπολεμηθεί ο ίος και να διευκολυνθεί η ανοσολογική απόκριση [148,149]. Σε ασθενείς που αναπτύσσουν πιο σοβαρή νόσο, ο SARS-CoV-2 προκαλεί μια παρεκκλίνουσα ανοσολογική απόκριση από τον ξενιστή. Η υπερπαραγωγή προφλεγμονώδων κυτοκινών οδηγεί στη λεγόμενη «καταιγίδα κυτοκινών», η οποία, εάν δεν αντιμετωπιστεί μέσω ανοσοτροποποιητών οι οποίοι ρυθμίζουν την παραγωγή κυτοκινών, μέσω φαρμάκων όπως η δεξαμεθαζόνη ή μέσω αραιωτικών του αίματος (αντιπηκτικά, αντιθρομβωτικά), μπορεί να οδηγήσει σε θάνατο [150-153]. Αυτή η «καταιγίδα κυτοκινών», μπορεί να επηρεάσει αρνητικά πολλά όργανα, συμπεριλαμβανομένων του ήπατος, των νεφρών και των πνευμόνων [150,154]. Ειδικότερα στους πνεύμονες μπορεί να οδηγήσει σε διάσπαση των επιθηλιακών κυττάρων τα οποία επενδύουν τις κυψελίδες, προκαλώντας τη συσσώρευση υγρού σε αυτές, προκαλώντας πνευμονία και ορισμένες σοβαρές περιπτώσεις, σύνδρομο οξείας αναπνευστικής δυσχέρειας (ARDS) το οποίο μπορεί να προκαλέσει πνευμονική ανεπάρκεια ή ακόμα και θάνατο. Μελέτες έχουν δείξει ότι τα επίπεδα ορισμένων κυτοκινών μπορούν να συσχετιστούν με τη σοβαρότητα των συμπτωμάτων του COVID-19 [150].

Ο SARS-CoV-2 έχει αυξημένο ρυθμό μετάδοσης σε σύγκριση με τον SARS-CoV του 2002. Αυτό οφείλεται αρχικά στα χαρακτηριστικά του ιού (SARS-CoV-2) όπως το ότι μπορεί να παραμείνει βιώσιμος (μολυσματικός), με χρόνο ημίζωνς 6,8 ώρες σε πλαστικό, 5,6 ώρες σε

ανοξείδωτο χάλυβα, 3,5 ώρες σε χαρτόνι και 1,1 ώρες σε αερολύματα. Εάν ένα υγιές άτομο αγγίζει αυτές τις μολυσμένες επιφάνειες και στη συνέχεια τα μάτια, το στόμα ή τη μύτη του, τα σωματίδια του ιού μπορεί να εισέλθουν στα κύτταρα του ξενιστή και να ξεκινήσουν μόλυνση [10]. Επίσης ο ίος μπορεί να παραμείνει βιώσιμος στην επιφάνεια του δέρματος για έως και 9 ώρες, αυξάνοντας έτσι την πιθανότητα έκθεσης σε αυτόν [155]. Ένας ακόμη λόγος που συμβάλει στην αύξηση του ρυθμού μετάδοσης αποτελεί το γεγονός ότι οι διαφορές στη σοβαρότητα και την έναρξη των συμπτωμάτων του COVID-19 σε σύγκριση με το SARS μπορεί να οδηγήσουν σε αυξημένη επαφή με μολυσμένα άτομα. Για παράδειγμα, κατά τη διάρκεια της επιδημίας του SARS-CoV το 2002, τα πτοσοστά θνησιμότητας ήταν σχεδόν διπλάσια από αυτά που παρατηρούνται επί του παρόντος για την πανδημία COVID-19 (9,6% έναντι 5,4%, αντίστοιχα) [156,157] και τα μολυσμένα άτομα εμφάνισαν σαφή συμπτώματα αναπνευστικής δυσχέρειας σχεδόν αμέσως μετά τη μόλυνση τους [158-160]. Αντίθετα, τα άτομα που έχουν μολυνθεί με SARS-CoV-2 μπορεί να μην παρουσιάζουν σημάδια μόλυνσης ή ασθένειας για μια εβδομάδα ή περισσότερο, παρά το γεγονός ότι μπορούν να μεταδίδουν τον ίο [161]. Τα ασυμπτωματικά άτομα και τα άτομα με ήπια συμπτώματα μπορεί να έχουν μεγάλες ποσότητες ιού (ϊϊκό φορτίο) στην ανώτερη αναπνευστική οδό, συμβάλλοντας έτσι στην ταχεία εξάπλωση του. Επίσης σε ορισμένες περιπτώσεις, ακόμη και μετά την εξαφάνιση των συμπτωμάτων, άτομα που έχουν μολυνθεί από τον SARS-CoV-2 εξακολουθούν να μεταδίδουν τον ίο. Επιπλέον σε συγκεκριμένες περιπτώσεις μπορεί να έχουμε και επαναμόλυνση, δεδομένου ότι οι ανθρώπινοι κορονοϊοί είναι καλά εξοπλισμένοι ώστε να ανατρέψουν την πλήρη ανοσία του ξενιστή. Χαρακτηριστικό παράδειγμα αποτελεί η εμφάνιση νέων παραλλαγών του ιού που είναι πιο μεταδοτικές και ενδέχεται να έχουν αυξημένο κίνδυνο θανάτου, όπως η παραλλαγή της Νότιας Αφρικής (B.1.351), η παραλλαγή του Ηνωμένου Βασιλείου (B.1.1.7), και η παραλλαγή της Ινδίας (B.1.617.2), γνωστή επίσης ως παραλλαγή Δέλτα. Οι παραπάνω παραλλαγές του ιού αποτελούν μόνο ένα κλάσμα αυτών που έχουν προκύψει μέσα στο χρονικό διάστημα από την αρχή της πανδημίας μέχρι και σήμερα, δείχνοντας έτσι τους ταχύτατους ρυθμούς εξέλιξης των μηχανισμών που συμβάλουν στην αύξηση της μεταδοτικότητας, καθώς και εκείνων που υπονομεύουν τις υπάρχουσες ανοσοποιητικές άμυνες [162,163]. Σαφώς, χρειάζονται πρόσθετες μελέτες για την κατανόηση των κινδύνων που ενέχουν αυτές οι παραλλαγές.

1.7 Συμπτώματα

Τα συμπτώματα της νόσου COVID-19, σύμφωνα με τον ΠΟΥ, μπορεί να εμφανιστούν 2 έως 14 ημέρες μετά την έκθεση στον ίο, με διάμεσο χρόνο τις 5,1 ημέρες. Η περίοδος μετά την έκθεση και πριν εμφανιστούν συμπτώματα ονομάζεται περίοδος επώασης. Το 97,5% των ατόμων θα εμφανίσουν συμπτώματα έως και 11,5 ημέρες από την μόλυνση. Σε σπάνιες περιπτώσεις τα συμπτώματα εμφανίζονται μετά τις 14 ημέρες [164].

Τα πιο συχνά συμπτώματα είναι ο πυρετός, ο ξηρός βήχας, η σωματική εξάντληση και η απώλεια γεύσης ή όσφρησης. Τα συμπτώματα της COVID-19 πτοικίλουν και χωρίζονται σε συστημικά, αναπνευστικά, οφθαλμών-μύτης- λαιμού, γαστρεντερικά, κεντρικού νευρικού συστήματος και οφθαλμών [165].

Τα συστημικά συμπτώματα της COVID-19 είναι ο πυρετός, η κόπωση, η αρθραλγία, οι μυαλγίες, και τα εξανθήματα. Στα αναπνευστικά συμπτώματα ανήκουν, ο βήχας, η δύσπνοια, ο πόνος στο στήθος, η αιμόπτυση και ο συριγμός της αναπνοής. Ο πονόλαιμος, η καταρροή, η ζάλη, η ρινική συμφόρηση, η απώλεια γεύσης και όσφρησης και η ωταλγία υπάγονται στα συμπτώματα οφθαλμών-μύτης-λαιμού. Συμπτώματα όπως η ναυτία, ο εμετός, η διάρροια και το κοιλιακό άλγος ανήκουν στα συμπτώματα του

γαστρεντερικού. Η COVID-19 επηρεάζει το κεντρικό νευρικό σύστημα (ΚΝΣ), προκαλώντας πτονοκεφάλους, σύγχυση και αταξία. Τέλος η επιπεφυκίτιδα, η οφθαλμαλγία και η φωτοφοβία [165] αποτελούν τα οφθαλμολογικά συμπτώματα της COVID-19.

Όταν εξελίσσεται σοβαρά η COVID-19 τα συμπτώματα που εμφανίζονται είναι δύσπνοια, απώλεια όρεξης, σύγχυση, επίμονος βήχας, πόνος στο στήθος και υψηλός πυρετός ($>38^{\circ}\text{C}$). Οποιοσδήποτε παρουσιάσει τα παραπάνω συμπτώματα, πρέπει να αναζητήσει άμεσα ιατρική βοήθεια και περίθαλψη [166].

Η σοβαρότητα των συμπτωμάτων της COVID-19 κυμαίνεται από πολύ ήπια έως σοβαρή. Μερικοί άνθρωποι μπορεί εμφανίζουν επιδεινωμένα συμπτώματα, όπως δύσπνοια και πνευμονία, περίπου μία εβδομάδα μετά την έναρξη των συμπτωμάτων. Υπάρχει επίσης το ενδεχόμενο κάποιοι που νοσούν να μην εμφανίζουν καθόλου συμπτώματα, αλλά να μεταδίδουν τον κανονικά τον SARS-CoV-2. Έχει παρατηρηθεί ότι σε παιδιά και νέους, η συμπτωματολογία είναι ηπιότερη [167]. Κάποια παιδιά μπορεί να εμφανίζουν πολυσυστημικό φλεγμονώδες σύνδρομο, το οποίο μπορεί να επηρεάσει ορισμένα όργανα και ιστούς, για αρκετές εβδομάδες μετά την εμφάνιση της νόσου. Η πιθανότητα εμφάνισης του πολυσυστημικού φλεγμονώδους συνδρόμου είναι αρκετά σπάνια σε ενήλικες.

1.8 Ανίχνευση και Διάγνωση του SARS-CoV-2

Η ταχεία διάγνωση της λοίμωξης SARS-CoV-2 είναι ο ακρογωνιαίος λίθος για τον περιορισμό και τον έλεγχο της νόσου. Στηρίζεται σε μεγάλο αριθμό κριτηρίων όπως το ιστορικό του εκάστοτε περιστατικού, τα κλινικά συμπτώματα, τον εργαστηριακό έλεγχο του ορού του αίματος (ορολογία), τη μοριακή διάγνωση και τη αξονική τομογραφία (CT). Στις 2 Μαρτίου 2020 ο ΠΟΥ εξέδωσε οδηγίες για τον εργαστηριακό έλεγχο των ύποπτων περιπτώσεων μόλυνσης από τον ίο, τις προφυλάξεις που θα πρέπει να παρθούν για τη συλλογή των δειγμάτων, τον τρόπο συσκευασίας και αποστολής και τις μεθόδους ενίσχυσης της ανίχνευσης του RNA και των ιικών γονιδίων (N, E, S και RdRp) [168]. Ο SARS-CoV-2 χρησιμοποιεί τον ίδιο υποδοχέα εισόδου κυττάρων (hACE2) με τον SARS-CoV, οπότε τα στοματικά επιχρίσματα, , το αίμα, το βρογχοκυψελιδικό υγρό πλύσης (BALF), καθώς και τα πρωκτικά επιχρίσματα αποτελούν τα καλύτερα δείγματα για τη διάγνωση του ιού [169]. Η σωστή διάγνωση εξαρτάται κυρίως από τους παράγοντες που περιγράφονται παρακάτω.

Επιδημιολογική Ιστορία

Η αυστηρή παρακολούθηση του ιστορικού περιστατικών σε κλινικά ύποπτους ασθενείς είναι το πρώτο βήμα στην έγκαιρη διάγνωση της λοίμωξης COVID-19. Κλινικά ύποπτοι ασθενείς θεωρούνται αυτοί που υποφέρουν από πυρετό και συμπτώματα λοίμωξης της κατώτερης αναπνευστικής οδού και διαμένουν ή έχουν ταξιδέψει σε περιοχές ενδημικές για τον ίο, ή είχαν στενή επαφή με επιβεβαιωμένο ή ύποπτο κρούσμα καθώς ο SARS-CoV-2 μπορεί να μεταδοθεί και από ασυμπτωματικούς φορείς [78].

Εργαστηριακή Διάγνωση

Τα προφύλ των δειγμάτων αίματος ασθενών που έπασχαν από λοίμωξη SARS-CoV-2 έδειξαν αυξημένα επίπεδα C-αντιδρώσας πρωτεΐνης, ερυθροκυττάρων, μυοαιμοσφαιρίνης, ηπατικών και μυϊκών ένζυμων, D-διμερούς (σε σοβαρές περιπτώσεις), εμφανίζουν επίσης φυσιολογικό ή μειωμένο αριθμό λευκών αιμοσφαιρίων και λεμφοκυττάρων στο πρώιμο στάδιο της νόσου, με προχωρημένη λεμφοκυτταροπενία σε σοβαρές περιπτώσεις [78]. Επίσης οι ασθενείς στις ΜΕΘ, εμφάνισαν υψηλά επίπεδα των κυτοκινών IP10, IL2, IL7, IL10,

TNF-α και MIP1a και του παράγοντα διέγερσης αποικιών κοκκιοκυττάρων πλάσματος (GCSF) [79].

Ανίχνευση του ιού

Ο SARS-CoV-2 απομονώθηκε επιτυχώς από τα επιθηλιακά κύτταρα του αναπνευστικού ή από δείγματα βρογχοκυψελιδικού υγρού πλύσης (BALF) μολυσμένων ασθενών με τη χρήση κυττάρων HuH7 και Vero E6. Η απομόνωση του ιού επιβεβαιώθηκε με τεχνικές ανοσοφθορισμού αντισωμάτων χρησιμοποιώντας το αντίσωμα διασταυρούμενης νουκλεοπρωτεΐνης (NP).

Ορολογική και Μοριακή Διάγνωση

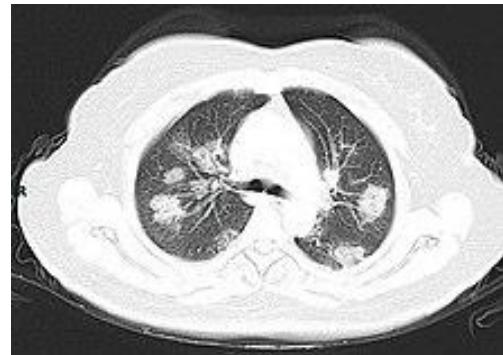
Τα κιτ ανίχνευσης ELISA IgM και IgG χρησιμοποιώντας τη νουκλεοπρωτεΐνη Rp3 του SARS-CoV νυχτερίδων, αναπτύχθηκαν χωρίς διασταυρούμενη αντίδραση έναντι των ανθρώπινων κορονοϊών εκτός του SARS-CoV [170]. Χρησιμοποιώντας αυτές τις τεχνικές ελέγχου του ορού του πλάσματος, εμφανίστηκε αύξηση των τίτλων των ΙΙΚών αντισωμάτων σε ασθενείς με COVID-19 [79]. Οι διαδικασίες της ELISA για την ανίχνευση του SARS-CoV-2 IgG περιγράφηκαν προηγουμένως [169]. Η ανίχνευση νουκλεϊκού οξέος είναι η κύρια, ταχύτερη και πιο ακριβής εξέταση για τη διάγνωση της λοίμωξης από SARS-CoV-2. Αναπτύχθηκαν δύο τεχνικές αναλύσεων αλυσιδωτής αντίδρασης πολυμεράσης μετά από αντίστροφη μεταγραφή (RT-PCR) και δύο τεχνικές RT-PCR πραγματικού χρόνου για την ανίχνευση των μολύνσεων [171]. Τρεις τεχνικές RT-PCR πραγματικού χρόνου έχουν σχεδιαστεί με βάση τα γονίδια E, RdRp και N [172]. Δημιουργήθηκαν ακόμα εργαλεία μοριακής ανίχνευσης για το SARS-CoV-2 με βάση το γονίδιο S της ακίδας [169]. Επίσης μέσω των τεστ τα αντιγόνου (rapid test) μπορεί να ανιχνευτεί η παρουσία μέρους της πρωτεΐνης επιφάνειας του ιού. Τα rapid test παρουσιάζουν μικρότερη ακρίβεια (ευαισθησία) και απαιτούν υψηλό ΙΙΚό φορτίο, με αποτέλεσμα σε ασυμπτωματικούς να μπορούν να βγουν ψευδώς αρνητικά. Τέλος υπάρχουν τα self test τα οποία μπορούν να διεξαχθούν από μη εξειδικευμένο προσωπικό, δηλαδή από το ευρύ κοινό καθώς ο τρόπος διεξαγωγής τους είναι αρκετά απλός.

Ακτινολογική Διάγνωση

Στα αρχικά στάδια της νόσου, η ακτινογραφία θώρακος (**Εικόνα 6**) εμφανίζει ενδιάμεσες αλλαγές και πολλαπλές μικρές σκιές στους πνεύμονες αλλά σύμφωνα με το Κέντρο Ελέγχου Λοιμώξεων (CDC) των ΗΠΑ δε συνιστάται η χρήση της ως μέσο διάγνωσης της Covid-19 λοίμωξης. Οι αξονικές τομογραφίες θώρακος (**Εικόνα 7**) διαδραματίζουν σημαντικό ρόλο στη διάγνωση του συνδρόμου οξείας αναπνευστικής νόσου (ARDS) και της πνευμονίας καθώς και στην έγκαιρη ανίχνευση ανωμαλιών του πνευμονικού παρεγχύματος σε ασθενείς σε κίνδυνο παρέχοντας μια εικόνα της δευτερογενούς λοίμωξης. Η αξιολόγηση αυτών των παρεγχυματικών ανωμαλιών των πνευμόνων εμφανίζει τη σοβαρότητα της νόσου.



Εικόνα 6: Ακτινογραφία πνευμόνων ασθενούς με πνευμονία λόγω της λοιμωξης Sars-CoV-2. [173]



Εικόνα 7: Αξονική τομογραφία ενός ατόμου με COVID-19 δείχνει βλάβες στους πνεύμονες του ασθενούς (φωτεινές περιοχές). [174]

Ο Pan και οι συνεργάτες του [175] διεξήγαγαν μια αναδρομική μελέτη ανασκόπησης για να επεξεργαστεί τη χρονική πορεία των αλλαγών στους πνεύμονες κατά τη διάρκεια της ανάρρωσης μετά από νόσηση COVID-19 και διαπίστωσαν πως στο περισσότερο από το 75% των ασθενών COVID-19 υπέστησαν αμφοτερόπλευρη προσβολή του πνεύμονα με το 71% να εμφανίζει πολυλοβική προσβολή. Επίσης το 86% των περιπτώσεων εμφάνισαν αμφίπλευρη πνευμονική προσβολή. Επιπλέον στο 80% των περιπτώσεων προσβλήθηκε το οπίσθιο τμήμα του πνεύμονα, ενώ στο 86%, η περιφέρεια του πνεύμονα [176].

1.9 Παθογένεση και Ανοσοπαθολογία του ιού

Οι εξαιρετικά παθογόνοι κορονοϊοί του ανθρώπου (HCoVs), συμπεριλαμβανομένων των SARS-CoV-2, MERS-CoV και SARS-CoV, μολύνουν την κατώτερη αναπνευστική οδό, τους πνεύμονες, τα νεφρά, το καρδιοαγγειακό σύστημα, τους ιστούς του δέρματος και το λεπτό έντερο [177]. Έπειτα η μόλυνση αυτή μπορεί να οδηγήσει σε σοβαρή πνευμονία, σύνδρομο οξείας αναπνευστικής δυσχέρειας (ARDS) και ανεπάρκειες οργάνων [178-181].

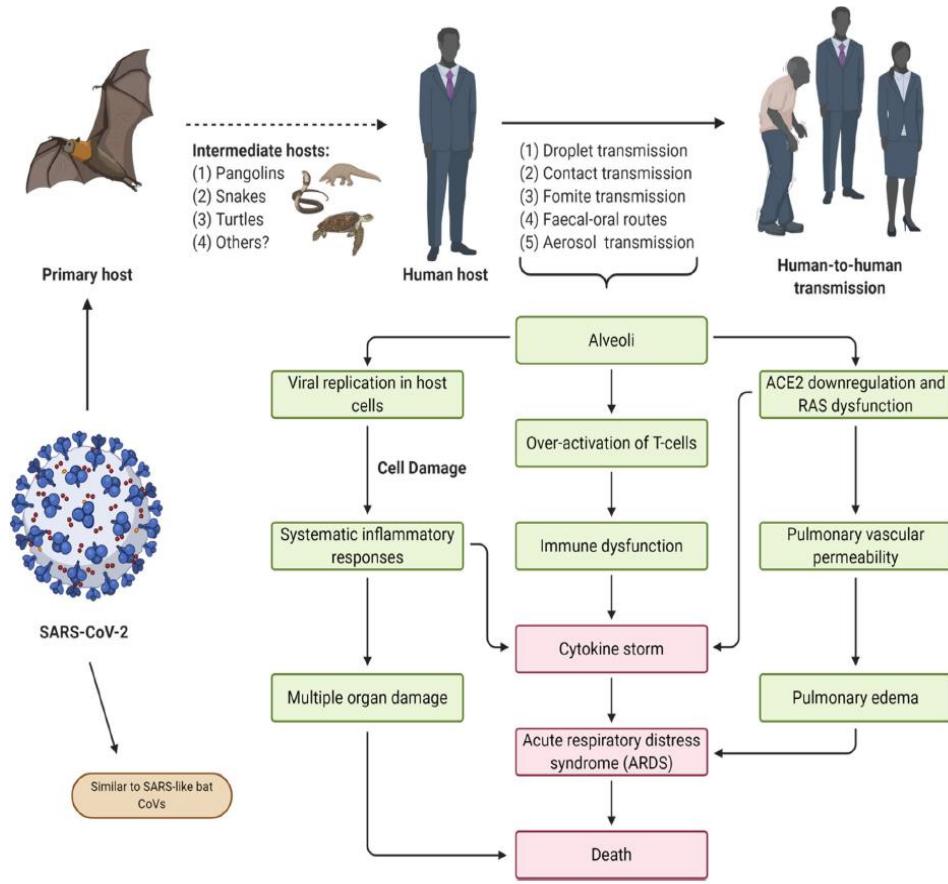
Ο SARS-CoV-2 χρησιμοποιεί τον υποδοχέα ACE2 του ξενιστή που εντοπίζεται στα κυψελιδικά κύτταρα των πνευμόνων, στα κύλικα κύτταρα του ρινικού βλεννογόνου και στα απορροφητικά εντεροκύτταρα της λάχνης σε συνεργασία με την πρωτεάση σερίνης TMPRSS2 του ξενιστή για να εισέλθει στα κύτταρα-ξενιστές [182-184]. Το «αστάρωμα» της ίικής γλυκοπρωτεΐνης S πυροδοτεί τη διάσπαση της πρωτεΐνης S από φουρινικές πρωτεάσεις του ξενιστή στη θέση διάσπασης S1/S2 και φέρνει τις μεμβράνες του ιών και του κυττάρου ξενιστή αρκετά κοντά διευκολύνοντας τη σύντηξη τους [185-188]. Στη συνέχεια, το ίικό γονιδίωμα απελευθερώνεται στα κύτταρα ξενιστές και ακολουθεί η αντιγραφή, η μετάφραση και η συσκευασία των νέων ίικών RNA στους φακέλους (E) των ιών απόγονων. Η ενεργή αντιγραφή και η απελευθέρωση ίικών σωματιδίων προκαλεί πυρόπτωση, μια φλεγμονώδη μορφή προγραμματισμένου κυτταρικού θανάτου, στα μολυσμένα κύτταρα [189,190]. Κατά την είσοδο του ιού στο κυττάρου μέσω της σύνδεσης στον υποδοχέα ACE2 του ξενιστή, ο ίος μειώνει την έκφραση του ACE2, που απαιτείται για τη διάσπαση της αγγειοτενσίνης-II (Ang II) ώστε να παραχθεί η αγγειοτενσίνης-1-7 [Ang (1-7)]. Η αύξηση της συγκέντρωσης της Ang II καθώς και το γεγονός πως αυτή αποτελεί το βασικό τελεστή του συστήματος ρενίνης-αγγειοτενσίνης (RAS), προκαλεί απορρύθμιση της σηματοδότησης του

RAS, οδηγώντας σε φλεγμονή πολλαπλών συστημάτων και σοβαρό τραυματισμό των πνευμόνων [191].

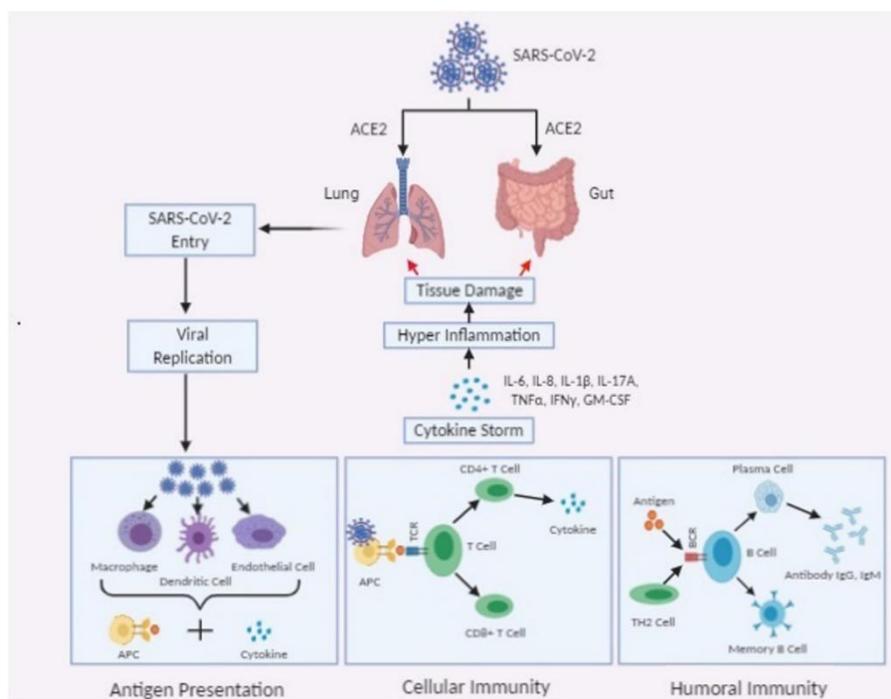
Ως συνέπεια της αύξησης της ρύθμισης του Ang II, τα μολυσμένα κυψελιδικά κύτταρα εκκρίνουν ειδικούς φλεγμονώδεις μεσολαβητές NLRP3, προάγοντας την έκκριση ιντερλευκινών IL-18, IL-1b και των μοριακών μοτίβων που σχετίζονται με ζημιές (DAMPs). Η απελευθέρωση αυτών των φλεγμονώδων κυτοκινών και των DAMP έχει ως αποτέλεσμα τον κυτταρικό θάνατο από πυρόπτωση. Αυτά τα μόρια αναγνωρίζονται επίσης από μακροφάγα των κυψελίδων και από τα γειτονικά κυψελιδικά κύτταρα, τα οποία προάγουν πολλαπλές τοπικές φλεγμονώδεις αποκρίσεις εκκρίνοντας ειδικές προφλεγμονώδεις κυτοκίνες και χημειοκίνες. Αυτές περιλαμβάνουν την ιντερφερόνη IFN- γ , την ιντερλευκίνη IL-6, την χημειοκίνη IP-10 και την κυτοκίνη MCP1, οι οποίες ευθύνονται για την προσέλκυση των T λεμφοκυττάρων, των μακροφάγων και των μονοκύτταρων στο σημείο του τραυματισμού [192-197]. Η διήθηση των κυττάρων του ανοσοποιητικού συστήματος στους αεραγωγούς προκαλεί έτσι περαιτέρω φλεγμονή δημιουργώντας έναν βρόχο ανατροφοδότησης, εξαπλώνοντας έτσι τη φλεγμονώδη αντίδραση και πυροδοτώντας καταιγίδα κυτοκινών [193]. Η προκύπτουσα καταιγίδα κυτοκινών όχι μόνο τραυματίζει τους πνεύμονες αλλά έχει επιπτώσεις και σε πολλά άλλα όργανα, προκαλώντας σηπτικό σοκ και πολλαπλή ανεπάρκεια οργάνων [198-200]. Εκτός από τις άμεσες βλάβες που σχετίζονται με μια ιογενή λοίμωξη, η φλεγμονώδης κυτταρική διήθηση, από τα κύτταρα του ανοσοποιητικού, προκαλεί σχηματισμό υαλώδους μεμβράνης, καταστροφή κυψελιδικών κυττάρων και πνευμονικό οίδημα που οδηγεί σε ARDS [195,196].

Για να ενισχύσουν την αντιϊκή ανοσία, τα κυψελιδικά μακροφάγα φαγοκυτταρώνουν τους ιούς, τους επεξεργάζονται και παρουσιάζουν μέσω των μορίων του κυρίου συμπλέγματος (MHC), τους αντιγονικούς επιτόπους στην επιφάνειά τους και ειδικά τους επιτόπους από τις γλυκοπρωτεΐνες της ακίδας. Αυτά τα επεξεργασμένα αντιγόνα αναγνωρίζονται από τα ειδικά κυτταροτοξικά T-λεμφοκύτταρα (CTL), εκκινώντας έτσι προσαρμοστικές ανοσολογικές αποκρίσεις που μεσολαβούνται από τα ειδικά για τον ιό B-λεμφοκύτταρα και T-λεμφοκύτταρα [201]. Μετά τη μόλυνση, τα επίπεδα των ειδικών για τον ιό αντισώματα IgG και IgM ήταν αυξημένα, όπως έδειξαν κλινικές μελέτες σε 285 ασθενείς. Όλοι οι ασθενείς πέτυχαν ορομετατροπή των IgM και IgG, διαδοχικά ή ταυτόχρονα εντός 20 ημερών μετά την εμφάνιση των κλινικών συμπτωμάτων [202]. Παρόμοια με τη μόλυνση του SARS-CoV, ο ειδικός για τον ιό τίτλος αντισωμάτων IgM αρχίζει να μειώνεται τρεις εβδομάδες μετά την έναρξη των συμπτωμάτων. Ωστόσο, τα αυξημένα επίπεδα αντισωμάτων IgG διαρκούν για μεγαλύτερη περίοδο, υποδηλώνοντας έναν αυξημένο προστατευτικό ρόλο που προσφέρει η IgG έναντι των ιών [202,203]. Εκτός από τις χυμικές αποκρίσεις, τόσο οι ασθενείς με SARS-CoV όσο και οι ασθενείς με SARS-CoV-2 υποφέρουν από λεμφοκυτταροπενία, όπου ο αριθμός των κυτταρικών ανοσοποιητικών CD8+ και CD4+ T-κύτταρων μειώνεται σημαντικά [196,204,205]. Μελέτες παρακολούθησης σε ασθενείς με SARS-CoV δείχνουν ότι τα CD8+ και CD4+ T-κύτταρων μνήμης παραμένουν για πιο εκτεταμένη περίοδο [206,207]. Έτσι, είναι σαφές ότι τα ειδικά για τον SARS-CoV T-κύτταρα είναι απαραίτητα τόσο για την εξάλειψη των ιών όσο και για τον έλεγχο της εξέλιξης της νόσου. Αντίθετα παραμένει ακόμα ασαφές το αν οι αποκρίσεις των T κυττάρων από μόνες τους μπορούν να ανακόψουν την πρόοδο της μόλυνσης. Περιέργως τα πρόσφατα στοιχεία δεν αποκλείουν την πιθανότητα η επαναμόλυνση με SARS-CoV-2 να παρουσιάζει πιο σοβαρά κλινικά συμπτώματα σε σύγκριση με την αρχική μόλυνση, θέτοντας υπό αμφισβήτηση την ικανότητα των προϋπαρχουσών ανοσολογικών αποκρίσεων να παρέχουν προστασία κατά τη δεύτερη έκθεση στον ιό. Επιπλέον, παρατηρήθηκαν διακυμάνσεις μεταξύ των απομονωθέντων ιών από τα σενάρια πρωτοπαθούς μόλυνσης και επαναμόλυνσης, παρατήρηση ή οποία εγείρει ανησυχίες σχετικά με το εάν ένα μεμονωμένο εμβόλιο μπορεί να προσδώσει ανοσία τόσο στο άτομο όσο και στον γενικό πληθυσμό [208,209]. Περαιτέρω μελέτες θα ρίξουν φως στην κατανόηση των διαδικασιών της ανοσοαπόκρισης έναντι του SARS-CoV-2. Στις παρακάτω

ΕΙΚΟΝΕΣ απεικονίζεται συνοπτικά η Μετάδοση, η Παθογένεση (Εικόνα 8) και η Ανοσοπαθογένεση της λοίμωξης (Εικόνα 9) από SARS-CoV-2



Εικόνα 8: Μετάδοση και Παθογένεση του SARS-CoV-2. Οι κύριοι ξενιστές εικάζεται ότι είναι νυχτερίδες, από τις οποίες ο ίος μπορεί να έχει μεταδοθεί σε έναν άνθρωπο ξενιστή μέσω ενδιάμεσων φορέων. Ένας μολυσμένος άνθρωπος ξενιστής μπορεί να μεταδώσει εύκολα τον ίο, σε άλλα άτομα μετά από ζωνόσωση. Τα πράσινα κουτιά υποδεικνύουν τα κύρια συμβάντα της λοίμωξης από SARS-CoV-2, συμπεριλαμβανομένων των αναπνευστικών ενοχλήσεων, της δυσλειτουργίας του συστήματος ρενίνης-αγγειοτενσίνης (RAS) και της απορρύμισης του ανοσοποιητικού συστήματος. Αν δεν υπάρχει η κατάλληλη ιατρική παρέμβαση, αυτά μπορεί να καταλήξουν σε δυνητικά θανατηφόρες καταστάσεις όπως το σύνδρομο οξείας αναπνευστικής δυσχέρειας (ARDS) και οι καταιγίδες κυτοκίνης. [210]



Εικόνα 9: Σχηματική αναπαράσταση της ανοσοπαθογένεσης της λοίμωξης από SARS-CoV-2. [211]

1.10 Δείκτες Αυξημένου Κινδύνου Σοβαρής Νόσησης

Υπάρχουν πολλοί παράγοντες που αυξάνουν τον κίνδυνο επιπλοκών και τη σοβαρότητα νόσησης με COVID-19 που μπορεί να καταλήξουν και σε θάνατο του πάσχοντος. Βασικός δείκτης κινδύνου είναι η ηλικία καθώς οι ηλικιωμένοι είναι πιο πιθανό να αρρωστήσουν σοβαρά από COVID-19. Πάνω από το 81% των θανάτων από COVID-19 είναι άτομα ηλικίας άνω των 65 ετών. Ο αριθμός των θανάτων μεταξύ ατόμων ηλικίας άνω των 65 ετών είναι 80 φορές υψηλότερος από τον αριθμό θανάτων μεταξύ ατόμων ηλικίας 18-29 ετών [212].

Οι ιατρικές καταστάσεις που αυξάνουν την σοβαρότητα της νόσησης όπως έχουν καταγραφεί μέχρι τώρα από τον ΠΟΥ είναι οι έξης:

- Καρκίνος καθώς οι θεραπείες για πολλούς τύπους καρκίνου μπορούν να αποδυναμώσουν το ανοσοποιητικό σύστημα. Αυτή τη στιγμή, με βάση τις διαθέσιμες μελέτες, το ιστορικό καρκίνου μπορεί να αυξήσει τον κίνδυνο [212].
- Καρδιακές παθήσεις, όπως καρδιακή ανεπάρκεια, στεφανιαία νόσο ή μυοκαρδιοπάθεια και η υψηλή αρτηριακή πίεση (υπέρταση) αυξάνουν τον κίνδυνο προσβολής και σοβαρών επιπλοκών [212]. Επιπλέον ο ίδιος μπορεί να προκαλέσει βλάβη στον καρδιακό μυ ή στα αγγεία, με αποτέλεσμα σοβαρά προβλήματα.
- Διαβήτης τύπου I ή τύπου II αυξάνουν τον κίνδυνο εισαγωγής στο νοσοκομείο και εμφανίζουν μεγαλύτερη θνησιμότητα σε σύγκριση με μη διαβητικά άτομα που έχουν μολυνθεί από SARS-CoV-2 [213,214].
- Το Υπερβολικό βάρος και η παχυσαρκία είναι από τις πιο σημαντικές καταστάσεις που αυξάνουν εκθετικά τον κίνδυνο θνησιμότητας του SarS-CoV-2 ασθενείς [215].
- Το κάπνισμα είναι ένας από τους παράγοντες κινδύνου για την ανάπτυξη και επιδείνωση πολλαπλών παθήσεων του αναπνευστικού, συμπεριλαμβανομένων των λοιμώξεων. Είναι ένας σημαντικός παράγοντας κινδύνου για λοίμωξη του ανώτερου και κατώτερου αναπνευστικού [216,217]. Τα διαθέσιμα στοιχεία υποδηλώνουν ότι το κάπνισμα σχετίζεται με αυξημένη σοβαρότητα της νόσου και θνησιμότητα σε νοσηλευόμενους ασθενείς με COVID-19.
- Η Χρόνια Νεφρική Νόσος (XNN) είναι ο πιο κοινός παράγοντας κινδύνου για θάνατο σε ασθενείς με COVID-19 παγκοσμίως με τον κίνδυνο να αυξάνεται ανάλογα με το στάδιο της νόσου, με τον υψηλότερο κίνδυνο να εμφανίζεται σε εκείνους με νεφρική ανεπάρκεια που λαμβάνουν θεραπεία υποκατάστασης [218-220]. Επιπλέον, οι ασθενείς που λαμβάνουν αιμοκάθαρση στο κέντρο διατρέχουν σημαντικά υψηλότερο κίνδυνο για COVID-19 από εκείνους που λαμβάνουν θεραπείες αιμοκάθαρσης στο σπίτι [221]. Αυτό εξηγείται από τον αυξημένο κίνδυνο έκθεσης στον ίο, λόγω της μεταφοράς τους, προς και από τη μονάδα αιμοκάθαρσης [222-224].
- Η Δρεπανοκυτταρική αναιμία ως μια ανοσοκατεσταλμένη κατάσταση, θέτει τους ασθενείς σε υψηλότερο κίνδυνο για λοιμώξεις του αναπνευστικού όπως η COVID-19 και επακόλουθες πνευμονικές επιπλοκές όπως το οξύ θωρακικό σύνδρομο (ACS) [225,226].
- Το εξασθενημένο ανοσοποιητικό σύστημα από μεταμοσχεύσεις οργάνων, μεταμοσχεύσεις μυελού των οστών ή από HIV, καθώς τα άτομα αυτά λαμβάνουν ανοσοκαταστατικά. Έχοντας ένα εξασθενημένο ανοσοποιητικό σύστημα αυξάνεται την πιθανότητα σοβαρής νόσησης από COVID-19 [212].
- Η Εγκυμοσύνη αποτελεί παράγοντα αυξημένου κινδύνου καθώς οι έγκυες πρόσφατα διατρέχουν αυξημένο κίνδυνο σοβαρής ασθένειας από COVID-19 σε σύγκριση με το γενικό πληθυσμό. Οι έγκυες που νοσούν από COVID-19 κατά τη διάρκεια της εγκυμοσύνης διατρέχουν επίσης αυξημένο κίνδυνο για πρόωρο τοκετό (γεννώντας το

μωρό νωρίτερα από 37 εβδομάδες) και θνησιγένεια και μπορεί να διατρέχουν αυξημένο κίνδυνο για άλλες επιπλοκές της εγκυμοσύνης. Η περίοδος κατά την οποία οι έγκυες έχουν αυξημένο κίνδυνο για σοβαρές επιπλοκές από νόσηση με COVID-19, διαρκεί από την αρχή της κύησης έως και 42 ημέρες μετά το τέλος αυτής [227].

- Οι χρόνιες πνευμονικές παθήσεις όπως η κυστική ίνωση ή η πνευμονική υπέρταση είναι παράγοντες αυξημένου κινδύνου λόγω του ότι ο SARS-CoV-2 πλήττει περισσότερο τους πνεύμονες και το άνω αναπνευστικό. Επομένως, οι θεραπείες που χρησιμοποιούνται από αυτούς τους ασθενείς θα πρέπει να συνεχιστούν και εκείνοι θα πρέπει να παρακολουθούνται όσο το δυνατόν περισσότερο [228].
- Η Χρόνια Ήπατική νόσος, όπως αυτή που σχετίζεται με το αλκοόλ, η μη αλκοολική λιπτώδης νόσος του ήπατος και η αυτοάνοση ηπατίτιδα, και ιδιαίτερα η κίρρωση ή οι ουλές στο ήπαρ, κάνει πιο πιθανό το ενδεχόμενο νόσησης από COVID-19 [212].
- Η Άνοια αποτελεί παράγοντα που μπορεί να αυξήσει την πιθανότητα μόλυνσης και βαριάς νόσησης από COVID-19 [212]. Γι' αυτό και θα πρέπει να παρέχεται πρόσθετη φροντίδα και στενή παρακολούθηση σε ασθενείς με άνοια για να ελαχιστοποιηθεί ο κίνδυνος λοιμώξεων, αποτρέποντας την ανάπτυξη σοβαρών εκβάσεων και θνητότητας [229].
- Το Σύνδρομο Down παρουσιάζει σημαντικά υψηλότερα ποσοστά ιατρικών επιπλοκών και θνησιμότητας από COVID-19. Μελέτες της Anke Huls και των συνεργατών της έδειξαν ότι ενώ τα κύρια σημεία και συμπτώματα που σχετίζονται με τον COVID-19 είναι παρόμοια με εκείνα του γενικού πληθυσμού, οι ασθενείς με σύνδρομο Down που νοσηλεύονται με COVID-19 ήταν πιο πιθανό να εμφανίσουν σοβαρή πορεία νόσου με υψηλότερα ποσοστά ιατρικών επιπλοκών και θνησιμότητας από το γενικό πληθυσμό, ιδιαίτερα ενήλικες άνω των 40 ετών [230].
- Οι Διαταραχές χρήσης ουσιών έχουν σοβαρές επιπτώσεις στον οργανισμό αυξάνοντας την πιθανότητα μόλυνσης και βαριάς νόσησης από COVID-19. Αυτό συμβαίνει κυρίως καθώς τα άτομα που κάνουν χρήση ναρκωτικών και ζουν σε συναθροίσεις, διατρέχουν αυξημένο κίνδυνο έκθεσης στον ίδιο. Επίσης η χρήση διεγερτικών, όπως η κοκαΐνη, η αμφεταμίνη και η μεθαμφεταμίνη, μπορεί να προκαλέσει οξέα προβλήματα υγείας όπως εγκεφαλικό επεισόδιο, καρδιακή προσβολή, μη φυσιολογικό καρδιακό ρυθμό και επιληπτικές κρίσεις, καθώς και πιο χρόνιες παθήσεις, όπως καρδιακή ή πνευμονική βλάβη. Η χρήση ναρκωτικών μέσω του καπνίσματος (μαριχουάνα κ.α.) μπορεί να προκαλέσει ή να επιδεινώσει ήδη υπάρχουσες πνευμονικές παθήσεις. Ακόμα παθήσεις που μειώνουν την δράση του ανοσοποιητικού όπως ο HIV, είναι πιο συχνές μεταξύ των ατόμων που κάνουν χρήση ναρκωτικών, ειδικότερα μεταξύ εκείνων που κάνουν ενέσιμα ναρκωτικά [231].

Τέλος οι μακροχρόνιες συστημικές ανισότητες υγείας και οι κοινωνικές ανισότητες έχουν θέσει διάφορες ομάδες ανθρώπων σε αυξημένο κίνδυνο να αρρωστήσουν και να πεθάνουν από τον COVID-19, συμπεριλαμβανομένων ατόμων από ορισμένες φυλετικές και εθνοτικές μειονότητες και άτομα με αναπηρίες. Μελέτες έχουν δείξει ότι άνθρωποι από φυλετικές και εθνοτικές μειονότητες πεθαίνουν επίσης από COVID-19 σε μικρότερες ηλικίες. Τα άτομα που ανήκουν σε μειονοτικές ομάδες είναι συχνά νεότερα όταν αναπτύσσουν χρόνιες ιατρικές παθήσεις και μπορεί να είναι πιο πιθανό να υποφέρουν από περισσότερες από μία παθήσεις [212].

1.11 Πρόληψη

Η εξάπλωση του ιού θα είναι εκθετικά αυξανόμενη αν δεν ληφθούν μέτρα πρόληψης. Τα μέτρα πρόληψης που πρέπει να λαμβάνονται για την ελαχιστοποίηση μετάδοσης του SAR-CoV-2 περιλαμβάνουν τη κοινωνική αποστασιοποίηση, τον αερισμό εσωτερικών χώρων, την κάλυψη του προσώπου μας όταν βήχουμε ή φτερνίζόμαστε, το πλύσιμο των χεριών και την υποχρεωτική χρήση μάσκας προσώπου σε εσωτερικούς χώρους και φυσικά τον εμβολιασμό. Ο εμβολιασμός δεν αναιρεί τα μετρά προφύλαξης, που παραμένουν σε ισχύ.

Στα άτομα που έχουν εκτεθεί ή είναι πιθανοί φορείς του ιού εμφανίζοντας ήπια συμπτώματα, συνίσταται να μείνουν εντός της οικίας τους και να αναζητήσουν ιατρική συνδρομή από την πλησιέστερη δομή υγείας (Νοσοκομείο, Κέντρο Υγείας) και τον ΕΟΔΥ.

Κοινωνική αποστασιοποίηση

Η κοινωνική αποστασιοποίηση στοχεύει στην επιβράδυνση της εξάπλωσης του ιού, ελαχιστοποιώντας την επαφή μεταξύ των ανθρώπων. Οι μέθοδοι της περιλαμβάνουν καραντίνα, ταξιδιωτικούς περιορισμούς, κλείσιμο σχολείων, σταδίων, θεάτρων, εμπορικών κέντρων, χώρων εργασίας (με χρήση της τηλεργασίας όπου είναι εφικτό) και αποφυγή συναθροίσεων και αποφυγή στενών επαφών μεταξύ των πολιτών(αποφυγή χειραψίας, εναγκαλισμών και κρατώντας αποστάσεις [232,233]. Σε περιοχές που έχουν υψηλό αριθμό κρουσμάτων επιβάλλεται lockdown από την εκάστοτε αρχή [234-236].

Αυτοαπομόνωση

Τα άτομα που νοσούν ή υποψιάζονται πως έχουν μολυνθεί με COVID-19, θα πρέπει σύμφωνα με τις οδηγίες από τον ΠΟΥ να μπαίνουν σε ατομική απομόνωση (καραντίνα). Πιο συγκεκριμένα θα πρέπει να απέχουν από την εργασία η αν είναι μαθητές, από το σχολείο, να μην κυκλοφορούν σε δημόσιους χώρους, να αποφεύγουν τη χρήση δημοσίων μέσων μεταφοράς ή ταξί [237]. Το χρονικό διάστημα της αυτοαπομόνωσης είναι 2 εβδομάδες (14 ημέρες) από την πιθανή έκθεση στον ιό [238].

Αναπνευστική Υγιεινή

Συμφωνά με τις οδηγίες του ΠΟΥ οι πολίτες θα πρέπει να καλύπτουν το στόμα και τη μύτη τους με ένα μαντήλι όταν βήχουν ή φταρνίζονται, το οποίο θα πρέπει έπειτα να απορρίπτεται άμεσα [239].

Οι πρακτικές που έχει εκδώσει ο ΠΟΥ για την χρήση της μάσκας για τους απλούς πολίτες είναι οι εξής:

- Η μάσκα πρέπει να καλύπτει μύτη και στόμα
- Η μάσκα θα πρέπει να δένεται με ασφάλεια ώστε να ελαχιστοποιούνται τα κενά μεταξύ του προσώπου και αυτής
- Η μάσκα θα πρέπει να αντικαθίσταται όταν θα είναι υγρή
- Ο χρήστης της μάσκας να αποφεύγει να αγγίζει το μπροστινό μέρος της μάσκας όταν την αφαιρεί.
- Μετά την αφαίρεση της μάσκας θα πρέπει να πλύνει τα χεριά του με σαπούνι ή απολυμαντικό.
- Οι μάσκες μιας χρήσης θα πρέπει να πετιούνται αμέσως μετά την αφαίρεση τους.

- Οι υγειονομικοί πρέπει να φορούν τουλάχιστον μάσκα με την πιστοποίηση κατά το πρότυπο NIOSH N95, πρότυπο EU FFP2 ή ισοδύναμο, πέραν των άλλων ατομικών προστατευτικών εξοπλισμών [240,241].

Τέλος οι μάσκες (**Εικόνα 10**) με βαλβίδα προστατεύουν μόνο το άτομο που τις φοράει (**Εικόνα 11**), ενώ αν ο χρήστης είναι φορέας δεν αποτρέπουν τη διασπορά του ιού.



Εικόνα 10: Τύποι Μασκών Προστασίας. [242]



Εικόνα 11: Αναπαράσταση αερολυμάτων μεταξύ μάσκας με βαλβίδα, μάσκας χωρίς βαλβίδα και χωρίς κανένα μέτρο προστασίας. [243]

Απολύμανση των χεριών

Το Κέντρο Ελέγχου Μολύνσεων (CDC) συνιστά το συχνό πλύσιμο των χεριών με σαπούνι (απορρυπαντικό) και νερό για τουλάχιστον 20 δευτερόλεπτα, όταν είναι ορατά βρώμικα, μετά από δραστηριότητες όπως ο καθαρισμός της μύτης, ο βήχας, το φτάρνισμα, η τουαλέτα καθώς και πριν από κάθε γεύμα. Όταν δεν είναι δυνατός ο καθαρισμός με σαπούνι, μπορούμε να χρησιμοποιήσουμε αλκοολούχο διάλυμα με τουλάχιστον 60% αλκοόλη. Ο ΠΟΥ συνιστά να μην αγγίζουμε τα μάτια, τη μύτη ή το στόμα [239].

Καθαρισμός επιφανειών

Οι επιφάνειες μπορούν να καθαριστούν με πλειάδα προϊόντων απολύμανσης περιεκτικότητας 60-70% αιθανόλης, 0,2-7,5% ποβιδόνη-ιώδιο, 50-100% ισοπροπανόλης, 0,5%, υπεροξείδιο υδρογόνου και 0,1% υποχλωριώδες νάτριο. Επιπλέον σε υγειονομικές μονάδες μπορεί να χρησιμοποιηθεί και η ακτινοβόληση με υπεριώδη ακτινοβολία ως ένας ακόμα αποτελεσματικός τρόπος απολύμανσης [244].

Αερισμός και φιλτράρισμα αέρα

Τα κεντρικά συστήματα θέρμανσης, εξαερισμού και κλιματισμού θα πρέπει να ρυθμίζονται, να καθαρίζονται και να απολυμαίνονται επειδή τα ιοσωμάτια του SAR-CoV-2 παραμένουν αερομεταφερόμενα για αρκετές ώρες. Επίσης οι αεραγωγοί απόρριψης θα πρέπει να έχουν την μεγαλύτερη δυνατή απόσταση από τα σημεία αναρρόφησης φρέσκου αέρα [245].

Εάν δεν υπάρχει σύστημα εξαερισμού, συνιστάται να ανοίγονται παράθυρα και πόρτες ανά τακτά χρονικά διαστήματα [245]. Επίσης η ρύθμιση της υγρασίας των εσωτερικών χώρων είναι αρκετά σημαντικός παράγοντας. Η διατήρηση της υγρασίας σε επίπεδα 40%-60% συμβάλει στην μείωση του χρόνου παραμονής των παθογόνων ιοσωματίων, εντός των σταγονιδίων, στους εσωτερικούς χώρους [245].

Υγιεινός τρόπος ζωής

Ο ΠΟΥ συστήνει υγιεινή διατροφή, σωματική άσκηση για τουλάχιστον 3 φορές την εβδομάδα από τουλάχιστον 30 λεπτά ανά ημέρα και επαρκείς ώρες ύπνου ως άμυνα απέναντι στην COVID-19 [246]. Η υγιεινή διατροφή περιλαμβάνει τροφές όπως άπαχο κρέας, ψαριά, αυγά, όσπρια, λαχανικά, φρούτα και χυμούς, τον περιορισμό της πρόσληψης ζάχαρης και αλκοόλ [247]. Επίσης εμφανίζεται ισχυρή συσχέτιση μεταξύ της έλλειψης βιταμίνης D και εκδήλωσης σοβαρής νόσησης από SARS-CoV-2 [248].

Εμβόλια

Από τη στιγμή που ή COVID-19 χαρακτηρίστηκε ως πανδημία στις αρχές του 2020, τα εμβόλια αποτελούν την πιο ρεαλιστική προσέγγιση για την πρόληψη της μόλυνσης από τον SARS-CoV-2 και τον έλεγχο της πανδημίας. Τα εμβόλια αναπτύχθηκαν με ταχύτατους ρυθμούς, μέσα σε μόνο αρκετούς μήνες και πάρα τις απρόβλεπτες αλλαγές στο γονιδίωμα του SARS-CoV-2 (VOCs και VOIs) η αποτελεσματικότητα στη πρόληψη εκδήλωση σοβαρής νόσου, διατηρείται προς το παρόν [249].

Τα εμβόλια κατά της COVID-19 (**Εικόνα 12**) έχουν αναπτυχθεί με πολλές διαφορετικές μεθόδους. Οι βασικότερες μέθοδοι ανάπτυξης εμβολίων στηρίζονται είτε στη χρήση ζωντανών ή εξασθενημένων ιών, είτε στη χρήση ανασυνδυασμένων πρωτεΐνων και mRNA. Η ανάπτυξη και η χορήγηση εμβολίων κατά του SARS-CoV-2, είναι το ισχυρότερο και αποτελεσματικότερο όπλο, που διαθέτουμε για τον έλεγχο και την εξάλειψη της πανδημίας.

Μέχρι σήμερα, δέκα εμβόλια έχουν λάβει έγκριση από τουλάχιστον μία εθνική ρυθμιστική αρχή για χρήση στο ευρύ κοινό. Αυτά περιλαμβάνουν τα δύο εμβόλια mRNA BNT162b2 των Pfizer και BioNTech και το εμβόλιο mRNA-1273 της Moderna, τα τέσσερα εμβόλια llκού φορέα ChAdOx1 nCoV-19/AZD1222 της Oxford-AstraZeneca, το Ad26.COv2.S της Johnson & Johnson, το Sputnik V του Ερευνητικού Ινστιτούτου Επιδημιολογίας και Μικροβιολογίας της Gamaleya στη Ρωσία, το AD5-nCOV της CanSino Biologics στην Κίνα, τα τρία συμβατικά αδρανοποιημένα εμβόλια το BBIBP-CorV της Sinopharm στην Κίνα, το Covaxin της Bharat Biotech στην Ινδία και το CoronaVac της SinovacBiotech στην Κίνα και ένα εμβόλιο πεπτιδίου το EpiVacCorona του Κρατικού Ερευνητικού Κέντρου Ιολογίας και Βιοτεχνολογίας Vektor στη Ρωσία [250]. Στην παρούσα εργασία θα εστιάσουμε στα εμβόλια των Pfizer και BioNTech, Moderna, AstraZeneca και Johnson & Johnson, που ήταν διαθέσιμα στην Ελλάδα.



Εικόνα 12: Τα πιο διαδεδομένα εμβόλια (COVID-19) των Moderna, AstraZeneca, Pfizer-BioNTech και Johnson & Johnson που δημιουργήθηκαν για την καταπολέμηση της πανδημίας. [251]

Εμβόλια mRNA

Τα δυο εμβόλια mRNA, BNT162b2 των Pfizer και BioNTech και το εμβόλιο mRNA-1273 της Moderna, αναπτυχθήκαν με βάση τη νέου τύπου τεχνολογία mRNA, με σκοπό την πρόληψη της μόλυνσης από τον SARS-CoV-2. Οι μέχρι τώρα κλινικές δοκιμές έχουν δείξει ότι είναι εξαιρετικά αποτελεσματικά όσον αφορά την ανάπτυξη ανοσίας και το περιορισμό στην εκδήλωση σοβαρής νόσου. Επίσης σε προγράμματα μαζικού εμβολιασμού εμφάνισαν ανεκτά επίπεδα αντιδραστικότητας (παρενεργειών) [252].

BNT162b2 (Pfizer και BioNTech)

Πρόσφατες μελέτες έχουν δείξει ότι υπάρχει συσχέτιση της ανοσολογικής απόκρισης του εμβολίου BNT162b2 με διάφορα επιδημιολογικά και κλινικά χαρακτηριστικά [253,254]. Επιπλέον, έχουν εκφραστεί ανησυχίες σχετικά με την αποτελεσματικότητα των εμβολίων έναντι των αναδυόμενων παραλλαγών του ιού.

Τα παραγόμενα αντισώματα (Nabs) του εμβολίου BNT162b2 φαίνεται να αναγνωρίζουν και να δρουν κατά της παραλλαγής Alpha, αλλά με μια ελαφρά μείωση της εξουδετερωτικής δραστηριότητας (NAc) σε σύγκριση με αυτή που παρατηρήθηκε στο άγριου τύπου του ιού, δηλαδή αυτόν που παρατηρήθηκε για πρώτη φορά στο Wuhan [250,255]. Η μείωση της NAc έναντι της παραλλαγής Alpha μπορεί να αποδοθεί στην μετάλλαξη υποκατάστασης της ακίδας p.E484K, καθώς αυτή η μετάλλαξη έχει προηγουμένως αποδειχθεί πως δεν αναγνωρίζεται από πολλά μονοκλωνικά αντισώματα [250].

Το BNT162b2 εξουδετερώνει την παραλλαγή Beta που φέρει αρκετές μεταλλάξεις στην περιοχή RBD πρωτεΐνης Spike. Μια πρόσφατη μελέτη έδειξε ότι οι δείκτες εξουδετέρωσης στον ορό ατόμων που εμβολιάστηκαν ήταν τρεις έως τέσσερις φορές χαμηλότεροι για το B.1.351 σε σύγκριση με αυτούς για τον άγριο τύπο του ιού (Wuhan), αλλά ήταν υψηλότεροι από αυτούς των ατόμων τα οποία αναρρώνουν μετά από νόσηση από τον ιό άγριου τύπου [256].

Σε μια μελέτη από το Κατάρ που περιλάμβανε περισσότερα από 265.000 εμβολιασμένα άτομα κατά την περίοδο που επικρατούσε τόσο η Alpha όσο η Beta, η αποτελεσματικότητα του BNT162b2 υπολογίστηκε στο 90% για τη μόλυνση από την Alpha και 75% για μόλυνση Beta [257]. Ωστόσο, μια εφάπταξ δόση BNT162b2 δεν εμφάνισε υψηλή αποτελεσματικότητα και έναντι των δύο παραλλαγών [257].

Σε μια πολυκεντρική μελέτη κοορτής, στην οποία συμμετείχαν πάνω από 23.000 εργαζόμενοι στον τομέα της υγείας στο Ηνωμένο Βασίλειο, για μια χρονική περίοδο κατά την οποία επικρατούσε η παραλλαγή Alpha, η αποτελεσματικότητα του BNT162b2 κατά της

λοίμωξης (ασυμπτωματική ή με συμπτώματα) εκτιμήθηκε στο 70% στις 21 ημέρες ή περισσότερες μετά την πρώτη δόση και στο 85% στις 7 ημέρες ή περισσότερες μετά τη δεύτερη δόση του BNT162b2, αντίστοιχα [258]. Η ανοσολογική απόκριση που επιτυγχάνεται μέσω του BNT162b2 ήταν τρεις έως πέντε φορές χαμηλότερη έναντι της παραλλαγής Delta σε σύγκριση με την Alpha λόγω των στρατηγικών διαφυγής ανοσίας από την αδυναμία αναγνώρισης της RBD περιοχής της Delta, η οποία περιέχει δύο νέες μεταλλάξεις (p.L452R και p.T478K) σε σχέση με την Alpha [259]. Σε μια μελέτη περίπτωσης ελέγχου, η αποτελεσματικότητα δύο δόσεων του BNT162b2 ήταν 94% έναντι της Alpha και 88% έναντι της Delta [260].

mRNA-1273 (Moderna)

Παρόμοια με το BNT162b2, το mRNA-1273 προσφέρει επαρκή κάλυψη σε επίπεδα αντισωμάτων εξουδετέρωσης έναντι της Alpha. Επιπλέον εμφανίζει 6–9 φορές χαμηλότερους δείκτες αντισωμάτων εξουδετέρωσης έναντι της Beta σε σχέση με τους δείκτες του ιού άγριου τύπου [261]. Πρόσφατα μελέτες του Wu και των συνεργατών του ανήνευσαν μειώσεις των δεικτών εξουδετερωτικής δραστηριότητας των αντισωμάτων κατά 1,2 βαθμούς έναντι της Alpha, 6,4 έναντι της Beta και 3,5 έναντι της Gamma [262]. Σε σύγκριση με προηγούμενα κυκλοφορούντα στελέχη, φαίνεται ότι η παραγωγή επιπέδων αντισωμάτων εξουδετέρωσης (Nabs) είναι χαμηλότερη έναντι της Beta και της παραλλαγής ενδιαφέροντος (VOI) Epsilon (B.1.617.2) [263,264]. Πρόσφατα ο McCallum και οι συνεργάτες του έδειξαν μείωση περίπου 2,5 φορές της εξουδετερωτικής δράσης έναντι της Epsilon σε ορό (πλάσμα) εμβολιασμένων με το mRNA-1273, αντίστοιχα [265].

Εμβόλια πικών φορέων

Τα εμβόλια Ad26.COV2.S (Johnson & Johnson) και ChAdOx1 nCoV-19/AZD1222 (AstraZeneca) βασίζονται στη χρήση αδενοϊού ως ανασυνδυασμένο φορέα και ανίκανο για αναπαραγωγή. Έχουν εγείρει αμφιλεγόμενα ζητήματα στην επιστημονική κοινότητα, κυρίως όσον αφορά τη συσχέτισή τους με τη αρκετά σπάνια παρενέργεια της θρομβωτικής θρομβοκυτοπενίας, παρά τα υψηλά ποσοστά ανοσογονικότητας, εξουδετερωτικής δραστηριότητας(NAc) καθώς και τη μεγάλη χρονική διάρκεια της ανοσίας του έναντι των πιο κοινών VOCs του SARS-CoV-2 [266-268].

ChAdOx1 nCoV-19/AZD1222 (AstraZeneca)

Σε τυχαιοποιημένη κλινική δοκιμή εμφάνισε σημαντικά ποσοστά αποτελεσματικότητας του εμβολίου ChAdOx1 nCoV-19 έναντι παραλλαγών που έχουν την δυνατότητα να αποφύγουν τις ανθρώπινες ανοσολογικές αποκρίσεις [269]. Παρά την *in vitro* επαγωγή χαμηλών αποκρίσεων αντισωμάτων εξουδετέρωσης (Nab) έναντι της Alpha, η αποτελεσματικότητά του ChAdOx1 nCoV-19 (70,4%) σε συμπτωματικούς ασθενείς που μολύνθηκαν από Alpha δεν παρουσίασε στατιστική διαφορά σε σχέση με την αποτελεσματικότητα του σε άλλες παραλλαγές (81,5%) [269]. Στη Νότια Αφρική, το ChAdOx1 nCoV-19 δεν μείωσε το ποσοστό της ήπιας έως μέτριας εκδήλωσης της νόσου COVID-19 όταν κυριαρχούσε η Beta, με την αποτελεσματικότητα του εμβολιασμού (*Vaccine Effectiveness, VE*) σε συμπτωματικούς ασθενείς που είχαν μολυνθεί από την Alpha να φτάνει το 70,4% και το 81,5% για όσους μολύνθηκαν από άλλες παραλλαγές εκτός της Alpha [269]. Ωστόσο, η *VE* σε ασυμπτωματικούς ασθενείς που μολύνθηκαν από την Alpha μειώθηκε στο 28,9% [269].

Όσον αφορά την *VE* έναντι των παραλλαγών VOC του SARS-CoV-2, σε μελέτη 63 εμβολιασμένων, εντοπίστηκε εξουδετερωτική δραστηριότητα (NAc) στο 60% των ατόμων, λόγω των στρατηγικών ανοσοδιαφυγής της παραλλαγής Delta [270]. Σε μελέτη περίπτωσης

ελέγχου, το ChAdOx1 nCoV-19 εμφάνισε VE σε ποσοστό 74% για την παραλλαγή Alpha και 67% για την παραλλαγή Delta [260].

Ad26.COV2.S (Johnson & Johnson)

Στην τυχαιοποιημένη δοκιμή ελέγχου φάσης III του εμβολίου Ad26.COV2.S, η αποτελεσματικότητα του εμβολιασμού (VE) υπολογίστηκε σε 66,9% στην πρόληψη της μέτριας έως σοβαρής λοίμωξης SARS-CoV-2. Για παράδειγμα τα εκτιμώμενα ποσοστά VE ήταν 74% και 72% στις 14 και 28 ημέρες αντίστοιχα, μετά τον εμβολιασμό κατά την περίοδο που στις ΗΠΑ επικρατούσαν παραλλαγές με την μετάλλαξη p.D614G (Beta, Gamma, Delta) [266]. Στη Νότια Αφρική, παρά τον υψηλό επιπολασμό της Beta κατά τη διάρκεια της δοκιμαστικής περιόδου, η VE του Ad26.COV2.S ενάντια στην εκδήλωση της μέτριας έως σοβαρής νόσησης, ήταν 52,0% στις 14 ημέρες και 64,0% στις 28 ημέρες μετά τον εμβολιασμό [266]. Στη Βραζιλία, όπου επικρατούσε η παραλλαγή Zeta (P.2), η VE ήταν 66% στις 14 ημέρες και 68% στις 28 ημέρες μετά τη χορήγηση του εμβολίου [266]. Αυτά τα δεδομένα είναι ενδεικτικά μιας σημαντικής ετερογένειας στα ποσοστά αποτελεσματικότητας του Ad26.COV2.S μεταξύ διαφορετικών περιοχών παγκοσμίως.

1.12 Τρόποι Αντιμετώπισης-Θεραπείας της νόσου

Αν ο ασθενής δεν αντιμετωπίζει σοβαρά συμπτώματα ο πιο σωστός τρόπος αντιμετώπισης είναι η απομόνωση στο σπίτι, όπως θα γινόταν για ένα κρυολόγημα ή τη γρίπη. Οι περισσότεροι ασθενείς με COVID-19 αναρρώνουν μετά από 2 εβδομάδες, χωρίς να είναι απαραίτητη η νοσοκομειακή περίθαλψη, οι πιο σοβαρές περιπτώσεις τείνουν να διαρκούν 3 έως 6 εβδομάδες και συνήθως χρειάζεται νοσηλεία σε υγειονομική μονάδα. Οι έρευνες για την παρασκευή νέων φαρμάκων για την αντιμετώπιση του COVID-19 βρίσκονται ακόμα σε εξέλιξη. Εντούτοις υπάρχουν πολλά πράγματα που μπορούν να ανακουφίσουν τα συμπτώματα, τόσο στο σπίτι όσο και στο νοσοκομείο [271].

Θεραπεία στο σπίτι

Εάν τα συμπτώματά που αντιμετωπίζει ο ασθενής, είναι αρκετά ήπια ώστε να μπορείτε να αναρρώσει στο σπίτι, οι οδηγίες που πρέπει να ακολουθήσει είναι:

- Η Ξεκούραση, καθώς συμβάλει στην ταχύτερη ανάρρωση.
- Η Απομόνωση στο σπίτι και η αποχή από την εργασία, το σχολείο και από άλλους δημόσιους χώρους, ώστε να αποφευχθεί η διασπορά του SARS-CoV-2.
- Η Ενυδάτωση, διότι κατά τη διάρκεια της νόσησης αυξάνεται ή απώλεια υγρών και ενδέχεται να προκληθεί αφυδάτωση, η οποία με τη σειρά της μπορεί να επιδεινώσει τη σοβαρότητα της νόσησης και να προκαλέσει και άλλα προβλήματα υγείας.
- Η Παρακολούθηση των συμπτωμάτων τα οποία αν επιδεινωθούν θα πρέπει μετά από ιατρικό έλεγχο και επαναξιολόγηση να κριθεί το αν θα χρειαστεί να συνεχιστεί η νοσηλεία στο σπίτι ή αν θα χρειαστεί να γίνει διακομιδή σε νοσοκομείο.
- Η Λήψη φαρμάκων όπως η ακεταμινοφαίνη (ή παρακεταμόλη) να μειώσει τον πυρετό σας, πάντα σε συνεννόηση με τον θεράποντα ιατρό.

Οι οδηγίες του CDC όσο αφορά το πότε μπορεί το άτομο που νόσησε να βγει από την απομόνωση αναφέρουν πως θα πρέπει να μην έχει εμφανίσει πυρετό για 3 ημέρες, τα αναπνευστικά σας συμπτώματα, όπως ο βήχας, η δύσπνοια να έχουν μειωθεί ή να έχουν σταματήσει εντελώς και τέλος θα πρέπει να έχουν περάσει τουλάχιστον 10 ημέρες από την έναρξη των συμπτωμάτων και να έχουν πραγματοποιηθεί δύο τεστ COVID-19, αρνητικά με διαφορά 24 ωρών.

Σε περίπτωση επιδείνωσης των συμπτωμάτων εάν ο ασθενής αρχίσει να έχει προβλήματα αναπνοής, πόνο ή πίεση στο στήθος, σύγχυση ή σοβαρή υπνηλία ή εάν τα χείλη ή το πρόσωπό εμφανίζουν μια κυανή απόχρωση, θα πρέπει μετά από άμεσο έλεγχο από τον θεράποντα ιατρό και αν κριθεί απαραίτητο να γίνει εισαγωγή σε νοσοκομείο ώστε να συνεχίσει ή θεραπεία εκεί.

Θεραπεία στο νοσοκομείο

Μετά την εισαγωγή στο νοσοκομείο ο ασθενής ενυδατώνεται ενδοφλέβια μέσω ορού, ώστε να αποφευχθεί αφυδάτωση. Επίσης παρακολουθείται στενά η αναπνοή και ο κορεσμός του οξυγόνου.

Εντός του νοσοκομείου ο ασθενής μπορεί να χρειαστεί να λάβει επιπλέον οξυγόνο είτε μέσω συμβατικής οξυγονοθεραπείας, είτε μέσω ρινικού καθετήρα υψηλής ροής (HFNC) [272]. Σε πολύ σοβαρές περιπτώσεις, ενδέχεται να γίνει διασωλήνωση κατά την οποία εισάγεται ένας ειδικός σωλήνας μέσω του στόματος στην τραχεία, ώστε να εξασφαλιστεί απευθείας ροή αέρα από και προς τους πνεύμονες. Προκειμένου να επιτευχθεί αυτό, πρέπει ο ασθενής να έχει απώλεια συνείδησης, κάτι που επιτυγχάνεται μέσω αναισθησιολογικών φαρμάκων, τα οποία προκαλούν χημικό κώμα [273].

Ενδέχεται να χορηγηθούν αντιπηκτικά φάρμακα με σκοπό την αραίωση του ορού του αίματος ώστε να αποτραπεί η δημιουργία θρόμβων, όπως και αναστολείς του ενζύμου μετατροπής της αγγειοτενσίνης (ACE), αναστολείς των υποδοχέων της αγγειοτενσίνης (ARBs) ή στατίνες.

Η Remdesivir (Veklury) είναι ένα αντιικό φάρμακο που εγκρίθηκε από τον Οργανισμό Τροφίμων και Φαρμάκων (FDA) για τη θεραπεία ασθενών με COVID-19 άνω των 12 ετών οι οποίοι νοσηλεύονται καθώς σε πολλές περιπτώσεις συμβάλει σε ταχύτερη ανάρρωση. Το Remdesivir παρασκευάστηκε για την καταπολέμηση του Έμπολα, αλλά ο FDA εξέδωσε απόφαση έκτακτης ανάγκης, ώστε οι γιατροί να το χρησιμοποιήσουν κατά της COVID-19.

Επίσης άδεια χρήσης έκτακτης ανάγκης, από τον FDA έχει δοθεί και στη χρήση μονοκλωνικών αντισωμάτων όπως το Sotrovimab ή ο συνδυασμός Casirivimab και Imddevimab (REGEN-COV) που μπορούν να χορηγηθούν σε ασθενείς υψηλού κινδύνου που έχουν πρόσφατα διαγνωστεί με ήπια έως μέτρια εκδήλωση της νόσου [271].

Ένας ακόμα τρόπος αντιμετώπισης της νόσου είναι και η χρήση παράγωγων αίματος, δηλαδή πλάσματος από ασθενείς που έχουν αναρρώσει. Από τα μέχρι τώρα δεδομένα φαίνεται ότι το μεγαλύτερο όφελος από τη χορήγηση πλάσματος το είχαν οι ασθενείς, ηλικίας 80 ετών και άνω που έλαβαν πλάσμα με υψηλό τίτλο αντισωμάτων εντός 72 ωρών από τη στιγμή διάγνωσης της λοίμωξης [272].

Έρευνα και ανάπτυξη φαρμάκων και θεραπευτικών τεχνικών

Πολλές από τις τρέχουσες μελέτες για την καταπολέμηση του SARS-CoV-2 επικεντρώνονται στην εύρεση τρόπων για τον αποκλεισμό των αλληλεπιδράσεων μεταξύ του υποδοχέα ACE2 και της ιικής πρωτεΐνης S [274]. Η δυσκολία να γίνει αυτό είναι ότι ο υποδοχέας ACE2 υπάρχει επίσης σε κύτταρα που βρίσκονται και σε άλλα μέρη του σώματος που δεν στοχεύει

ο ιός. Επομένως, η μη ειδική αναστολή του ACE2 θα μπορούσε να οδηγήσει σε ανεπιθύμητες παρενέργειες [275,276] και σε διάφορες επιπλοκές, συμπεριλαμβανομένων των καρδιαγγειακών παθήσεων και της υπέρτασης [277]. Ως αποτέλεσμα, διερευνώνται εναλλακτικές στρατηγικές για τον αποκλεισμό της αλληλεπίδρασης μεταξύ της πρωτεΐνης S SARS-CoV-2 και του ACE2 απευθείας χρησιμοποιώντας θεραπευτικά μονοκλωνικά αντισώματα. Ένα ιδανικό θεραπευτικό αντίσωμα θα δέσμευε ειδικά την S-πρωτεΐνη και θα την εμπόδιζε από το να αλληλεπιδράσει με τον υποδοχέα ACE2, επιτυγχάνοντας παρόμοιο αποτέλεσμα με τον αποκλεισμό του ACE2 απευθείας χωρίς τις αρνητικές παρενέργειες [278]. Εναλλακτικά, ο σχεδιασμός μικρών πτερυϊδών ή μικρών μορίων που συνδέονται με διαφορετικές υπομονάδες της S-πρωτεΐνης SARS-CoV-2 θα αποτρέψει την αλληλεπίδραση των υπομονάδων με το ACE2 και την είσοδο του ιού στο κύτταρο. Όπως αναφέρθηκε προηγουμένως, η S-πρωτεΐνη αρχικά διασπάται σε S1 και S2 υπομονάδες από την πρωτεάση του κυττάρου ξενιστή, τη φουρίνη. Το S1 διευκολύνει τη δέσμευση του υποδοχέα με τη μεμβράνη του κυττάρου ξενιστή, ενώ το S2 παρέχει δομική υποστήριξη και ρυθμίζει τη σύντηξη με τη μεμβράνη των επιθηλιακών κυττάρων του ξενιστή. Για το λόγο αυτό, η παρεμπόδιση της διάσπασης της S-πρωτεΐνης θα μπορούσε επίσης να είναι μια πιθανή στρατηγική για την αναστολή της εισόδου του ιού [278-283].

Χρησιμοποιώντας τη γνώση που αποκτήθηκε από τον HKU1, έναν ιό του κοινού κρυολογήματος, και τον ιό MERS, οι επιστήμονες βρήκαν δύο μικρούς βρόχους αμινοξέων που συγκρατούν τις δύο α-έλικες στην ακίδα. Όταν η S-πρωτεΐνη δεσμεύεται στα ανθρώπινα κύτταρα, η δομή που μοιάζει με πηνίο απελευθερώνεται και οι δύο έλικες και ο ένας βρόχος επιμηκύνονται για να φέρουν το ανθρώπινο κύτταρο και τον ιό κοντά μεταξύ τους για σύντηξη. Προσθέτοντας έως και έξι υπολείμματα προλίνης στην S-πρωτεΐνη SARS-CoV-2, οι επιστήμονες θεωρούν ότι μπορούν να αποτρέψουν την απελευθέρωση του «ελατηρίου», αποτρέποντας τη σύντηξη του SARS-CoV-2 με ανθρώπινα κύτταρα [274].

Μια ακόμα μελέτη διαπίστωσε ότι η δεξαμεθαζόνη, ένα κοινό στεροειδές φάρμακο, μπορεί να βοηθήσει ασθενείς που νοσηλεύονται με σοβαρές επιπλοκές από COVID-19. Τα ευρήματα είναι προκαταρκτικά και οι ερευνητές δεν έχουν δημοσιεύσει ακόμα την πλήρη μελέτη [271].

Κλινικές δοκιμές βρίσκονται σε εξέλιξη για άλλα φάρμακα, συμπεριλαμβανομένης της tocilizumab, η οποία έχει χρησιμοποιηθεί για τη θεραπεία αυτοάνοσων καταστάσεων και της φλεγμονώδους κατάστασης του Συνδρόμου Απελευθέρωσης Κυτοκίνης [271].

Αποθεραπεία

Οι μέχρι τώρα μελέτες έχουν δείξει ότι για πλήρη αποθεραπεία μπορεί να χρειαστούν μέχρι και 2 εβδομάδες μετά από ήπια νόσηση ή μέχρι έως 6 εβδομάδες για σοβαρή νόσηση. Επίσης σε περίπτωση που κάποιο άτομο χρειάστηκε εισαγωγή σε μονάδα εντατικής θεραπείας (ΜΕΘ) για αρκετό χρονικό διάστημα, άνω των 2 εβδομάδων, θα πρέπει να ακολουθήσει ένα πρόγραμμα αποκατάστασης (αναπνευστικές και μυϊκές φυσιοθεραπείες), ώστε να ανακτήσει την μυϊκή του δύναμη και αντοχή. Νεότερα δεδομένα δείχνουν ότι η ανάρρωση εξαρτάται από πολλούς παράγοντες όπως η ηλικία και τα τυχόν υποκείμενα νοσήματα [284].

Επιπτώσεις Νόσησης (Post-COVID-19)

Έρευνες για την εικόνα και την πορεία της υγείας σε ασθενείς που νόσησαν από COVID-19 έχει καταγράψει την αύξηση του κινδύνου εμφάνισης νέων παθήσεων μετά από σοβαρή νόσηση από COVID-19, πράγμα το οποίο παρατηρείται σε μικρότερη συχνότητα σε άλλες ιογενείς ασθενείες. Πέραν των ατόμων μεγαλύτερης ηλικίας που είχαν προϋπάρχουσες

καταστάσεις και νεότεροι ενήλικες, ηλικίας κάτω των 50 ετών, είχαν επίσης αυξημένο κίνδυνο ανάπτυξης νέων παθήσεων [285].

Πιο συγκεκριμένα το 14% των ατόμων ηλικίας 18 έως 65 ετών, που είχαν μολυνθεί από τον SARS-CoV-2 και έφτασαν στην οξεία φάση της νόσου, εμφάνισαν τουλάχιστον έναν νέο τύπο χρόνιας πάθησης μετά από την αποθεραπεία, ποσοστό που ήταν 5% υψηλότερο σε σχέση με αυτό των ασθενών που νόσησαν από άλλου τύπου ιογενή νόσο του κατώτερου αναπνευστικού συστήματος. Οι παθήσεις που μπορεί να παρουσιαστούν μετά από σοβαρή νόσηση από COVID-19 είναι το σύνδρομο χρόνιας αναπνευστικής ανεπάρκειας (ΧΑΠ), οι καρδιακές αρρυθμίες, η μυοκαρδίτιδα, η υπερπηκτικότητα του αίματος, η εγκεφαλοπάθεια, οι δυσκολίες στην μνήμη, η νεφροπάθεια, ο διαβήτης, η ηπατική και η κόπωση [285].

1.13 Επιδημιολογικά Δεδομένα

Ο πρώτος ασθενής με COVID-19 εντοπίστηκε την 1η Δεκεμβρίου 2019, αλλά χωρίς να έχει εμφανείς συνδέσεις με την Αγορά Χονδρικού Εμπορίου Θαλασσινών της Χουνάν [286,287], πράγμα που έρχεται σε αντίθεση με το γεγονός πως από τα αρχικά κρούσματα που εμφανίστηκαν το Δεκέμβριο, τα δύο τρίτα βρέθηκαν να έχουν σχέση με την αγορά της Χουνάν [288-290]. Η ανάλυση μέσω της τεχνικής του μοριακού ρολογιού έδειξε πως το πιο πιθανό χρονικό διάστημα έναρξης της πανδημίας θεωρείται πως ξεκινάει από τα μέσα του Οκτωβρίου και διαρκεί έως τα μέσα του Νοεμβρίου 2019 [291,292].

Κρούσματα

Έως και τις 3 Δεκεμβρίου του 2021 έχουν καταγραφεί περισσότερα από 263.000.000 επιβεβαιωμένα κρούσματα [293]. Ως επίσημες μετρήσεις κρουσμάτων θεωρούμε τον αριθμό των ατόμων που έχουν υποβληθεί σε εξετάσεις για COVID-19 και των οποίων το τεστ έχει επιβεβαιωθεί θετικό σύμφωνα με τα επίσημα πρωτόκολλα (RT-PCR tests, rapid test) [294,295]. Αρχικά πολλές χώρες, είχαν επίσημες πολιτικές να μην γίνονται διαγνωστικά τεστ σε πολίτες με ήπια μόνο συμπτώματα [296,297]. Μια ανάλυση της πρώιμης φάσης της επιδημίας, έως τις 23 Ιανουαρίου, υπολόγισε ότι το 86% των λοιμώξεων COVID-19 δεν είχαν εντοπιστεί και ότι αυτές οι μη καταγεγραμμένες λοιμώξεις ήταν η πηγή για το 79% των τεκμηριωμένων περιπτώσεων [298]. Αρκετές άλλες μελέτες, είχαν υπολογίσει πως ο αριθμός των λοιμώξεων σε πολλές χώρες είναι πιθανό να είναι σημαντικά μεγαλύτερος από τις επίσημα καταγεγραμμένες περιπτώσεις [299,300]. Στις 9 Απριλίου 2020, προκαταρκτικοί έλεγχοι διαπίστωσαν ότι από τους πολίτες που εξετάστηκαν στο Gangelt της Γερμανίας, επίκεντρο σημαντικών εξάρσεων κρουσμάτων COVID-19, το 15% βρέθηκαν θετικοί για αντισώματα έναντι του ιού [301]. Επιπλέον διαγνωστικοί έλεγχοι για COVID-19 σε έγκυες γυναίκες στην πόλη της Νέας Υόρκης και σε αιμοδότες στην Ολλανδία, εμφάνισαν επίσης υψηλά ποσοστά θετικών τεστ αντισωμάτων υποδηλώνοντας πως οι πραγματικές μολύνσεις ήταν αρκετά περισσότερες από τις επίσημες [302,303].

Μια ανάλυση στις αρχές του 2020 των περιπτώσεων ανά ηλικία στην Κίνα έδειξε ότι ένα σχετικά χαμηλό ποσοστό περιπτώσεων εμφανίστηκε σε άτομα κάτω των 20 ετών [304], χωρίς να αποσαφηνίζει, εάν αυτό οφειλόταν στο ότι οι νέοι είχαν λιγότερες πιθανότητες να μολυνθούν ή αν ήταν λιγότερο πιθανό να αναπτύξουν σοβαρά συμπτώματα ώστε να αναζητήσουν ιατρική βοήθεια και να υποβληθούν σε εξετάσεις [305]. Μια αναδρομική μελέτη κορτής στην Κίνα διαπίστωσε ότι τα παιδιά και οι ενήλικες είχαν τις ίδιες πιθανότητες να μολυνθούν [306].

Οι αρχικές εκτιμήσεις του βασικού αριθμού αναπαραγωγής (R_0) για τον COVID-19 τον Ιανουάριο του 2020 ήταν μεταξύ 1,4 και 2,5 [307], αλλά μια μεταγενέστερη ανάλυση κατέληξε στο συμπέρασμα ότι μπορεί να είναι περίπου από 3,8 έως 8,9, με μέσο όρο 5,7 [308]. Το R_0 μπορεί να ποικίλλει μεταξύ των πληθυσμών και δεν πρέπει να συγχέεται με τον αριθμό αποτελεσματικής αναπαραγωγής (R), ο οποίος λαμβάνει υπόψη επιπτώσεις όπως η κοινωνική απόσταση και η ανοσία της αγέλης. Μέχρι τα μέσα Μαΐου 2020, το πραγματικό R ήταν κοντά ή κάτω από το 1,0 σε πολλές χώρες, πράγμα που σημαίνει ότι η εξάπλωση της νόσου σε αυτές τις περιοχές εκείνη την εποχή ήταν σταθερή ή μειούμενη.

Πρόσφατη μελέτη του Yu, Cheng-Jun MD και των συνεργατών του που κάλυψε 5 χώρες από την Ασία, 5 χώρες από την Ευρώπη, 12 χώρες από την Αφρική και 1 χώρα από καθεμία από τις ηπείρους της Αυστραλίας της Βόρειας Αμερικής και τη Νότιας Αμερικής εμφάνισε πως το συνολικό παγκόσμιο R_0 κυμαίνεται μεταξύ 3,09–5,39 με μέσο το 4,08, με τις μελλοντικές εκτιμήσεις του R_0 για νέα και μεταβαλλόμενα επίκεντρα να αναμένεται να είναι στα ίδια ή και υψηλότερα επίπεδα από εκείνα του αρχικού επίκεντρου της Γουχάν στην Κίνα [309].

Θάνατοι

Μέχρι και τις 3 Δεκεμβρίου του 2021 έχουν καταγραφεί περισσότεροι από 5.200.000 επιβεβαιωμένοι θάνατοι από COVID-19 [293]. Ο πρώτος επιβεβαιωμένος θάνατος ήταν στη Γουχάν στις 9 Ιανουαρίου 2020 [310], με τον πρώτο επιβεβαιωμένο θάνατο εκτός Κίνας να σημειώνεται την 1η Φεβρουαρίου 2020 στις Φιλιππίνες [311] και τον πρώτο αναφερόμενο θάνατο εκτός Ασίας (ΗΠΑ), να σημειώνεται στις 6 Φεβρουαρίου 2020 [312]. Οι επίσημοι θάνατοι από COVID-19 αναφέρονται σε άτομα που πέθαναν, αφού βρέθηκαν θετικά με COVID-19, σύμφωνα με τα πρωτόκολλα. Σε αυτές τις μετρήσεις ενδέχεται να μην έχουν υπολογιστεί οι θάνατοι εκείνων που δεν είχαν υποβληθεί σε εξετάσεις [313]. Αντίθετα, οι θάνατοι ατόμων που είχαν υποκείμενες παθήσεις μπορεί να οδηγήσουν σε υπερβολική καταμέτρηση [314]. Οι συγκρίσεις των στατιστικών για τους θανάτους για όλες τις αιτίες έναντι του εποχιακού μέσου όρου δείχνουν υπερβολική θνησιμότητα σε πολλές χώρες [315]. Αυτό μπορεί να περιλαμβάνει θανάτους λόγω των τεταμένων συστημάτων υγειονομικής περίθαλψης και τις αναστολές των προγραμματισμένων τακτικών χειρουργείων [316].

Θνητότητα

Περισσότερο από το 95 τοις εκατό των ανθρώπων που προσβάλλονται από τον COVID-19 αναρρώνουν. Ο χρόνος μεταξύ της έναρξης των συμπτωμάτων και του θανάτου κυμαίνεται συνήθως από 6 έως 41 ημέρες, με μέσο όρο τις 14 ημέρες [317]. Τα άτομα που διατρέχουν τον μεγαλύτερο κίνδυνο από τον COVID-19 τείνουν να είναι αυτά με υποκείμενες παθήσεις ή άτομα 65 ετών και άνω [212]. Οι σοβαρότεροι παράγοντες κινδύνου για την εκδήλωση σοβαρής νόσησης COVID-19 να είναι η παχυσαρκία και οι επιπλοκές του διαβήτη και οι αγχώδεις διαταραχές.

Το κλάσμα του αριθμού των θανάτων προς τον αριθμό των διαγνωσμένων κρουσμάτων σε ένα δεδομένο χρονικό αποτελεί την Θνητότητα (Mortality) του COVID-19.

Τον Δεκέμβριο του 2020, μια συστηματική ανασκόπηση και μετα-ανάλυση που δημοσιεύτηκε στο European Journal of Epidemiology υπολόγισε ότι η Θνητότητα ήταν 0,5% έως 1% σε ορισμένες χώρες (Γαλλία, Ολλανδία, Νέα Ζηλανδία και Πορτογαλία), 1% έως 2% σε πολλές άλλες χώρες (Αυστραλία, Αγγλία, Λιθουανία και Ισπανία) και περίπου 2,5% στην Ιταλία. Αυτές οι εκτιμήσεις περιελάμβαναν θανάτους σε εγκαταστάσεις φροντίδας ηλικιωμένων [318]. Η ίδια η μελέτη διαπίστωσε επίσης ότι οι περισσότερες από τις διαφορές στη Θνητότητα μεταξύ τοποθεσιών αντικατοπτρίζουν αντίστοιχες διαφορές στην ηλικιακή

σύνθεση του πληθυσμού της περιοχής, καθώς τα παιδιά και οι νεότεροι ενήλικες εμφανίζουν πολύ χαμηλή Θνητότητα (0,002% στην ηλικία 10 ετών και στο 0,01% στην ηλικία των 25 ετών) με τα ποσοστά της να αυξάνονται σταδιακά στους ενήλικες μεγαλύτερης ηλικίας (0,4% στην ηλικία 55 ετών, 1,4% στην ηλικία 65 ετών, 4,6% στην ηλικία 75 ετών και 15% στην ηλικία των 85 ετών) [318]. Αυτά τα αποτελέσματα επισημάνθηκαν επίσης σε μια έκθεση του Δεκεμβρίου 2020 που εκδόθηκε από τον Παγκόσμιο Οργανισμό Υγείας. Ανάλυση αυτών των δεδομένων δείχνει πως ο COVID-19 είναι επικίνδυνος όχι μόνο για τους ηλικιωμένους αλλά και για τους μεσήλικες ενήλικες, για τους οποίους η θανατηφόρα μόλυνση από COVID-19 είναι δύο τάξεις μεγέθους πιο πιθανή από τον ετήσιο κίνδυνο θανατηφόρου τροχαίου ατυχήματος και πολύ πιο επικίνδυνη από την εποχική γρίπη [318].

Με βάση τα στατιστικά του ΠΟΥ, η Θνητότητα σε παγκόσμια κλίμακα είναι 1,98% (5.232.562 θάνατοι για 263.563.622 περιπτώσεις) έως και τις 3 Δεκεμβρίου 2021 [293] με την Θνητότητα στην Ελλάδα να είναι 1,92%. (18.325 θάνατοι για 951.351 περιπτώσεις) [293]. Η Θνητότητα ποικίλλει ανάλογα την περιοχή, τον όγκο των διαγνωστικών τεστ, την ποιότητα του συστήματος υγειονομικής περίθαλψης, τις επιλογές θεραπείας, την ταχύτητα ανταπόκρισης της κυβέρνησης, τον χρόνο από την αρχική έκρηξη κρουσμάτων και τα χαρακτηριστικά του πληθυσμού (ηλικία, φύλο) [317]. Οι επίσημοι αριθμοί θανάτων έχουν επικριθεί για την υποαναφορά του πραγματικού αριθμού θανάτων, επειδή οι συγκρίσεις των ποσοστών θανάτων πριν και κατά τη διάρκεια της πανδημίας δείχνουν αύξηση των θανάτων που δεν εξηγείται μόνο από τους θανάτους από τον COVID-19 [319]. Χρησιμοποιώντας τέτοια δεδομένα, οι εκτιμήσεις του πραγματικού αριθμού θανάτων από COVID-19 παγκοσμίως έχουν συμπεριλάβει ένα εύρος από 9,5 έως 18,6 εκατομμύρια από το The Economist, καθώς και πάνω από 10,3 εκατομμύρια από το Ινστιτούτο Μετρήσεων και Αξιολόγησης Υγείας (IHME) [319]. Στον **Πίνακα 1** (βλ. **Παράρτημα Α**) απεικονίζονται τα κρούσματα και οι θάνατοι ανά χώρα έως τις 3 Δεκεμβρίου 2021 [321].

Εμβολιασμοί

Έως και σήμερα, δέκα εμβόλια έχουν λάβει έγκριση από τουλάχιστον μία εθνική ρυθμιστική αρχή για χρήση στο ευρύ κοινό. Οι εμβολιασμοί ξεκίνησαν στα τέλη 2020 και έως και τις 2 Δεκεμβρίου 2021, έχουν χορηγηθεί περισσότερες από 7,8 δισεκατομμύρια δόσεις (7.864.123.038) [293], ενώ ο ρυθμός εμβολιασμού είναι πάνω από τα 27,5 εκατομμύρια εμβολιασμούς ημερησίως [320].

Τεχνητή Νοημοσύνη

Η Νοημοσύνη ορίζεται ως το σύνολο των γνωστικών ικανοτήτων ενός ανθρώπου, όπως αντίληψη, μνήμη, φαντασία, συνειρμός, διανόηση και ειδικότερα η ικανότητα προσαρμογής και αντιμετώπισης νέων καταστάσεων αξιοποιώντας προηγούμενες εμπειρίες, διακρίνοντας ομοιότητες και διαφορές με αυτές [322].

Η Τεχνητή Νοημοσύνη-TN (AI) διατυπώθηκε ως όρος για πρώτη φορά στο πανεπιστήμιο του Ντάρτμουθ το 1956 από τον John McCarthy, σε συνέδριο ερευνητών, από τους τομείς των Μαθηματικών, της Ηλεκτρονικής και της Ψυχολογίας για τη μελέτη δυνατοτήτων χρήσης των υπολογιστών με στόχο την προσομοίωση της ανθρώπινης νοημοσύνης [323,324], η TN ως έννοια είχε ήδη κάνει την εμφάνισή της το 1950 όπου ο φημισμένος Άγγλος μαθηματικός, Άλαν Τούρινγκ επινόησε μια δοκιμασία που έμεινε στην ιστορία της Πληροφορικής ως η δοκιμασία Τούρινγκ [325]. Στο test παίρνουν μέρος ένας εξεταστής ένα φυσικό πρόσωπο και μια υπολογιστική μηχανή και ο καθένας έχει το δικό του χώρο, απομονωμένο από τους άλλους. Ο εξεταστής απευθύνει ερωτήσεις και στους δύο και δεν δύναται να αντιληφθεί ποιος από τους άλλους δυο απαντά κάθε φορά. Ο εξεταστής θα πρέπει να συμπεράνει σωστά αν αυτός που απάντησε ήταν το φυσικό πρόσωπο ή η υπολογιστική μηχανή. Ο Τούρινγκ προέβλεψε ότι το 2000 θα έχει αναπτυχθεί τεχνητή νοημοσύνη που θα μπορεί να ξεγελάσει το 30% των εξεταστών έπειτα από 5 λεπτά συζήτησης. Το 2014, ένα πρόγραμμα υπολογιστή, το Eugene Goostman, πέρασε για πρώτη φορά το πλήρες τεστ Τούρινγκ, αφού κατάφερε να ξεγελάσει το 33% των εξεταστών [326].

Ο όρος Τεχνητή Νοημοσύνη σήμερα, αναφέρεται στον τομέα της Πληροφορικής που ασχολείται με το σχεδιασμό και την υλοποίηση υπολογιστικών συστημάτων τα οποία προσπαθούν να μιμηθούν χαρακτηριστικά της ανθρώπινης συμπεριφοράς που προϋποθέτουν κάποια υποτυπώδη ευφυΐα όπως η μάθηση, η προσαρμοστικότητα, η εξαγωγή συμπερασμάτων, η επίλυση προβλημάτων κ.α. Η TN αποτελεί σημείο τομής των επιστημονικών πεδίων, της Πληροφορικής, της Μηχανικής, της Ψυχολογίας, της Νευρολογίας, της Γλωσσολογίας και της Φιλοσοφίας με σκοπό τη σύνθεση μιας ευφυούς συμπεριφοράς, που να εμπεριέχει στοιχεία συλλογιστικής με τη δυνατότητα μάθησης και προσαρμογής σε μεταβαλλόμενο περιβάλλον.

Η TN χωρίζεται στη Συμβολική και την Υποσυμβολική TN. Στην Συμβολική TN έχουμε την προσπάθεια εξομοίωσης της ανθρώπινης νοημοσύνης με τη χρήση αλγορίθμιμικών συμβόλων και υψηλού επιπέδου λογικούς κανόνες. Η Υποσυμβολική TN έχει ως στόχο την αναπαράσταση της ανθρώπινης ευφυΐας, χρησιμοποιώντας στοιχειώδη αριθμητικά μοντέλα τα οποία συνθέτονται από επαγωγικά νοήμονες συμπεριφορές οι οποίες προκύπτουν από την διαδοχική αυτοοργάνωση απλούστερων δομικών συστατικών (Συμπεριφορική TN), προσομοιώνοντας βιολογικές διαδικασίες όπως η εξέλιξη των ειδών και η εγκεφαλική λειτουργία (Υπολογιστική Νοημοσύνη), ή εφαρμόζοντας στατιστικές μεθοδολογίες σε προβλήματα TN [327].

2.1 Ιστορική Αναδρομή

Ημερομηνίες σταθμοί της εξέλιξης της Τεχνητής Νοημοσύνης:

- 1943: Οι McCulloch και Pitts προτείνουν ότι μοντέλο τεχνητών νευρώνων που έχει τη δυνατότητα να μαθαίνει και να υπολογίζει κάθε υπολογίσιμη συνάρτηση [328].
- 1950: Ο Alan Turing, που θεωρείται ο πατέρας της TN, επινόησε «το παιχνίδι της μίμησης» [325] για την ανάδειξη ευφυών μηχανών.
- 1951: Οι Minsky και Edmonds υλοποιούν το πρώτο νευρωνικό δίκτυο, το SNARC (*Stochastic Neural Analog Reinforcement Calculator*), το οποίο έχει 40 νευρώνες και χρησιμοποιεί 3000 λυχνίες και γράφονται τα πρώτα προγράμματα για τον υπολογιστή Ferranti Mark I στο Πανεπιστήμιο του Μάντσεστερ: ένα πρόγραμμα που παίζει ντάμα από τον Κρίστοφερ Στράκλι και ένα που παίζει σκάκι από τον Ντίτριχ Πρίνζ.

1956-70: Πρώτη Φάση ανάπτυξης της Τεχνητής Νοημοσύνης

- 1956: Συνάντηση στο Dartmouth College ερευνητών από το χώρο των Μαθηματικών, της Ηλεκτρονικής και Ψυχολογίας (McCarthy, Allen Newell, Herbert Simon, Marvin Minsky) με κοινό στόχο τη μελέτη δυνατοτήτων χρήσης των υπολογιστών με σκοπό την προσομοίωση της ανθρώπινης νοημοσύνης [323].
- 1958: Ο Friedberg πρότεινε μια τεχνική, τη μηχανική εξέλιξη (machine evolution) ή όπως ονομάζεται τώρα, γενετικοί αλγόριθμοι (genetic algorithms).
- 1958: Δημιουργία της γλώσσας Lisp από τον McCarthy. Πρότεινε ένα υποθετικό σύστημα (advice taker), που χρησιμοποιούσε γνώση (όπως το LT) αλλά αφορούσε γενικά, καθημερινά, προβλήματα.
- 1960-1969: Στο Stanford υλοποιήθηκε το πρώτο robot, το Shakey robot.
- 1962: Βελτιώσεις της μεθόδου μάθησης των νευρωνικών δικτύων του Hebb από τον Rosenblatt με τα perceptrons.
- 1965: Ο Έντουαρντ Φάιγκενμπαουμ ξεκινά το Dendral, μια δεκαετή προσπάθεια ανάπτυξης λογισμικού που θα συμπεράνει τη μοριακή δομή οργανικών ενώσεων χρησιμοποιώντας ενδείξεις επιστημονικών οργάνων. Ήταν το πρώτο έμπειρο σύστημα (expert system).
- 1965: Πρώτος ο Zadeh εισάγει τους όρους «Ασαφή Σύνολα» (Fuzzy Sets) και «Ασαφείς Αλγόριθμοι» (Fuzzy Algorithms) [329,330].
- 1966: Ιδρύεται το Εργαστήριο Μηχανικής Νοημοσύνης στο Εδιμβούργο – το πρώτο από μια σημαντική σειρά εγκαταστάσεων που οργανώνονται από τον Ντόναλντ Μίτσι και άλλους, επίσης ο Weizenbaum ύστερα από έρευνα γύρω από την κατανόηση γλώσσας και την αντίληψη μηχανής, δημιουργεί το ELIZA το οποίο μπορούσε να κάνει συζήτηση για οποιοδήποτε θέμα χρησιμοποιώντας και παραφράζοντας τις προτάσεις που έδινε σαν ερώτηση ο χρήστης .
- 1970: Αναπτύσσεται το Planner και χρησιμοποιείται στο SHRDLU, μια εντυπωσιακή επίδειξη αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστή.
- 1971: Ξεκινά η εργασία πάνω στο σύστημα αυτόματης απόδειξης θεωρημάτων Boyer-Moore στο Εδιμβούργο.

1970-80: Ωρίμαση της συμβολικής και υπολογιστικής Τεχνητής Νοημοσύνης.

Το χαρακτηριστικό των συστημάτων της εποχής ήταν ότι περιείχαν ελάχιστη ή καθόλου γνώση για το πεδίο του προβλήματος (weak methods). Το εύρος εφαρμογών των νευρωνικών δικτύων ήταν μικρό.

- 1972: a. Οι Colmerauer και Roussel από το Πανεπιστήμιο της Μασσαλίας σε συνεργασία με τον R. Kowalski από το Πανεπιστήμιο του Εδιμβούργου καταλήγουν στη δημιουργία της γλώσσας λογικού προγραμματισμού PROLOG και O Winograd εμβαθύνει στην κατανόηση φυσικής γλώσσας [331].
- 1973: Ρομπότ συναρμολόγησης «Φρέντι» στο Εδιμβούργο: ένα ευπροσάρμοστο σύστημα συναρμολόγησης που ελέγχεται από υπολογιστές.
- 1973: Μελέτη του Rechenberg για τη βελτιστοποίηση των τεχνικών συστημάτων και τις αρχές της βιολογικής εξέλιξης [332].
- 1974: Ο Τέντ Σόρτλιφ γράφει τη διατριβή του για το πρόγραμμα MYCIN (Στάνφορντ), το οποίο κατέδειξε μια πολύ πρακτική προσέγγιση στην ιατρική διάγνωση που βασίζεται σε κανόνες, ενώ λειτουργεί ακόμα και με παρουσία αβεβαιότητας. Άν και δανείστηκε στοιχεία από το DENDRAL, οι δικές του συνεισφορές επηρέασαν έντονα το μέλλον των έμπειρων συστημάτων, ένα μέλλον με πολλαπλές εμπορικές εφαρμογές.
- 1975: Μελέτη του Holland για την προσαρμοστικότητα στα φυσικά και τεχνητά συστήματα [333].
- 1975 & 1977: Ο M. Minsky δημοσιεύει κεφάλαια περί αναπαράστασης της γνώσης σε βιβλία [334,335].
- 1976: Οι Newell & Simon υποστηρίζουν την υπόθεση ότι ένα φυσικό συμβολικό σύστημα διαθέτει τα απαραίτητα χαρακτηριστικά για νοήμονες ενέργειες [336].

1980-90: Μελέτη για Αναγέννηση των Τεχνητών Νευρωνικών Δικτύων

- 1983: Ο Sugeno διατυπώνει την «Ασαφή Θεωρία» [337].
- 1986: Οι Rumelhart and McClelland περιγράφουν τη δημιουργία προσομοιώσεων της αντίληψης στον υπολογιστή [338].
- 1987: 1ο Διεθνές Συνέδριο για τα Νευρωνικά Δίκτυα του IEEE (Institute of Electrical and Electronics Engineers).
- 1991: Η εφαρμογή σχεδίασης ενεργειών DART χρησιμοποιείται αποτελεσματικά στον Α' Πόλεμο του Κόλπου και ανταμείβει 30 χρόνια έρευνας στην TN του Αμερικανικού Στρατού.
- 1992: 1ο Συνέδριο του IEEE για τα Ασαφή Σύνολα.
- 1992: Μελέτη του Koza, για το Γενετικό Προγραμματισμό (Genetic Programming) [339].
- 1994: Ντίκμανς και Ντάιμλερ-Μπενζ οδηγούν περισσότερο από 1000 χλμ. σε μια εθνική οδό του Παρισιού υπό συνθήκες βαρείας κυκλοφορίας και σε ταχύτητες ως και 130 χλμ/ώρα, επιδεικνύοντας αυτόνομη οδήγηση σε ελεύθερες παρόδους, οδήγηση σε συνοδεία, αλλαγή παρόδων και αυτόματη προσπέραση άλλων οχημάτων.
- 1995: Μελέτη του Fogel για τον Εξελικτικό Υπολογισμό (Evolutionary Computation) [340].

- 1997: Ο υπολογιστής Deep Blue της IBM κερδίζει τον παγκόσμιο πρωταθλητή σκακιού Γκάρι Κασπάροφ [341].
- 1998: Κυκλοφορεί ο Φέρμπι της Tiger Electronics και γίνεται η πρώτη επιτυχημένη εμφάνιση TN σε οικιακό περιβάλλον.
- 1999: Η Sony λανσάρει το AIBO, που είναι ένα από τα πρώτα αυτόνομα κατοικίδια TN.
- 2000: Το ρομπότ Nomad εξερευνάει απομακρυσμένες περιοχές στην Ανταρκτική αναζητώντας δείγματα μετεωρίτων.
- 2004: Η DARPA ξεκινά το πρόγραμμα DARPA Grand Challenge («Μεγάλη Πρόκληση DARPA»), που προκαλεί τους συμμετέχοντες να δημιουργήσουν αυτόνομα οχήματα για ένα χρηματικό βραβείο.
- 2005: Γεννιέται το Blue Brain, ένα project που προσομοιάζει τον εγκέφαλο σε μοριακό επίπεδο.
- 2009: Η Google δημιουργεί το πρώτο αυτο-οδηγούμενο αυτοκίνητο.

2.2 Έρευνα και Εξέλιξη

Μολονότι έχουν περάσει περισσότερα από πενήντα χρόνια αδιάλειπτης έρευνας από την ίδρυση του κλάδου της TN, ο τομέας εξακολουθεί να προσφέρει ερευνητικές προκλήσεις, με βασικό στόχο τη δημιουργία όλο και πιο ευφύων μηχανών. Η εξάπλωση του internet και η αύξηση του όγκου της πληροφορίας, έχει ως αποτέλεσμα την ανάπτυξη πληθώρας συστημάτων βασισμένων σε τεχνικές της TN, όπως τα υβριδικά συστήματα μηχανικής μάθησης τα οποία συνδυάζουν νευρωνικά δίκτυα και γενετικούς αλγορίθμους. Ορισμένα τέτοια συστήματα είναι οι μηχανές ευφυούς αναζήτησης (intelligent search machines), οι νοήμονες πράκτορες (intelligent agents), τα συμβουλευτικά συστήματα (recommender systems), τα συστήματα περιρρέουσας νοημοσύνης (ambient intelligence systems), τα συστήματα εξόρυξης δεδομένων (data mining systems) κ.α.

Η Τεχνητή Νοημοσύνη αποτελεί πλέον αναπόσπαστο κομμάτι της καθημερινότητας μας καθώς βρίσκεται παντού διαμορφώνοντας και επηρεάζοντας όλο και μεγαλύτερο μέρος των αποφάσεων μας. Παράδειγμα αποτελούν τα smartphones με το μεγάλο πλήθος εφαρμογών, όπως εφαρμογές αναζήτησης (google, duckduckgo), εφαρμογές κοινωνικών δικτύων (facebook, linkedin), εφαρμογές ηλεκτρονικού ταχυδρομείου (gmail, yahoo), εφαρμογές για την εύρεση της βέλτιστης διαδρομής (google maps), εφαρμογές για την πρόβλεψη του καιρού (meteo) κ.α. Ακόμα ένα αρκετά επίκαιρο παράδειγμα είναι το Netflix όπου ανάλογα τις προτιμήσεις του κάθε χρήστη, προτείνει τις αντίστοιχες ταινίες και σειρές που θα τον ενδιαφέρουν. Παρατηρούμε λοιπόν πως αλγόριθμοι Μηχανικής Μάθησης καταφέρνουν να μιμηθούν την ανθρώπινη νοημοσύνη αναπτύσσοντας έτσι την δική τους.

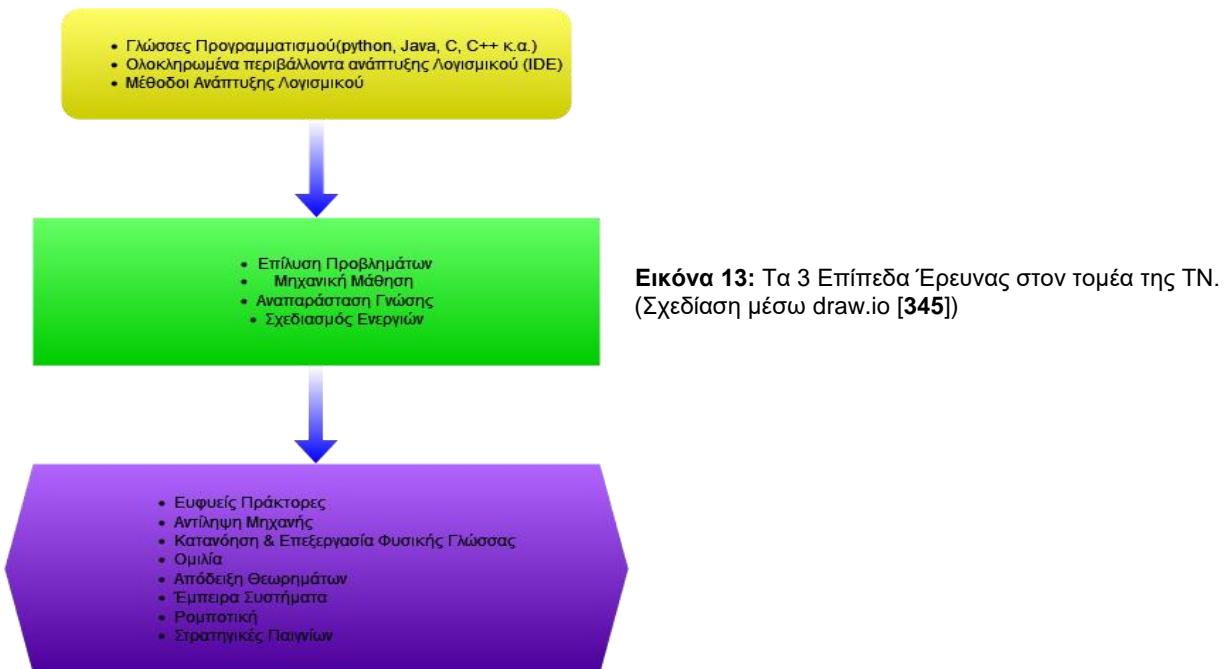
Μια σημαντική συμβολή της TN στην καθημερινότητα μας, αποτελούν οι δυνατότητες που προσφέρει, μέσω της δημιουργίας μοντέλων πρόβλεψης, με πρόσφατο παράδειγμα την επιτυχή πρόβλεψη της αναδίπλωσης των πρωτεϊνών, από τους αλγόριθμους Βαθέων Νευρωνικών Δικτύων (Deep Neural Networks,DNN) AlphaFold και DeepMind [342,343]. Άλλο παράδειγμα στην καθημερινότητα μας, αποτελούν οι αλληλεπιδράσεις με αλγόριθμους TN όπως οι Alexa, Cortana και Siri που βασίζονται στα DNN και συνεχίζουν να βελτιώνουν την απόδοση τους όσο περισσότερο χρησιμοποιούνται.

Επίσης η TN έχει αυτοματοποιήσει την επαναληπτική μάθηση και την ανακάλυψη μέσω δεδομένων. Πιο συγκεκριμένα μέσω της TN μπορούν να εκτελεστούν συχνά μεγάλου όγκου

μηχανογραφημένα έργα, αξιόπιστα και χωρίς κόπο, αντί των αυτοματοποιημένων χειροκίνητων έργων. Επιπλέον παρέχεται η δυνατότητα ανάλυσης περισσότερων και πιο πολύπλοκων δεδομένων μέσω νευρωνικών δικτύων με πολλά κρυφά επίπεδα. Σήμερα με την αύξηση της επεξεργαστικής ισχύος και της χωρητικότητας της μνήμης των υπολογιστών, μπορεί να επιτευχθεί μεγαλύτερη ακρίβεια από μεγαλύτερο όγκο δεδομένων και σε μικρότερο χρόνο, μέσω της μηχανικής μάθησης. Εκτός αυτού μέσω της TN επιτυγχάνεται η μέγιστη δυνατή αξιοποίηση των δεδομένων δημιουργώντας ένα ισχυρό ανταγωνιστικό πλεονέκτημα.

Άλλη μία δυνατότητα που προσφέρουν οι αλγόριθμοι προοδευτικής μάθησης είναι ότι προσαρμόζονται μέσω της συνεχούς ανατροφοδότησης των αποτελεσμάτων που έχουν οι προηγούμενες ενέργειες, όπως οι αλγόριθμοι μπορεί να διδάσκονται μονοί τους σκάκι. Η ίδια λογική χρησιμοποιείται και από τους αλγόριθμους που συστήνουν προϊόντα με βάση τις προηγούμενες επιλογές του κάθε χρήστη και προσαρμόζονται στο αν επιλέχθηκαν ή όχι οι προτάσεις τους [344].

Τέλος η αυτοματοποίηση, οι πλατφόρμες συνομιλίας, τα chatbots και οι έξυπνες μηχανές μπορούν να συνδυαστούν με τις τεράστιες και συνεχώς αυξανόμενες βάσεις δεδομένων προκειμένου να βελτιώσουν πολλές τεχνολογίες στο σπίτι και στο χώρο εργασίας από την ασφάλεια πληροφοριών και προσωπικών στοιχείων έως την ανάλυση επενδύσεων. Τα επίπεδα έρευνας της TN απεικονίζονται συνοπτικά στην (Εικόνα 13).



Όλα όσα έχουν επιτευχθεί μέχρι σήμερα είναι μόνο η αρχή, καθώς υπάρχει τόσο μεγάλο δυναμικό για την ανάπτυξη όλων των κλάδων της TN, όπου γίνεται όλο και πιο δύσκολο να φανταστεί κανείς ένα μέλλον χωρίς αυτή. Στο μέλλον ενδέχεται να υπάρχει και η δυνατότητα της ανάλυσης τεράστιων ποσοτήτων γονιδιωματικών δεδομένων που θα οδηγεί σε ακριβέστερη πρόληψη και θεραπεία των ιατρικών καταστάσεων σε ατομικό επίπεδο. Επίσης αλγόριθμοι μηχανικής μάθησης θα χρησιμοποιούνται για την πρόληψη της διαδικτυακής τρομοκρατίας και απάτης στις ηλεκτρονικές πληρωμές. Επιπροσθέτως εμφανίζεται μια αύξηση στην παραγωγικότητα στον εργασιακό χώρο χάρη στις προόδους της τεχνητής νοημοσύνης.

2.3 Μηχανική Μάθηση

Η Μάθηση είναι ενδογενής ανάγκη του ανθρώπου και χαρακτηριστικό της νοήμονος συμπεριφοράς του. Παρόλα αυτά ερευνητές της Γνωστικής Ψυχολογίας και της φιλοσοφίας δεν έχουν κατανοήσει πλήρως την έννοια της μάθησης. Μέσω της Επαγωγικής Μάθησης (Inductive Learning), ο άνθρωπος εκπαιδεύεται μέσω παρατηρήσεων και δημιουργώντας έτσι μια πιο απλουστευμένη εκδοχή η οποία ονομάζεται νοητικό μοντέλο (mental model). Επίσης ο άνθρωπος έχει τη ικανότητα να οργανώνει και να συσχετίζει τις παρατηρήσεις και τις εμπειρίες του έτσι ώστε να δημιουργήσει νέες δομές που ονομάζονται νοητικά πρότυπα (mental patterns), αξιοποιώντας τον επαγωγικό και του απαγωγικό συλλογισμό. Σε αυτή τη δημιουργία νέων προτύπων από παλαιότερα, στηρίζονται μέθοδοι μάθησης που εξαρτώνται από την προϋπάρχουσα γνώση γύρω από ένα πρόβλημα, όπως η μάθηση μέσω επεξηγήσεων ή μέσω περιπτώσεων.

Η Μηχανική Μάθηση δημιουργήθηκε ως εργαλείο για την ανάλυση δεδομένων και την αναγνώριση προτύπων. Οι προκύπτοντες αλγόριθμοι εκπαιδεύονται (μαθαίνουν) από διαθέσιμα σύνολα δεδομένων που αφορούν ένα συγκεκριμένο αντικείμενο μελέτης (πρόβλημα), και έπειτα αποδίδουν όσο πιο ακριβείς προβλέψεις για νέες περιπτώσεις που έχουν προκύψει. Η Μηχανική Μάθηση, ασχολείται κυρίως με τη μελέτη αλγορίθμων που βελτιώνουν τη απόδοση τους σε μια εργασία που τους έχει ανατεθεί χρησιμοποιώντας την εμπειρία τους [346].

Στον τομέα της Μηχανικής Μάθησης ανήκει και η Εξελικτική Μάθηση (Evolutionary Learning), η οποία μιμείται τις διαδικασίες της φυσικής εξέλιξης ενός είδους. Σε αυτήν ανήκουν οι γενετικοί αλγόριθμοι που χρησιμοποιούνται σε προβλήματα βελτιστοποίησης.

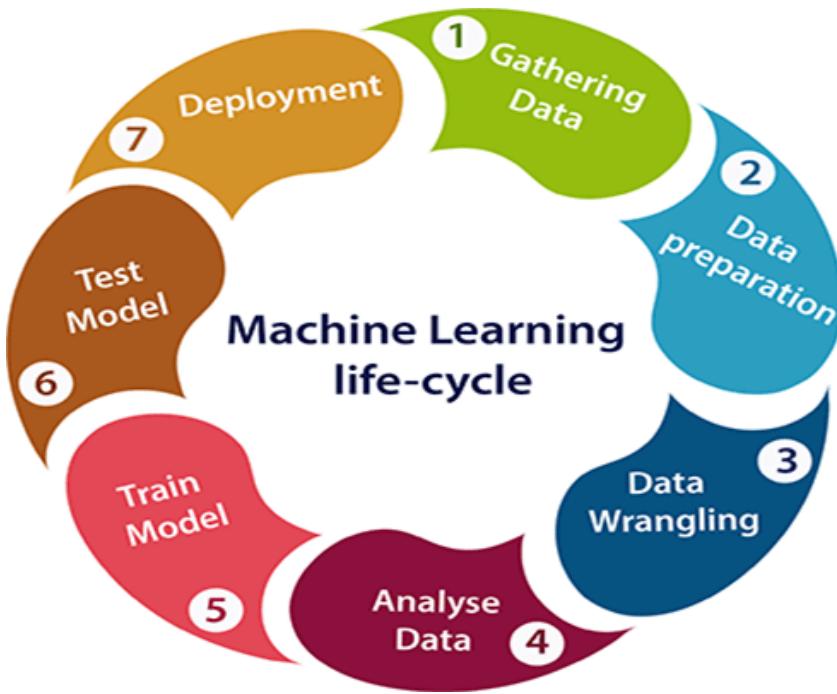
Μέχρι σήμερα έχουν αναπτυχθεί αλγόριθμοι, για συγκεκριμένες περιοχές μάθησης, που έχουν επιτρέψει την εμφάνιση σύγχρονων εμπορικών εφαρμογών με επιτυχία. Από τις εφαρμογές αυτές έχουμε αρχίσει να βλέπουμε ακριβή αποτελέσματα σε αναπάντητα και δυσεπίλυτα προβλήματα σε πολλούς επιστημονικούς τομείς, όπως η Εξόρυξη Δεδομένων, οι Πιθανότητες και η Στατιστική, η Αριθμητική Βελτιστοποίηση, η Νευροβιολογία, η Θεωρία Ελέγχου (Προσαρμοστική), η Ψυχολογία (Εξελικτική, Γνωστική), η Θεωρία της Πολυπλοκότητας κ.α.

Οι αλγόριθμοι Μηχανικής Μάθησης επεξεργάζονται τη γνώση και την αναπαριστούν με πιο μαθηματικοποιημένους τρόπους [346]. Ορισμένοι δέχονται ως είσοδο μόνο παρατηρήσεις ενώ άλλοι λαμβάνουν υπόψιν τους την ήδη υπάρχουσα γνώση.

Ο κύκλος της μηχανικής μάθησης περιλαμβάνει επτά βασικά βήματα (**Εικόνα 14**):

1. Συλλογή δεδομένων
2. Προετοιμασία δεδομένων
3. Προεπεξεργασία δεδομένων
4. Ανάλυση Δεδομένων
5. Εκπαίδευση του Μοντέλου(Αλγορίθμου)
6. Δοκιμασία του Μοντέλου(Αλγορίθμου)
7. Χρήση του Μοντέλου(Αλγορίθμου) σε πραγματικές καταστάσεις

Αρχικά αφού αποσαφηνιστούν οι στόχοι που πρέπει να επιτευχθούν και ολοκληρωθεί η συλλογή δεδομένων εκπαίδευσης σειρά έχει η επιλογή του σωστού αλγορίθμου με βάση τον οποίο θα εκπαιδευτεί το μοντέλο (σύστημα μηχανικής μάθησης) [347].



Εικόνα 14-Φάσεις Μηχανικής Μάθησης:

1. Συλλογή δεδομένων
2. Προετοιμασία δεδομένων
3. Προεπεξεργασία δεδομένων
4. Ανάλυση Δεδομένων
5. Εκπαίδευση του Μοντέλου
6. Δοκιμασία του Μοντέλου
7. Χρήση του Μοντέλου σε πραγματικές καταστάσεις [348]

1. Συλλογή δεδομένων

Η συλλογή δεδομένων είναι το πρώτο βήμα του κύκλου ζωής της μηχανικής μάθησης. Ο στόχος αυτού του βήματος είναι ο εντοπισμός και η λήψη όλων των δεδομένων που σχετίζονται με το συγκεκριμένο πρόβλημα. Σε αυτό το βήμα, πρέπει να προσδιορίσουμε τις διαφορετικές πηγές δεδομένων, καθώς αυτά μπορούν να συλλεχθούν από διάφορες πηγές, όπως αρχεία, βάσεις δεδομένων, διαδίκτυο ή κινητές συσκευές. Είναι ένα από τα πιο σημαντικά βήματα του κύκλου. Η ποσότητα και η ποιότητα των συλλεγόμενων δεδομένων θα καθορίσουν την απόδοση του μοντέλου. Όσο περισσότερα είναι τα δεδομένα, τόσο πιο ακριβής θα είναι η πρόβλεψη.

Αυτό το βήμα περιλαμβάνει τις παρακάτω διαδικασίες [347]:

- Τον Προσδιορισμό των διάφορων πηγών δεδομένων
- Την Συλλογή δεδομένων
- Την Ενσωμάτωση των δεδομένων που λαμβάνονται από διαφορετικές πηγές

Εκτελώντας την παραπάνω εργασία, λαμβάνουμε ένα συνεκτικό σύνολο δεδομένων, που ονομάζεται επίσης σύνολο δεδομένων που θα χρησιμοποιηθεί σε επόμενα βήματα.

2. Προετοιμασία δεδομένων

Η προετοιμασία δεδομένων είναι το βήμα όπου τα δεδομένα, μετά τη συλλογή τους, τοποθετούνται μας σε κατάλληλο μέρος (φάκελος υπολογιστή) ώστε να χρησιμοποιηθούν στην εκπαίδευση του μοντέλου. Σε αυτό το βήμα, πρώτα, συγκεντρώνουμε όλα τα δεδομένα και στη συνέχεια τυχαιοποιούμε τη σειρά τους.

3. Προεπεξεργασία δεδομένων

Είναι η διαδικασία καθαρισμού και μετατροπής ακατέργαστων δεδομένων σε μορφή που μπορεί να χρησιμοποιηθεί από το μοντέλο. Έπειτα γίνεται η επιλογή της μεταβλητής που θα χρησιμοποιηθεί και ύστερα τα δεδομένων μετατρέπονται σε τέτοια μορφή ώστε να είναι πιο κατάλληλα για ανάλυση στο επόμενο βήμα. Είναι ένα από τα πιο σημαντικά βήματα της διαδικασίας. Ο καθαρισμός δεδομένων είναι απαραίτητος για την αντιμετώπιση των ζητημάτων ποιότητας. Δεν είναι απαραίτητο να χρησιμοποιηθούν όλα τα δεδομένα που έχουμε συλλέξει, καθώς ορισμένα από τα δεδομένα μπορεί να μην είναι απαραίτητα. Τα δεδομένα που συλλέγονται ενδέχεται να έχουν διάφορα ζητήματα, όπως το να λείπουν τιμές, να υπάρχουν διπλότυπες ή μη έγκυρες τιμές, ή κάποιες τιμές να επηρεάζουν την ανάπτυξη του μοντέλου ως θόρυβος. Γι' αυτό και χρησιμοποιούμε διάφορες τεχνικές φιλτραρίσματος για να καθαρίσουμε τα δεδομένα. Είναι υποχρεωτικό να εντοπιστούν και να αφαιρεθούν τα παραπάνω ζητήματα γιατί μπορεί να επηρεαστεί αρνητικά η ποιότητα του αποτελέσματος [347].

4. Ανάλυση Δεδομένων

Σε αυτό το βήμα ξεκινά η ανάλυση των δεδομένων και περιλαμβάνει την επιλογή αναλυτικών τεχνικών, την κατασκευή των μοντέλων και την αξιολόγηση των αποτελεσμάτων. Ο στόχος αυτού του βήματος είναι η κατασκευή ενός μοντέλου μηχανικής μάθησης για την ανάλυση των δεδομένων χρησιμοποιώντας διάφορες αναλυτικές τεχνικές και την ανασκόπηση του αποτελέσματος. Ξεκινά με τον προσδιορισμό του τύπου των προβλημάτων, όπου επιλέγουμε τις τεχνικές μηχανικής μάθησης όπως η ταξινόμηση, η παλινδρόμηση, η συσταδοποίηση, συσχέτιση κ.α. Στη συνέχεια κατασκευάζεται το μοντέλο χρησιμοποιώντας τα δεδομένα που έχουν επεξεργαστεί προηγουμένως και αξιολογούμε το μοντέλο. Συγκεκριμένα σε αυτό το βήμα, παίρνουμε τα δεδομένα και χρησιμοποιούμε αλγόριθμους μηχανικής εκμάθησης για να δημιουργήσουμε το μοντέλο.

5. Εκπαίδευση του Μοντέλου

Το επόμενο βήμα είναι η εκπαίδευση του μοντέλου ώστε να βελτιωθεί η απόδοσή του και να επιτευχθεί μεγαλύτερη ακρίβεια στα αποτελέσματα. Χρησιμοποιούμε τα σύνολα των δεδομένων από τα προηγούμενα βήματα για να εκπαίδευσουμε το μοντέλο χρησιμοποιώντας διάφορους αλγόριθμους μηχανικής μάθησης με σκοπό την κατανόηση των διαφόρων μοτίβων, κανόνων και χαρακτηριστικών.

6. Δοκιμασία του Μοντέλου

Μόλις το μοντέλο μηχανικής εκμάθησης έχει εκπαίδευτεί σε ένα δεδομένο σύνολο δεδομένων, τότε δοκιμάζουμε το μοντέλο. Σε αυτό το βήμα, ελέγχουμε την ακρίβεια του μοντέλου μας παρέχοντας ένα δοκιμαστικό σύνολο δεδομένων σε αυτό. Η δοκιμή του μοντέλου καθορίζει την ποσοστιαία ακρίβεια του μοντέλου σύμφωνα με την απαίτηση του προβλήματος

7. Χρήση(Ανάπτυξη) του Μοντέλου σε πραγματικές καταστάσεις

Το τελευταίο βήμα του κύκλου ζωής της μηχανικής μάθησης είναι η Χρήση του Μοντέλου σε πραγματικές καταστάσεις, όπου αναπτύσσουμε το μοντέλο σε πραγματικό σύστημα. Εάν το μοντέλο που προετοιμάστηκε παραπάνω παράγει ένα ακριβές αποτέλεσμα σύμφωνα με τις απαιτήσεις μας με αποδεκτή ταχύτητα, τότε αναπτύσσουμε το μοντέλο στο πραγματικό σύστημα. Άλλα πριν από την ανάπτυξη του έργου, θα πρέπει να ελέγχουμε εάν βελτιώνει

την απόδοσή του χρησιμοποιώντας τα διαθέσιμα δεδομένα ή όχι [347]. Στη φάση αυτή έχουμε και τη σύνταξη της τελικής αναφοράς για το συγκεκριμένο πρόβλημα (έργο).

Είδη Μηχανικής Μάθησης

Οι κυριότερες διαδικασίες για την εκπαίδευση των αλγορίθμων είναι [346]:

1. Επιβλεπόμενη Μάθηση (*Supervised Learning*):

Μέσω αυτής της διαδικασίας μάθησης, ο αλγόριθμος κατασκευάζει μια συνάρτηση που θα απεικονίζει το σύνολο των γνωστών εισόδων και εξόδων (σύνολο εκπαίδευσης), με απώτερο στόχο τη γενίκευση αυτής της συνάρτησης για εισόδους με άγνωστη έξοδο. Η διαδικασία αυτή χρησιμοποιείται συνήθως σε προβλήματα Ταξινόμησης (Classification), Πρόγνωσης (Prediction) ή Διερμηνείας (Interpretation).

2. Μη Επιβλεπόμενη Μάθηση (*Unsupervised Learning*):

Σύμφωνα με τη διαδικασία αυτή ο αλγόριθμος κατασκευάζει ένα μοντέλο για διάφορες εισόδους χωρίς να γνωρίζει τις επιθυμητές εξόδους. Χρησιμοποιείται κυρίως σε προβλήματα, Ανάλυσης Συσχετισμών (Association Analysis) και Ομαδοποίησης (Clustering).

3. Ενισχυτική Μάθηση (*Reinforcement Learning*):

Εδώ ο αλγόριθμος καλείται να αλληλεπιδράσει μέσα σε ένα δυναμικό περιβάλλον όπου πρέπει να επιτευχθεί ένας προκαθορισμένος στόχος. Το πρόγραμμα-δάσκαλος δεν έχει ορίσει τις τεχνικές με τις οποίες ο αλγόριθμος θα φτάσει στον εν λόγω στόχο, αλλά τον επιβραβεύει όσο πλησιάζει προς την επίτευξη του στόχου. Στη συνέχεια ο αλγόριθμος αξιοποιεί αυτές τις πληροφορίες ώστε να επαναπροσαρμόσει την πορεία του για να φτάσει στη βέλτιστη ζητούμενη λύση. Ο τρόπος αυτός μάθησης, χρησιμοποιείται σε προβλήματα Σχεδιασμού (Planning), όπως είναι ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση των εργασιών σε εργοστασιακούς χώρους.

Η Μηχανική Μάθηση, διαθέτει πλήθος μεθόδων επίλυσης, όπως η Παλινδρόμηση (Regression Analysis), η Ταξινόμηση (Classification) και η Συσταδοποίηση (Clustering), με κάθε μία από αυτές να μπορεί να ανταποκριθεί καλύτερα σε συγκεκριμένους τύπους προβλημάτων. Κάθε αλγόριθμος που εκπαιδεύεται, συνοδεύεται και από κάποια ποιοτικά χαρακτηριστικά τα οποία αφορούν την αποτελεσματικότητά του, όπως τα ποσοστά της ακριβείας και του σφάλματος. Η αποτελεσματικότητα εξαρτάται σε μεγάλο βαθμό από την ποιότητα και το μέγεθος των δεδομένων εκπαίδευσης, την επιλογή του κατάλληλου αλγορίθμου εκπαίδευσης καθώς επίσης και από την παραμετροποίηση του.

2.4 Ταξινόμηση (*Classification*)

Μια μέθοδος επιβλεπόμενης μηχανικής μάθησης είναι η Ταξινόμηση η οποία χρησιμοποιείται στην αναγνώριση μοτίβων με στόχο τον προσδιορισμό και την καταχώρηση ενός αντικειμένου σε ένα προκαθορισμένο σύνολο ή υποσύνολο [349, 350].

Αρχικά το μοντέλο (αλγόριθμος) εκπαιδεύεται με ένα σύνολο δεδομένων, με γνωστές εξόδους (κλάσεις) οι οποίες αποτελούν τις κατηγορίες με βάσει τις οποίες θα ταξινομηθούν τα νέα δεδομένα, αφού ολοκληρωθεί το κομμάτι της εκπαίδευσης. Η τελευταία ενέργεια του εκπαίδευσης μοντέλου μπορεί να αφορά ταξινομήσεις δυαδικής ή πολλαπλής φύσεως. Στην περίπτωση της δυαδικής ταξινόμησης, το μοντέλο διαχειρίζεται δεδομένα που

κατατάσσονται με απαντήσεις τύπου ΝΑΙ/ΟΧΙ, ενώ στη περίπτωση της πολλαπλής φύσεως τα δεδομένα μπορούν να ταξινομούνται σε περισσότερες κλάσεις [349,350].

Ένα παράδειγμα χρήσης της ταξινόμησης είναι η υλοποίηση ενός μοντέλου το οποίο θα ταξινομεί φωτογραφίες οι οποίες θα απεικονίζουν κάποια είδη ζώων και φυτών. Αρχικά θα πρέπει να εκπαιδεύσουμε το μοντέλο μέσω ενός συνόλου δεδομένων με φωτογραφίες από τα είδη ζώων και φυτών που επιθυμούμε να διαχειριστούμε, με κάθε φωτογραφία να συνοδεύεται από περαιτέρω δεδομένα σημασμένα με ετικέτες (labels). Οι ετικέτες θα αναφέρουν αρχικά για το αν η φωτογραφία ανήκει σε ζώο ή σε φυτό και το όνομα του είδους που ανήκει. Αν είναι ζώο θα πρέπει να αναφέρουν κάποια από τα χαρακτηριστικά του όπως την ομοταξία του (θηλαστικό, πτηνό, έντομο κ.α.), τη διατροφή του (φυτοφάγο, σαρκοφάγο, παμφάγο). Στην περίπτωση που η φωτογραφία ανήκει σε φυτό οι ετικέτες επίσης θα πρέπει να αναφέρουν χαρακτηριστικά όπως τον τύπο (πόα, θάμνος, δέντρο κ.α.), το οικοσύστημα που αναπτύσσεται (έρημος, τούνδρα, τάιγκα κ.α.).

Εν συνεχεία θα πρέπει να οριστούν οι Κλάσεις σύμφωνα με τις οποίες θα ταξινομηθούν οι είσοδοι (αντικείμενα), προαιρετικά μπορεί να οριστεί και μια επιπλέον η οποία θα περιλαμβάνει τα μη ταξινομημένα αντικείμενα τα οποία δεν θα μπορέσουν να ταξινομηθούν από το εκπαιδευμένο μοντέλο (αλγόριθμο).

Αφού ολοκληρωθεί η εκπαίδευση του μοντέλου, ακολουθεί η δοκιμασία της ακρίβειας του μέσω της τροφοδότησης του με ένα νέο άγνωστο σύνολο εισόδων το οποίο αποτελείται από φωτογραφίες ζώων ή φυτών τις οποίες και θα πρέπει να τις ταξινομήσει.

Η μέθοδος της Ταξινόμησης μπορεί να εφαρμοστεί σε περιπτώσεις δεδομένων, κυρίως πολυπαραγοντικών, τα οποία είναι απαραίτητη η άμεση, γρήγορη και η ακριβής ταξινόμηση τους. Ορισμένα παραδείγματα οπού μπορεί να εφαρμοστεί η Ταξινόμηση είναι:

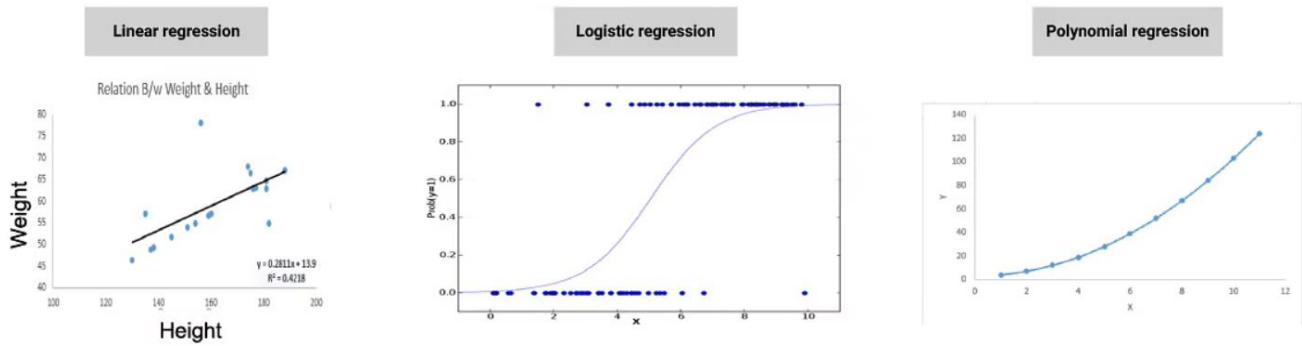
- Για την αναγνώριση ανεπιθύμητων email (spam)
- Για την ικανότητα αποπληρωμής ενός δανείου
- Για την αναγνώριση της αγοραστικής συμπεριφοράς χρήστη
- Για την αναγνώριση του γραφικού χαρακτήρα

2.5 Παλινδρόμηση (Regression)

Η Παλινδρόμηση (Regression) είναι η μέθοδος που εξετάζει τις σχέσεις ενός πλήθους μεταβλητών ενός αντικειμένου με σκοπό την πρόβλεψη μίας εξ' αυτών [350]. Η Παλινδρόμηση ανήκει στην κατηγορία των μεθόδων Επιβλεπόμενης Μηχανικής Μάθησης και έχει ως στόχο την εκπαίδευση ενός μοντέλου, από τα δεδομένα των αντικειμένων, ώστε να οριστεί συνάρτηση μέσω της οποίας θα προβλέπει με ακρίβεια συγκεκριμένες τιμές για άγνωστα σύνολα αντικειμένων. Η Παλινδρόμηση δεν περιορίζεται μόνο στον ορισμό μια ευθείας ή καμπύλης αλλά μπορεί να σχηματίσει και υπερεπίπεδα ανάλογα με το πλήθος των μεταβλητών του κάθε αντικειμένου.

Η Παλινδρόμηση διακρίνεται σε απλή και πολλαπλή. Στην απλή Παλινδρόμηση, χρησιμοποιείται μόνο μία ανεξάρτητη μεταβλητή ως είσοδος ώστε να προβλεφθεί η εξαρτημένη μεταβλητή, ενώ στην πολλαπλή Παλινδρόμηση οι ανεξάρτητες μεταβλητές μπορεί να είναι δύο ή περισσότερες.

Υπάρχουν αρκετοί τύποι Παλινδρόμησης, ανάλογα με την φύση του εκάστοτε προβλήματος. Οι κυριότεροι τύποι Παλινδρόμησης (**Εικόνα 15**) είναι η Γραμμική (Linear) Παλινδρόμηση, η Λογιστική (Logistic) και η Πολυωνυμική (Polynomial) [350].



Εικόνα 15 Οι κυριότεροι τύποι Παλινδρόμησης: Γραμμική (Linear) Παλινδρόμηση, Λογιστική (Logistic) Παλινδρόμηση και Πολυωνυμική (Polynomial) Παλινδρόμηση. [351]

Για να μπορέσει να επιλεγεί ο καλύτερος τύπος Παλινδρόμησης με τον οποίον θα γίνει η εκπαίδευση του μοντέλου (αλγόριθμου), ώστε οι προβλέψεις του να έχουν όσο το δυνατόν μεγαλύτερη ακρίβεια, θα πρέπει αρχικά να αποτυπωθούν σε ένα διάγραμμα διασποράς (scatter plot) όλες τις εξαρτημένες μεταβλητές συναρτήσει όλων των ανεξάρτητων για τις γνωστές εισόδους (training set). Κατόπιν είτε επιλέγεται εμπειρικά ο τύπος Παλινδρόμησης που ταιριάζει στα δεδομένα μας, είτε μετά από δοκιμές διαφορετικών τύπων Παλινδρόμησης επιλέγεται ο καταλληλότερος, ο οποίος μέσω παραμετροποίησης μπορεί να βελτιστοποιήσει την ακρίβεια του υπό εκπαίδευση μοντέλου. Στις περισσότερες περιπτώσεις για τον προσδιορισμό της τελικής εξίσωσης που θα αποδίδει την αποτελεσματικότερη καμπύλη προβλέψεων, θα πρέπει να δημιουργηθούν και να συγκριθούν πολλά μοντέλα, χρησιμοποιώντας πολλές διαφορετικές μαθηματικές μεθόδους, όπως η μέθοδος των ελαχίστων τετραγώνων και το άθροισμα απόλυτων αποκλίσεων.

Η Παλινδρόμηση θεωρείται η κατάλληλη μέθοδος για ακριβείς προβλέψεις τιμών βάσει συσχετισμών με συγκεκριμένες παραμέτρους, βρίσκοντας εφαρμογή στους κλάδους:

- στην Ιατρική, για την εκτίμηση της πορείας ασθενών κ.α.
- της Οικονομίας, για την πρόβλεψη της πορείας μετοχών
- των Κτηματομεσιτικών, για την εκτίμηση της τιμής ακινήτων
- του Εμπόριου, για την πρόβλεψη της αγοραστικής κίνησης

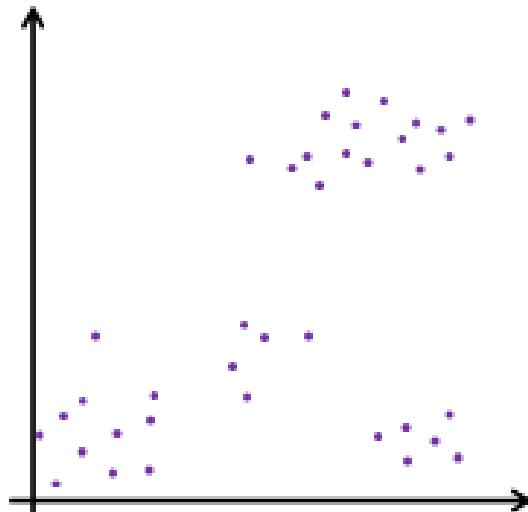
2.6 Συσταδοποίηση (Clustering)

Η Συσταδοποίηση (Clustering) ή Ομαδοποίηση, είναι μια μέθοδος μη επιβλεπόμενης μηχανικής μάθησης, όπου το μοντέλο (αλγόριθμός) αναλαμβάνει να εντοπίσει κοινά χαρακτηριστικά μεταξύ οντοτήτων (data points) και να τις ομαδοποιήσει [350].

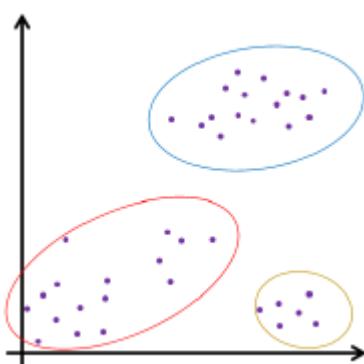
Εδώ ο αλγόριθμος εκπαίδευσης παραμετροποιείται, καθορίζοντας τον αριθμό των συστάδων (clusters), που θα καταταχθούν οι οντότητες των εκπαιδευτικών δεδομένων (training set). Εν συνέχεια ο αλγόριθμος ομαδοποιεί τις οντότητες του training set μέσω συσχετίσεων των δεδομένων που τις απαρτίζουν. Οι συστάδες που θα προκύψουν, θα αποτελέσουν τα σημεία αναφοράς για νέες (άγνωστες) οντότητες, που θα χρειαστεί να αντιστοιχιστούν από το μοντέλο μετά την εκπαίδευσή του. Οι αντιστοιχίσεις των νέων

οντότητων μπορούν να γίνουν με διάφορες προσεγγίσεις με τη βέλτιστη επιλογή να εξαρτάται από το εκάστοτε προς επίλυση πρόβλημα. Ορισμένες από τις πιο γνωστές μεθόδους συσταδοποίησης είναι η προσέγγιση με βάση τα κέντρα βάρους κάθε συστάδας ή με βάση την εγγύτερη γειτνιάζουσα οντότητα (nearest neighbour). Η προσέγγιση με βάση τα κέντρα βάρους κάθε συστάδας, ορίζει ένα κεντρικό σημείο στη συστάδα ως σημείο αναφοράς, ενώ στη προσέγγιση με βάση την εγγύτερη γειτνιάζουσα οντότητα ως σημείο αναφοράς λαμβάνεται σχεδόν οποιαδήποτε οντότητα (data point) βρίσκεται πιο κοντά στην υπό εξέταση οντότητα. Η τελευταία υλοποιείται συνήθως λαμβάνοντας υπόψη ένα πλήθος K κοντύτερων γειτόνων (K nearest neighbors, KNN) [352]. Ακολουθεί παράδειγμα για το πως η διαφορά στην παραμετροποίηση του αλγορίθμου εκπαίδευσης επηρεάζει το μοντέλο που προκύπτει.

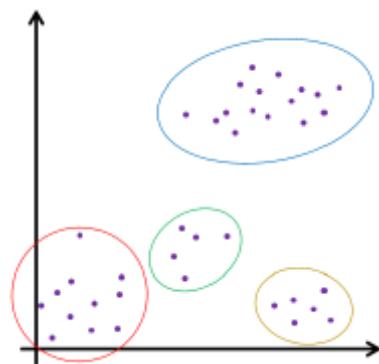
Στο παρακάτω διάγραμμα διασποράς (**Εικόνα 16**) ενός συνόλου δεδομένων, η πρώτη επιλογή στην παραμετροποίηση του αλγόριθμου εκπαίδευσης οπού πρέπει να γίνει είναι η επιλογή του αριθμού των clusters. Στην συγκεκριμένη περίπτωση οι καλύτερες επιλογές φαίνεται πως είναι η χρήση 3 ή 4 clusters, οπότε δημιουργούνται 2 μοντέλα με 3 (**Εικόνα 17**) και 4 (**Εικόνα 18**) clusters αντίστοιχα.



Εικόνα 16: Αρχικό διάγραμμα διασποράς [353]



Εικόνα 17: Διαχωρισμός με επιλογή 3 clusters [353]

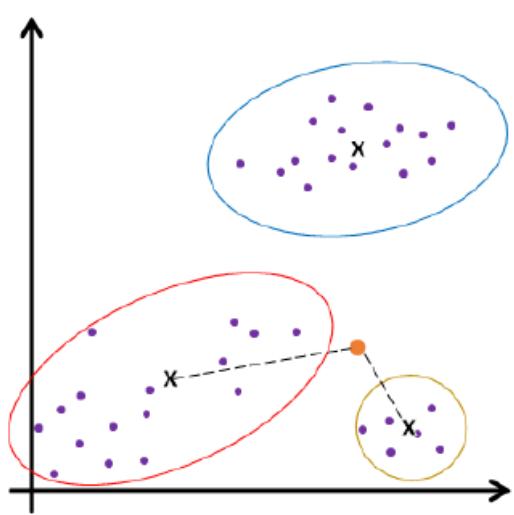


Εικόνα 18: Διαχωρισμός με επιλογή 4 clusters [353]

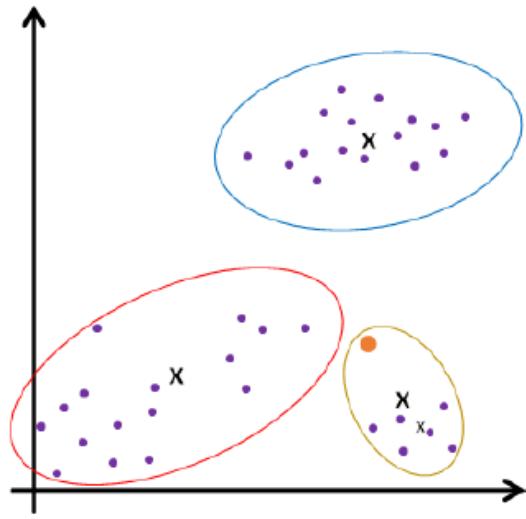
Έπειτα δοκιμάζουμε να εισάγουμε μια νέα οντότητα σε κάθε ένα από τα 2 λειτουργικά μοντέλα για να ελέγξουμε πως θα ομαδοποιηθεί. Για την ομαδοποίηση θα χρησιμοποιηθούν

οι μέθοδοι των K nearest neighbors (KNN) και η μέθοδος ομαδοποίησης με βάση τα κέντρα βάρους κάθε cluster.

Αρχικά εισάγουμε τη νέα οντότητα στο διάγραμμα που έχει διαχωριστεί σε 3 συστάδες. Στη συνέχεια χρησιμοποιούμε τη μέθοδο των κέντρων, όπου χαράσσουμε 2 ευθείες από τη νέα οντότητα προς το κέντρο των 2 εγγύτερων clusters (**Εικόνα 19**). Συγκρίνοντας τις 2 ευθείες, η μικρότερη είναι εκείνη που συνδέει τη νέα οντότητα με το κίτρινο cluster όπου και θα ενταχθεί αναπροσαρμόζοντας τα όρια του (**Εικόνα 20**).

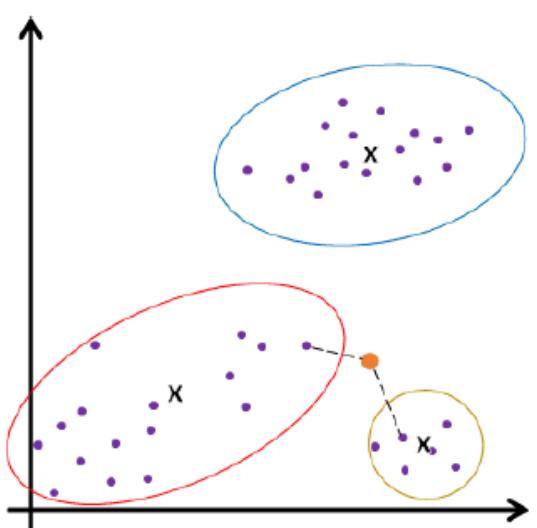


Εικόνα 19: Εισαγωγή της νέας οντότητας στο μοντέλο των 3 clusters [354]

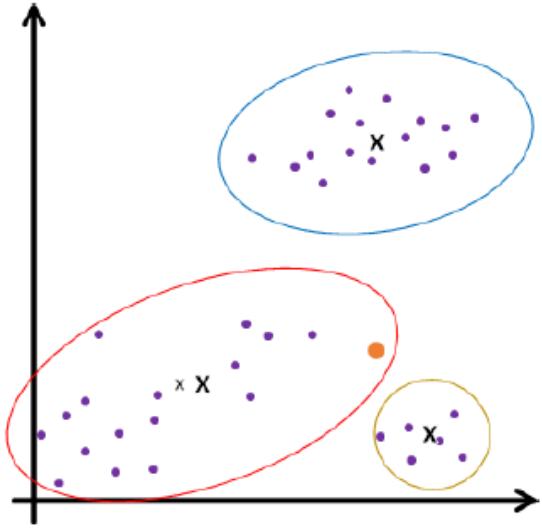


Εικόνα 20: Ομαδοποίηση της νέας οντότητας στο μοντέλο των 3 clusters μέσω της μεθόδου των κέντρων [354]

Στη συνέχεια χρησιμοποιούμε πρώτα την μέθοδο K nearest neighbors (KNN). Για να δούμε το που θα ενταχθεί, βάσει της KNN χαράσσουμε 2 ευθείες από την νέα οντότητα προς στην κοντινότερες οντότητες των 2 εγγυτέρων clusters (**Εικόνα 21**). Συγκρίνοντας τις 2 ευθείες, η μικρότερη είναι εκείνη που συνδέει τη νέα οντότητα με το κόκκινο cluster όπου και θα ενταχθεί αναπροσαρμόζοντας τα όρια του (**Εικόνα 22**).

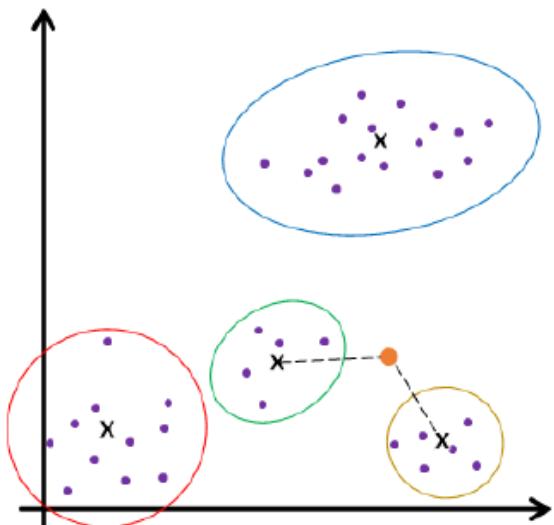


Εικόνα 21: Εισαγωγή της νέας οντότητας στο μοντέλο των 3 clusters [355]

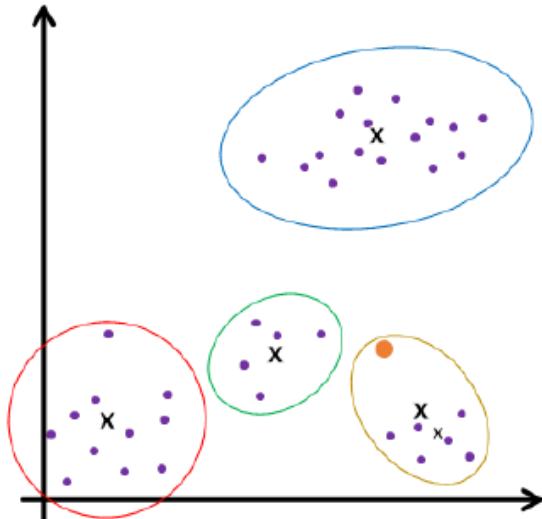


Εικόνα 22: Ομαδοποίηση της νέας οντότητας στο μοντέλο των 3 clusters μέσω της μεθόδου των KNN [355]

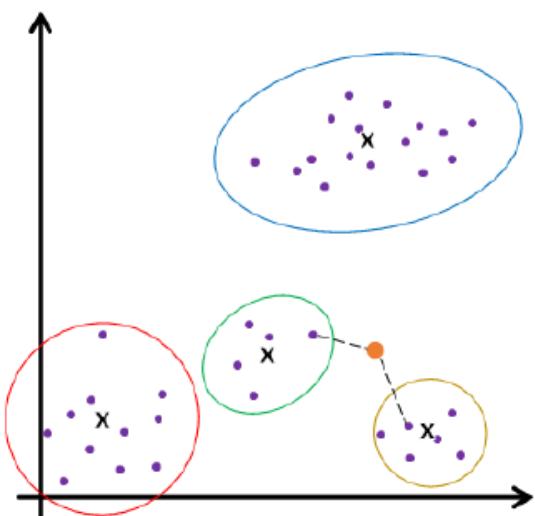
Επαναλαμβάνουμε τις παραπάνω ενέργειες και στο μοντέλο με τα 4 clusters (**Εικόνες 23-26**).



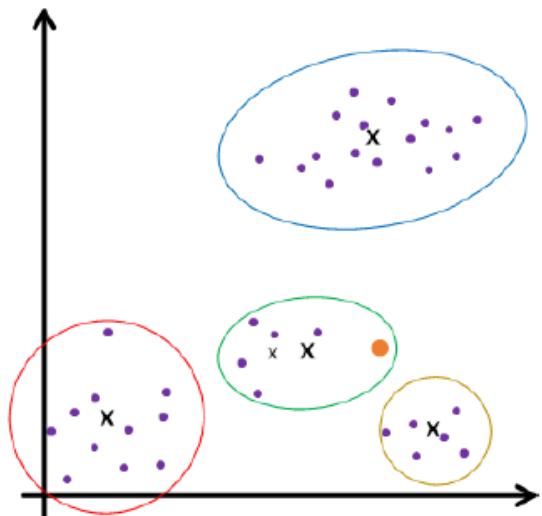
Εικόνα 23: Εισαγωγή της νέας οντότητας στο μοντέλο των 4 clusters [356]



Εικόνα 24: Ομαδοποίηση της νέας οντότητας στο μοντέλο των 4 clusters μέσω της κέντρων[356]



Εικόνα 25: Εισαγωγή της νέας οντότητας στο μοντέλο των 4 clusters[357]



Εικόνα 26: Ομαδοποίηση της νέας οντότητας στο μοντέλο των 4 clusters μέσω της KNN[357]

Παρατηρούμε πως μέσω της μεθόδου των κέντρων η νέα οντότητα ομαδοποιήθηκε οριακά στο κίτρινο cluster (**Εικόνα 24**), ενώ στο μοντέλο με τα 3 clusters ομαδοποιήθηκε στο κίτρινο αλλά με ευκρινή διαφορά (**Εικόνα 20**). Επίσης παρατηρούμε πως χρησιμοποιώντας την μέθοδο των KNN η νέα οντότητα ομαδοποιήθηκε στο πράσινο cluster (**Εικόνα 26**), ενώ στο μοντέλο με τα 3 clusters ομαδοποιήθηκε στο κόκκινο (**Εικόνα 22**).

Από το παραπάνω παράδειγμα είναι σαφές πως τόσο ο αριθμός των clusters με βάση τον οποίο θα γίνει η παραμετροποίηση όσο και η μέθοδος που θα χρησιμοποιηθεί για την εκπαίδευση του μοντέλου, παίζουν βασικό ρόλο στην λειτουργία και την απόδοση του

μοντέλου που προκύπτει και το οποίο θα καθορίσει την ομαδοποίηση των νέων οντοτήτων στα clusters.

Η Συσταδοποίηση μπορεί να εφαρμοστεί σε περιπτώσεις όπου υπάρχουν δεδομένα, κυρίως πολυπαραγοντικά, που απαιτείται άμεση, ακριβής και γρήγορη αξιολόγηση. Ορισμένα παραδείγματα χρήσης της Συσταδοποίησης είναι:

- Η κατηγοριοποίηση των email (επιβλαβής, διαφημιστική κ.α.)
- Η βέλτιστη κατηγοριοποίηση πελατών για στοχευμένες πρωθητικές ενέργειες(διαφημίσεις προϊόντων)
- Ο προσδιορισμός τύπου ψευδών ειδήσεων (fake news)
- Η αναγνώριση παράνομων και εγκληματικών δραστηριοτήτων

2.7 Δένδρα Απόφασης (Decision Trees)

Τα Δένδρα Απόφασης (Decision Trees) αποτελούν μια ευρέως χρησιμοποιούμενη μέθοδο Μηχανικής Μάθησης [358] όπου επιχειρείται η προσέγγιση μιας κατηγορικής συνάρτησης στόχου, μέσω της τεχνικής του «διαιρεί και βασίλευε» (Divide and Conquer). Το σύνολο των οντοτήτων (dataset) του προβλήματος χωρίζεται σε κατηγορίες που αποτελούνται από οντότητες οι οποίες έχουν την ίδια τιμή ως προς κάποια μεταβλητή (χαρακτηριστικό), με τη διαδικασία να επαναλαμβάνεται αναδρομικά, κατασκευάζουν μία οπτική παρουσίαση των κανόνων των επιμέρους κατηγοριοποιήσεων ως το δένδρο απόφασης του συγκεκριμένου μοντέλου [359].

Στον τομέα της εξόρυξης δεδομένων, τα ΔΑ αποτελούν ένα προγνωστικό μοντέλο που χρησιμοποιείται στην αναπαράσταση των ταξινομητών και των μοντέλων παλινδρόμησης. Αντίθετα στην επιχειρησιακή έρευνα, τα ΔΑ χρησιμοποιούνται για την κατασκευή ενός ιεραρχικού μοντέλου αποφάσεων, μαζί με τις συνέπειες που αυτές έχουν. Έτσι ανάλογα με τον εκάστοτε στόχο εκτελείται η διαδικασία ανάπτυξης του ΔΑ για το μοντέλο της στρατηγικής με τη μεγαλύτερη πιθανότητα για να επιτύχει ο στόχος αυτός. Τα ΔΑ τα οποία χρησιμοποιούνται για διαδικασίες ταξινόμησης, αναφέρονται συνήθως ως Δένδρα Ταξινόμησης, ενώ αυτά που χρησιμοποιούνται για διαδικασίες παλινδρόμησης, αναφέρονται ως Δένδρα παλινδρόμησης.

Τα ΔΑ συνθέτουν μία οπτική παρουσίαση των κανόνων, συμβάλλοντας σημαντικά στη προτίμησή τους από πολλούς ως μία από τις πιο διαδεδομένες μεθόδους ταξινόμησης. Οι μέθοδοι των ΔΑ δύναται να χρησιμοποιηθούν στην πρόβλεψη της τάξης που ανήκουν κάποια δεδομένα (ταξινόμηση), στην πρόβλεψη κάποιας συγκεκριμένης τιμής της εξαρτημένης μεταβλητής από τις ανεξάρτητες μεταβλητές (παλινδρόμηση), αλλά και στη μείωση του όγκου δεδομένων, μετά από τον μετασχηματισμό τους σε πιο συμπιεσμένη μορφή ενώ παράλληλα διατηρούν τα βασικά χαρακτηριστικά τους. Το κάθε ΔΑ συντίθεται με τη χρήση του συνόλου των δεδομένων/εγγραφών εκπαίδευσης (training set). Η κάθε εγγραφή χαρακτηρίζεται από το σύνολο χαρακτηριστικών (attributes) και την τάξη κατάταξης (label class). Η λογική πίσω από την κατασκευή ενός ΔΑ είναι η σωστή και ακριβής σχέση των χαρακτηριστικών αυτών και της τάξης κατάταξης.

Χαρακτηριστικά των ΔΑ

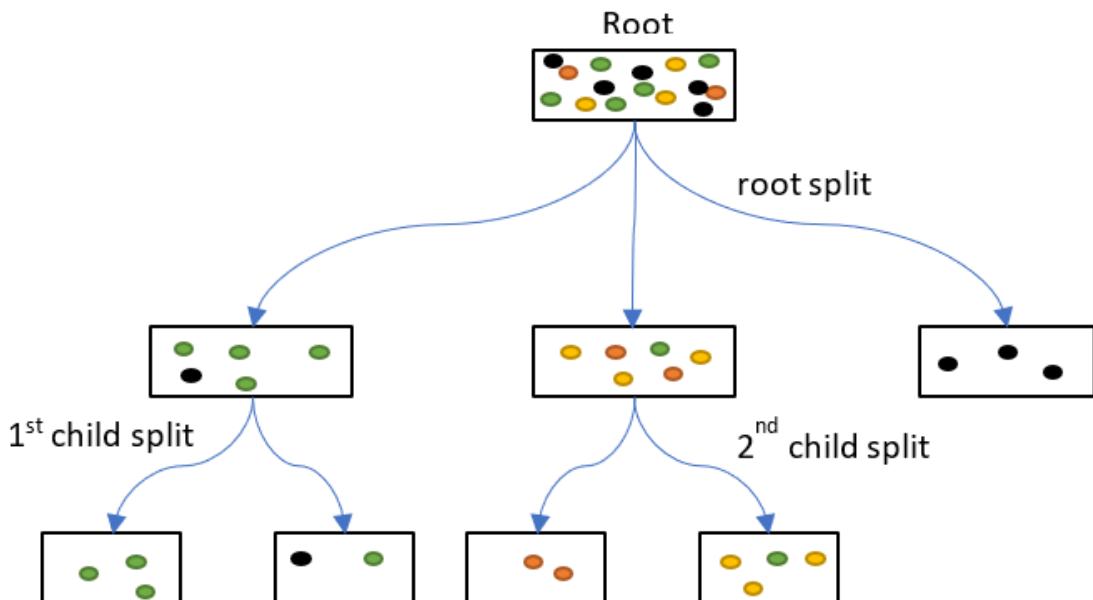
Κάθε Δένδρο Απόφασης είναι ένας ταξινομητής ο οποίος εκφράζεται ως ή αναδρομική διχοτόμηση του συνόλου των γεγονότων. Ένα ΔΑ αποτελείται από διασυνδεόμενους μεταξύ τους κόμβους με σημείο εκκίνησης τη «ρίζα» του, η οποία είναι ο κόμβος που δεν έχει καμία εισερχόμενη άκρη ως σημείο εκκίνησης. Όλοι οι υπόλοιποι κόμβοι έχουν ακριβώς μία

εισερχόμενη ακμή. Οι κόμβοι με εξερχόμενη άκρη ονομάζονται «εσωτερικοί» ή «εξέτασης». Τέλος οι υπόλοιποι κόμβοι που δεν έχουν εξερχόμενη ακμή ονομάζονται «φύλλα», «τερματικοί» ή «κόμβοι απόφασης» [360].

Ο κάθε εσωτερικός κόμβος χωρίζει το χώρο των γεγονότων σε δύο ή περισσότερους υπόχωρους (κλαδιά) ανάλογα με μία συνάρτηση κατηγοριοποίησης των τιμών των χαρακτηριστικών της εισόδου. Στην πιο απλή και συχνή περίπτωση, που το χαρακτηριστικό είναι κατηγορική μεταβλητή τότε ο χώρος του γεγονότος κατανέμεται σύμφωνα με τις διαφορετικές τιμές που λαμβάνει, ενώ στην περίπτωση που το χαρακτηριστικό είναι ποσοτική μεταβλητή, τότε η κάθε κατηγορία που χωρίζεται ο χώρος του γεγονότος, θα περιλαμβάνει ένα εύρος τιμών.

Κάθε φύλλο ως τερματικός κόμβος αποτελείται από μία τιμή της τάξης κατάταξης ή εναλλακτικά, το φύλλο μπορεί να περιλαμβάνει ένα διάνυσμα πιθανοτήτων που υποδηλώνει την πιθανότητα το χαρακτηριστικού στόχου να έχει μια ορισμένη τιμή.

Τα γεγονότα (αποφάσεις) ταξινομούνται, ανάλογα με την πλοιήγησή τους, ξεκινώντας από τη ρίζα του δέντρου και καταλήγοντας σε ένα φύλλο, ακολουθώντας τα αποτελέσματα των γεγονότων (αποφάσεων) κατά μήκος της διαδρομής. Πιο συγκεκριμένα, ξεκινάμε από τη ρίζα ενός ΔΑ και έπειτα ανάλογα με το χαρακτηριστικά που αντιστοιχούν σε αυτή εντοπίζουμε σε ποιο κλαδί αντιστοιχεί η παρατηρούμενη αξία για το δοσμένο χαρακτηριστικό, εν συνεχεία μεταβαίνουμε στον κόμβο όπου συνδέεται με το συγκεκριμένο κλαδί. Τη διαδικασία αυτή την επαναλαμβάνουμε έως ότου καταλήξουμε σε ένα φύλλο. Παρακάτω απεικονίζεται η δομή ενός ΔΑ (Εικόνα 27).



Εικόνα 27: Δομή ενός τυπικού Δένδρου Απόφασης [361]

Προϋποθέσεις για τη χρήση ΔΑ

Οι Βασικές προϋποθέσεις για τη χρήση ενός ΔΑ επαγωγικής μάθησης σε ένα σύνολο δεδομένων (γεγονότων) είναι:

- Ο καθορισμός ενός συνόλου χαρακτηριστικών πριν την εξαγωγή του κανόνα ταξινόμησης.
- Η ύπαρξη προκαθορισμένων διακριτών κατηγοριών ταξινόμησης (Classes) θα οδηγήσει στην αναγνώριση της δενδροειδούς μορφής που θα αναπτύξει ο αλγόριθμος:
- Η ύπαρξη επαρκούς αριθμού δειγμάτων για την δημιουργία του εκπαιδευτικού συνόλου (Training Set, TS).

Στόχοι των ΔΑ

Οι αντικειμενικοί στόχοι των ΔΑ είναι:

- να ταξινομήσουν σωστά όσο το δυνατόν περισσότερο ποσοστό από το σύνολο εκπαίδευσης.
- να μπορούν να γενικεύσουν πέρα από το δείγμα εκπαίδευσης, ώστε ένα άγνωστο δείγμα να μπορεί να ταξινομηθεί με όσο το δυνατό μεγαλύτερη ακρίβεια γίνεται.
- να έχουν την δυνατότητα να εντατικοποιούνται (update), όταν έχουν στη διάθεσή τους περισσότερα δεδομένα.
- να έχουν όσο το δυνατόν πιο απλή δομή.

Μερικοί από τους πιο δημοφιλείς αλγόριθμους δημιουργίας δέντρων αποφάσεων είναι ο CART, ο ID3 [362] και η τροποποιημένη έκδοσή του ο C4.5 [363].

Πλεονεκτήματα και Μειονεκτήματα

Πλεονεκτήματα

- Ο φυσικός και κατανοητός τρόπος αναπαράστασης της γνώσης.
- Αυτόματη παραγωγή εκτελέσιμου κώδικα, εφόσον κάθε δέντρο μεταφράζεται σε μια ακολουθία «**if-then**» εντολών.
- Αποτελεί παραδοσιακό μοντέλο μηχανικής μάθησης.

Μειονεκτήματα

- Υπάρχει δυσκολία στη μετατροπή αριθμητικών τιμών σε αντίστοιχη κατηγορία κατηγορημάτων.
- Είμαστε αναγκασμένοι να περιορίζουμε κλαδιά που αναπτύσσονται συνεχώς επειδή η υπερβολική ανάπτυξη μειώνει μεν το σφάλμα στο σύνολο εκπαίδευσης, αλλά αυξάνει το πλήθος των απαιτούμενων ελέγχων σφάλματος.
- Υπάρχει δυσκολία στη διαχείριση χαρακτηριστικών με πολλές τιμές.
- Ο αλγόριθμος δεν λειτουργεί σωστά, όταν στο σύνολο εκπαίδευσης υπάρχουν δείγματα χωρίς τιμές σε ορισμένα από τα χαρακτηριστικά τους, σύνηθες φαινόμενο κατά τη συγκέντρωση και καταγραφή δειγμάτων.

Εφαρμογές των ΔΑ

Τα Δένδρα Απόφασης αποτελούν τον πιο διαδεδομένο τρόπο Επιβλεπόμενης Επαγωγικής Μάθησης για την Εξόρυξη Δεδομένων (Data Mining). Ενδεικτικά ορισμένες από τις εφαρμογές των ΔΑ είναι [364,365]:

- στην Ιατρική μέσω της ανάπτυξης συστημάτων Λήψης Ιατρικών Αποφάσεων όπως η διαλογή στα επείγοντα, διαλογή ασθενών με COVID-19 κ.α. όπου μπορούν να χρησιμοποιηθούν ΔΑ για να δοθεί προτεραιότητα στη φροντίδα του ασθενούς, με βάση παράγοντες όπως η ηλικία, το φύλο, τα συμπτώματα κ.α.
- στον τομέα των Επιχειρήσεων μέσω της ανάπτυξης συστημάτων που αποφασίζουν τους τρόπους αξιοποίησης του διαφημιστικού προϋπολογισμού, με βάση τα δημογραφικά δεδομένα τους, την ανάλυση προηγούμενων δεδομένων πωλήσεων κ.α.
- στον τομέα της Οικονομίας μέσω συστημάτων όπου Οι τράπεζες και οι πάροχοι στεγαστικών δανείων θα χρησιμοποιούν ιστορικά δεδομένα για να προβλέψουν πόσο πιθανό είναι ένας δανειολήπτης να αποτληρώσει τις οφειλές του κ.α.
- στην Εκπαίδευση δημιουργώντας συστήματα αξιολόγησης υποτροφιών με βάσει τις προηγούμενες επιδόσεις των σπουδαστών.
- στην ανάπτυξη αυτοματοποιημένων τηλεφωνικών συστημάτων που θα καθοδηγούν εκείνον που καλεί ανάλογα με το ζήτημα το οποίο αντιμετωπίζει και τις ενέργειες που έχει κάνει.

2.8 Υπολογιστική Νοημοσύνη

Η Υπολογιστική Νοημοσύνη (Computational Intelligence, CI), αποτελεί έναν κλάδο της TN που βασίζεται στη μίμηση στοιχείων ευφυίας που παρατηρούνται στα έμβια όντα, από ένα υπολογιστικό σύστημα, και προσφέρει τεχνικές ικανές να επιλύουν πολύπλοκα προβλήματα, με υπολογιστικές μεθόδους, που δεν θα μπορούσαν να λυθούν με συμβατικές μεθοδολογίες, χωρίς να είναι απαραίτητο το υπολογιστικό σύστημα, να επιδεικνύει γενική νοημοσύνη [366].

Η ΥΝ χαρακτηρίζεται κυρίως από ιδιότητες, όπως:

- προσαρμοστικότητα (adaptation)
- αυτοοργάνωση (self-organization)
- μάθηση-εξέλιξη (learning-evolution)

Η Υπολογιστική Νοημοσύνη είναι ένας συνδυασμός της TN με τη Θεωρία Συστημάτων με στόχο την αναπαραγωγή στοιχείων της ανθρώπινης ευφυίας με υπολογιστικές τεχνικές και μέσα, ώστε τα υπολογιστικά συστήματα να μπορέσουν να εκτελέσουν έργο για το οποίο ένας άνθρωπος τα καταφέρνει καλύτερα μέχρι τώρα [366]. Επίσης διαθέτει αρκετές μεθόδους ανάλογα με το εκάστοτε πρόβλημα.

Η Υπολογιστική Νοημοσύνη αρχικά περιελάμβανε:

- την Ασαφή Λογική (Fuzzy Logic)
- τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)
- τους Γενετικούς Αλγόριθμους (Genetic Algorithms)

Αργότερα προστέθηκαν και επιπλέον τεχνολογίες όπως:

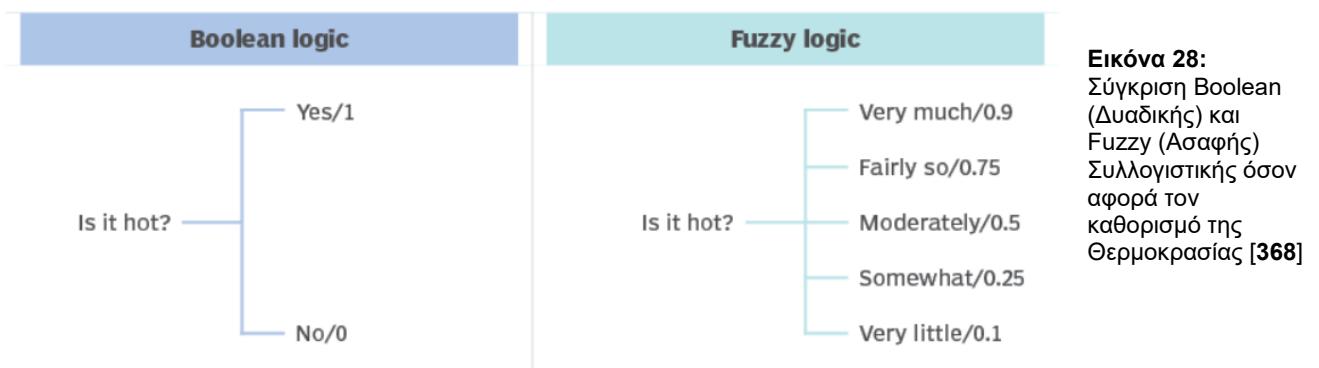
- Τα Συστήματα Στήριξης Αποφάσεων
- Η Μηχανική Μάθηση
- Η Εξόρυξη Δεδομένων

Πρόσφατες δημοσιεύσεις τοποθετούν την Υπολογιστική Νοημοσύνη στον πυρήνα τεχνολογιών αιχμής που ασχολούνται με την επεξεργασία τεράστιου όγκου δεδομένων, την αλληλεπίδραση ανθρώπου-μηχανής, το Ίντερνετ των Αντικειμένων (Internet Of Things, IoT) κ.α.

2.9 Ασαφής Λογική (Fuzzy Logic)

Η ασαφής λογική είναι τομέας της που βασίζεται στη λογική των «βαθμών αλήθειας» και όχι στη συνηθισμένη αληθής (1) ή ψευδής (0) λογική Boolean στην οποία βασίζεται ο σύγχρονος υπολογιστής. Η ιδέα της ασαφούς λογικής εισήχθη για πρώτη φορά από τον Lotfi Zadeh του Πλανετιστημίου της Καλιφόρνια στο Μπέρκλεϋ τη δεκαετία του 1960, όταν αυτός εργαζόταν πάνω στο πρόβλημα της κατανόησης της φυσικής γλώσσας από τους υπολογιστές [367]. Η φυσική γλώσσα όπως οι περισσότερες άλλες δραστηριότητες στη ζωή δεν είναι δυνατόν να μεταφραστεί εύκολα στους απόλυτους όρους του 0 και του 1, με αποτέλεσμα πολλά δεδομένα που χρειάζεται να τροφοδοτήσουμε έναν υπολογιστή να είναι σε κάποια ενδιάμεση κατάσταση. Έτσι η Ασαφής Λογική θα μπορούσε να αποτελεί τον βασικό τρόπο που λειτουργεί η συλλογιστική με την Boolean λογική να αποτελεί απλά μια ειδική περίπτωση της.

Στα συστήματα TN, η Ασαφής Λογική χρησιμοποιείται για τη μίμηση του ανθρώπινου συλλογισμού και της γνώσης. Αντί για αυστηρά δυαδικές περιπτώσεις αλήθειας, η ασαφής λογική περιλαμβάνει το 0 και 1 ως ακραίες περιπτώσεις αλήθειας αλλά με διάφορους ενδιάμεσους βαθμούς αλήθειας. Στην **Εικόνα 28** έχουμε την σύγκριση της Ασαφούς Λογικής με την Διαδικτή με σκοπό τον καθορισμό της θερμοκρασίας.



Τα χαρακτηριστικά της ασαφούς λογικής είναι [366,369]:

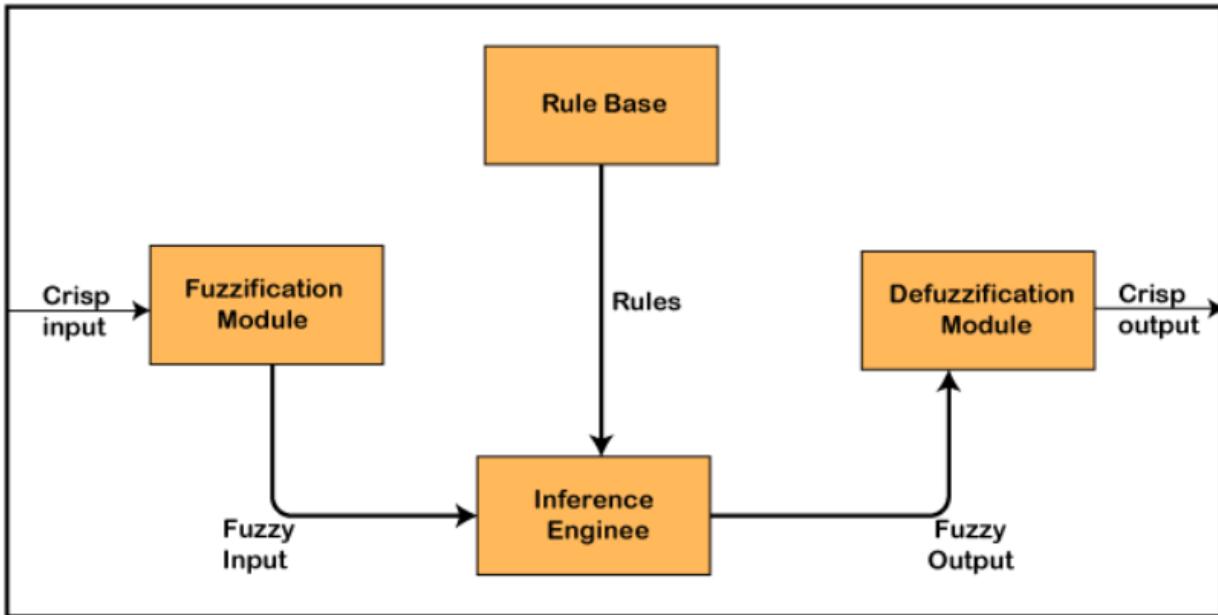
- Είναι ευέλικτη και μπορούμε εύκολα να την κατανοήσουμε και να την εφαρμόσουμε.
- Βοηθάει στην ελαχιστοποίηση των λογικών που δημιουργούνται από τον άνθρωπο.
- Είναι η καλύτερη μέθοδος για την εύρεση της λύσης εκείνων των προβλημάτων που είναι κατάλληλα για κατά προσέγγιση ή αβέβαιο συλλογισμό.
- Προσφέρει πάντα δύο τιμές, οι οποίες δηλώνουν τις δύο πιθανές λύσεις για ένα πρόβλημα και μια δήλωση.
- Επιτρέπει στους χρήστες να δημιουργήσουν λογικές συναρτήσεις που είναι μη γραμμικές και αυθαίρετης πολυπλοκότητας.
- Στη ασαφή λογική, όλα είναι βαθμονομημένα.
- Στην Ασαφή λογική, κάθε σύστημα που είναι λογικό μπορεί εύκολα να ασαφοποιηθεί (fuzzified).
- Βασίζεται στην επεξεργασία φυσικής γλώσσας.
- Χρησιμοποιείται επίσης από τους ποσοτικούς αναλυτές για τη βελτίωση της εκτέλεσης του αλγορίθμου τους.

Αρχιτεκτονική ενός Συστήματος Ασαφής Λογικής

Ένα Σύστημα Ασαφούς Λογικής (Fuzzy Logic System, FLS), όπως αυτό της **Εικόνας 29**, αποτελείται από τέσσερα δομικά μέρη [366,369]:

1. Τον **Ασαφοποιητή (Fuzzification)** ο οποίος μετατρέπει την ευκρινή τιμή της κάθε εισόδου του συστήματος σε Ασαφή Δεδομένα. Αυτό το πετυχαίνει διαιρώντας τα σήματα εισόδου στις ακόλουθες πέντε καταστάσεις:
 - Μεγάλη Θετική (Large Positive, LP)
 - Μέση Θετική (Medium Positive , MP)
 - Μικρή (Small, S)
 - Μεσαία Αρνητική (Medium Negative, MN)
 - Μεγάλη αρνητική (Large negative, LN)
2. Τη **Βάση των Κανόνων(Rule Base, RB)**, η οποία περιέχει τα Ασαφή Δεδομένα (σύνολα) και τους Ασαφείς Κανόνες σύμφωνα με τους οποίους το σύστημα θα κληθεί να λάβει αποφάσεις. Το πόσο αποτελεσματική είναι η λήψη αποφάσεων στηρίζεται σε μεγάλο βαθμό από τον εμπλουτισμό της RB, με σωστά ορισμένα σύνολα κανόνων τα οποία αντανακλούν τους ανθρώπινους κανόνες και γνώσεις καθώς και με Ασαφείς Κανόνες οι οποίοι αναπαριστούν με αποτελεσματικά την ανθρώπινη εμπειρία.
3. Τη **Μηχανή Ασαφούς Συμπεράσματος (Fuzzy Inference Engine, FIE)**, η οποία βασιζόμενη στις αρχές της Ασαφούς Λογικής, λαμβάνει ασαφή σύνολα από την είσοδό της και αντιστοιχεί το κάθε ένα στον κατάλληλο γι' αυτό Ασαφή Κανόνα, από τη Βάση των Κανόνων, δημιουργώντας έτσι ένα Ασαφές Σύνολο ως έξοδο (output). Υπάρχουν αρκετές Μηχανές Ασαφούς Συμπεράσματος (Larsen, Tsukamoto, Sugeno κ.α.), κάθε μία με τα δικά της πλεονεκτήματα. Η επιλογή της σωστότερης εξαρτάται από το είδος της εφαρμογής που θα υλοποιηθεί.

4. Τον **Αποασαφοποιητή (Defuzzification)**, που λαμβάνει ως είσοδο το Ασαφές Σύνολο που δημιουργείται από τη Μηχανή Ασαφούς Συμπεράσματος και στη συνέχεια το μετατρέπει σε μια Ευκρινή Τιμή (Crisp Value). Αποτελεί το τελευταίο βήμα της διαδικασίας του Σύστηματος Ασαφούς Λογικής. Η Ευκρινής Τιμή είναι ο τύπος τιμής ο οποίος είναι αποδεκτός από τον χρήστη. Υπάρχουν αρκετές διαφορετικές τεχνικές, από τις οποίες ο εκάστοτε χρήστης θα πρέπει να επιλέξει εκείνη που περιέχει τα λιγότερα σφάλματα.



Εικόνα 29: Διάγραμμα ενός Συστήματος Ασαφούς Λογικής (FLS) [370]

Σύμφωνα με τα παραπάνω, θα ήταν δυνατό να εφαρμόσουμε ένα Σύστημα Ασαφούς Λογικής (FLS) για τον έλεγχο ενός συστήματος θέρμανσης. Το FLS λαμβάνει στην είσοδο του, από τον αισθητήρα της θερμοκρασίας και θα καθορίζει το πόσο ψυχρή ή θερμή θα είναι. Έτσι ανάλογα με το ποιοι Ασαφείς Κανόνες ενεργοποιούνται κάθε φορά, το FLS, σύμφωνα με τις τιμές εισόδου, θα επιστρέψει ως έξοδό μία συγκεκριμένη σαφή τιμή η οποία θα καθοδηγεί την ηλεκτροβάνα, στο να αυξήσει η να μειώσει τη ροή του θερμού νερού στις σωληνώσεις.

Ορισμένοι από τους τους τομείς όπου μπορεί να εφαρμοστεί η Ασαφής Λογική είναι οι εξής:

- Στον τομέα των Ηλεκτρονικών για τον έλεγχο της θερμοκρασίας και της υγρασίας του χώρου, του χρονισμού των πλυντηρίων ρούχων κ.α.
- Στον Δευτερογενή τομέα για τη βελτιστοποίηση της παραγωγής τυριού, γάλακτος κ.α.
- Στην Αυτοκινητοβιομηχανία για τον έλεγχο της ταχύτητας, των αυτοοδηγούμενων αμαξιών κ.α.
- Στον τομέα των Μαζικών Μεταφορών για την βελτίωση της αυτόματης λειτουργίας υπόγειων συρμών, τον έλεγχο του χρονοδιαγράμματος των συρμών κ.α..
- Στην Αεροναυπηγική με την ανάπτυξη συστημάτων ελέγχου υψομέτρου για διαστημικά σκάφη, τη ρύθμιση της ροής και του μείγματος των οχημάτων απόψυξης των αεροσκαφών κ.α.

- Στον Θαλάσσιο τομέας μέσω της ανάπτυξης του αυτόματου πιλότου πλεύσης, αυτόνομων υποβρύχιων οχημάτων κ.α.
- Στον τομέα της Πολεμικής Άμυνας μέσω της ανάπτυξης συστημάτων αναγνώρισης στόχων, συστημάτων λήψης αποφάσεων κ.α.
- Στον τομέας των Επιχειρήσεων μέσω της ανάπτυξης συστημάτων λήψης αποφάσεων, αξιολόγησης προσωπικού κ.α.
- Στην Οικονομία μέσω της ανάπτυξης συστημάτων ελέγχου μεταφοράς τραπεζογραμματίων, χρηματιστηριακών προβλέψεων κ.α.
- Στην Ιατρική με την ανάπτυξη υποστηρικτικών διαγνωστικών συστημάτων, συστημάτων ελέγχου της αρτηριακής πίεσης κατά την αναισθησία, την μοντελοποίηση νευροπαθολογικών ευρημάτων σε ασθενείς με Alzheimer κ.α.
- Στον τομέας της Ψυχολογίας μέσω της κατασκευής FLS ανάλυσης της ανθρώπινης συμπεριφοράς κ.α.
- Στην Αναγνώριση και ταξινόμηση μοτίβων μέσω FLS, μέσω της ανάπτυξης μοντέλων αναγνώρισης ομιλίας, εικόνων, χειρόγραφων κ.α.

2.10 Εξελικτικοί Αλγόριθμοι (*Evolutionary Algorithms*)

Οι Εξελικτικοί Αλγόριθμοι (Evolutionary Algorithms, EA) αποτελούν μια κατηγορία μεθόδων βελτιστοποίησης οι οποίοι αποπειρώνται να μιμηθούν, ως ένα βαθμό, την ίδια την εξελικτική διαδικασία. Αποτελούν ένα υπερσύνολο μεθόδων το οποίο περιλαμβάνει τον Εξελικτικό Προγραμματισμό, τους Γενετικούς Αλγόριθμους με τον Γενετικό Προγραμματισμό, τα Συστήματα Ταξινόμησης, και τις Στρατηγικές Εξέλιξης [371].

Για την εκπαίδευση ενός μοντέλου μέσω των Εξελικτικών Αλγόριθμων γίνεται η χρήση μεθόδων Ενισχυμένης Μάθησης και για την υλοποίησή τους εφαρμόζονται τεχνικές εμπνευσμένες από τη βιολογική εξέλιξη των ειδών και γι' αυτό και χρησιμοποιούν όρους από την επιστήμη της Βιολογίας. Σκοπός των Εξελικτικών Αλγόριθμων είναι η εξοικείωση με τις φυσικές διεργασίες ώστε να αναπαρασταθούν αλγορίθμικά για την επίλυση προβλημάτων αναζήτησης και βελτιστοποίησης.

Οι όροι της Βιολογίας που χρησιμοποιούνται για την περιγραφή των διαδικασιών των Εξελικτικών Αλγορίθμων είναι οι ακόλουθοι:

- **Γονιδιακή Δεξαμενή** (gene pool) καλείται το σύνολο της γενετικής ποικιλομορφίας ενός υπό εξέταση πληθυσμού.
- **Πληθυσμός** (population) καλείται το σύνολο των ομοειδών αλληλο-αναπαραγόμενων οντοτήτων.
- **Γενεά** (generation) είναι το κάθε νέο σύνολο οντοτήτων (απόγονοι) το οποίο προκύπτει από το προηγούμενο (πρόγονοι).
- **Γενότυπος** (genotype) ονομάζεται το σύνολο των γονιδίων τα οποία έχει μια οντότητα, συχνά ο όρος χρησιμοποιείται ως αναφορά στην **οντότητα** (entity) ή στο **άτομο** (individual) για την περιγραφή ενός στοιχείου του πληθυσμού.
- **Χρωμόσωμα** (chromosome) αποτελεί μια οργανωμένη δομή γονιδίων.
- **Γονίδιο** (gene) είναι μια ιδιότητα ή χαρακτηριστικό, το οποίο κατέχει συγκεκριμένη θέση σε ένα χρωμόσωμα και έχει την δυνατότητα να κληροδοτηθεί.

- **Γονιδιακή Θέση** ή **Γενετικός Τόπος** (*locus*) ονομάζεται η θέση ενός γονιδίου σε ένα χρωμόσωμα.
- **Αλληλόμορφα** (*alleles*) ονομάζονται οι διαφορετικές μορφές ή παραλλαγές ενός γονιδίου, τα οποία ελέγχουν το ίδιο γνώρισμα, αλλά με διαφορετικό τρόπο και εντοπίζεται σε συγκεκριμένο γενετικό τόπο (*locus*), ενός χρωμοσώματος.

Οι Εξελικτικοί Αλγόριθμοι είναι μέθοδοι αναζήτησης των βέλτιστων λύσεων, από ένα σύνολο το οποίο περιέχει όλες τις δυνατές λύσεις του διθέντος προβλήματος. Σε πρώτο στάδιο ένας ΕΑ ξεκινάει δημιουργώντας έναν αρχικό σύνολο από τις δυνατές λύσεις, τις οποίες στη συνέχεια αξιολογεί, μέσα από διεργασίες που μιμούνται τους βιολογικούς μηχανισμούς της εξέλιξης (διασταύρωση, μετάλλαξη κ.α.), με αποτέλεσμα να προκύψουν νέες λύσεις από τις οποίες επιλέγονται οι καταλληλότερες από τις οποίες θα προκύψουν οι επόμενες. Η διαδικασία επαναλαμβάνεται είτε μέχρι να βρεθεί η βέλτιστη λύση σύμφωνα με τα προεπιλεγμένα κριτήρια, είτε μέχρι να ολοκληρωθεί ένας ορισμένος αριθμός επαναλήψεων (*iterations*).

Τα κύρια βήματα τα οποία ακολουθεί ένας Εξελικτικός Αλγόριθμος (**Εικόνα 30**) για τον προσδιορισμό της βέλτιστης λύσης είναι [372]:

1. **Η Αρχικοποίηση** (*initialization*) που είναι η διαδικασία για τη δημιουργία του αρχικού πληθυσμού (1^η γενιά οντοτήτων). Από την πρώτη γενιά (**g**) θα προκύψουν οι επόμενες. Ο καθορισμός της 1^{ης} γενιάς γίνεται είτε τυχαία είτε με τη χρήση προϋπάρχουσας γνώσης, με σκοπό να κατευθύνει και να διευκολύνει την αναζήτηση.
2. **Αξιολόγηση** (*evaluation*) κατά την οποία η κάθε οντότητα αξιολογείται ως προς το κατά πόσο πλησιέστερα βρίσκεται στο επιθυμητό αποτέλεσμα, λαμβάνοντας μια τιμή καταλληλότητας (*fitness score*). Όσο καταλληλότερη θεωρείται μια οντότητα ως βέλτιστη λύση, τόσο μεγαλύτερη είναι αυτή η τιμή καταλληλότητας.
3. **Επιλογή** (*selection*) είναι το στάδιο το οποίο έπεται της αξιολόγησης και στο οποίο επιλέγονται οι οντότητες που θα διασταυρωθούν μεταξύ τους για να δημιουργηθεί η επόμενη γενιά. Η πιθανότητα που έχει μια οντότητα να επιλεγεί, για να περάσει τα γονίδια της στην επόμενη γενιά, έχει άμεσή συσχέτιση με τη βαθμολογία που έλαβε κατά τη διαδικασία της αξιολόγησης.
4. **Η Δημιουργία της νέας γενιάς** ολοκληρώνεται αφού ο Γενετικός Αλγόριθμος αξιοποιήσει τις βαθμολογίες των οντοτήτων του τρέχοντος πληθυσμού. Κάθε γενιά αποτελείται από έναν προκαθορισμένο αριθμό οντοτήτων ο οποίος συνήθως είναι ίσος με αυτόν του αρχικού πληθυσμού. Ο αριθμός επιλέγεται με διάφορα κριτήρια, όπως το πόσο κοστοβόρος ανάλογα με τους πόρους του εκάστοτε υπολογιστικού συστήματος στο οποίο εκτελείται ο αλγόριθμος.

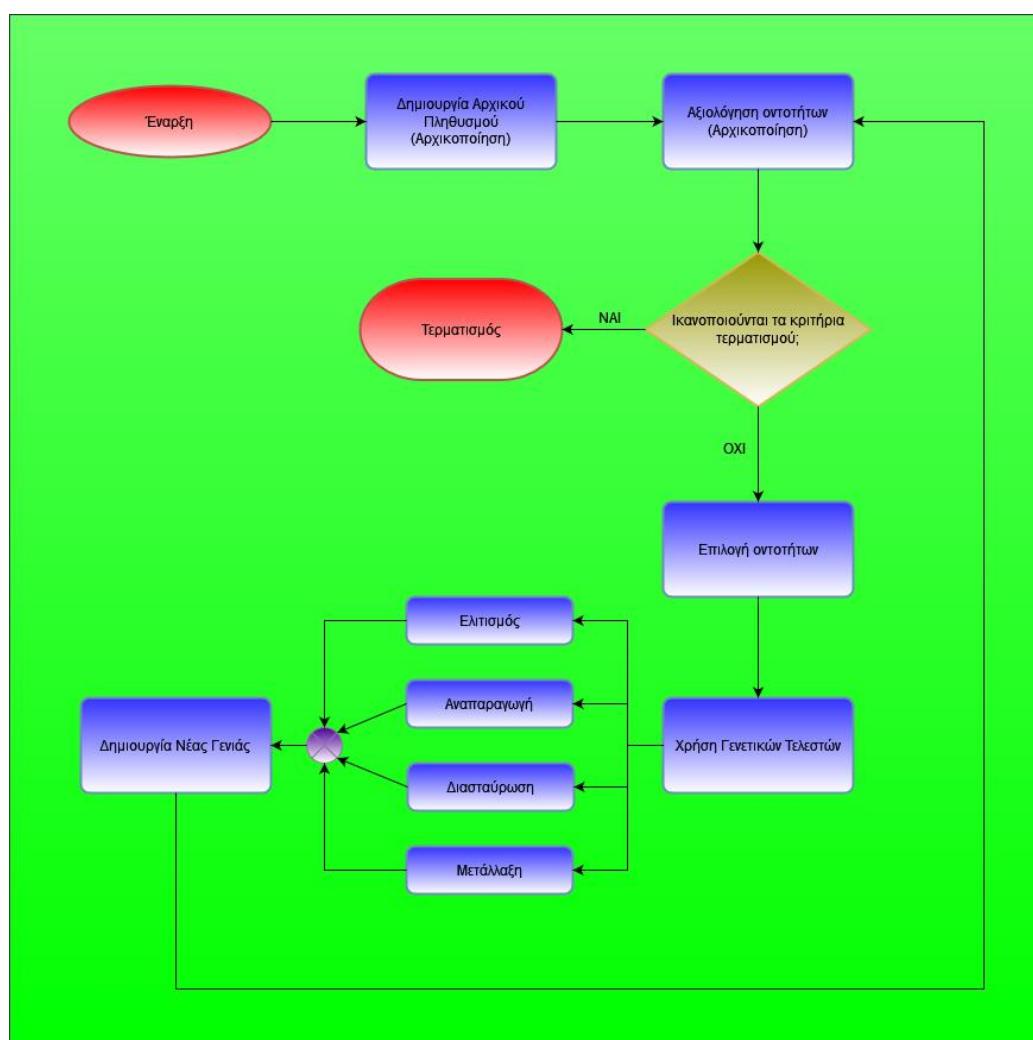
Οι μέθοδοι που ακολουθούνται για τη δημιουργία μιας νέας γενιάς ενός νέου πληθυσμού ονομάζονται Γενετικοί Τελεστές με τους πιο ευρέως χρησιμοποιούμενους να είναι:

- **Ελιτισμός** (*elitism*) είναι η μέθοδος με την οποία μια οντότητα μπορεί να επιλέγει ώστε να περάσει αυτούσια στην επόμενη γενιά αν πληροί πολύ συγκεκριμένα κριτήρια, όπως το να πετυχαίνει ιδιαίτερα μεγάλη βαθμολογία έναντι των υπολοίπων.
- **Αναπαραγωγή** (*replication*) καλείται η μέθοδος κατά την οποία μια οντότητα αντιγράφεται αυτούσια στην επόμενη γενιά μέσω τυχαίας επιλογής και

χαρακτηρίζεται και ως «ρουλέτα επιλογής». Η πιθανότητα μια οντότητα να επιλεγεί είναι ανάλογη με το πόσο υψηλή βαθμολογία έλαβε κατά την αξιολόγηση.

- **Διασταύρωση (crossover)** είναι η διαδικασία στην οποία έχουμε ανταλλαγή γενετικής πληροφορίας ανάμεσα σε 2 οντότητες, μιμούμενη την βιολογική αναπαραγωγή κατά την οποία ο απόγονος κληρονομεί διαφορετικά γονίδια από το κάθε γονέα.
- **Μετάλλαξη (mutation)** είναι η μέθοδος σύμφωνα με την οποία μια αλλαγή στην τιμή ενός γονιδίου περνάει από τον πρόγονο στον απόγονο του. Η συχνότητα εμφάνισης μιας μετάλλαξης (mutation rate) ορίζεται κατά την παραμετροποίηση του αλγορίθμου.

Σύμφωνα με τα παραπάνω, υπάρχει πιθανότητα μια οντότητα να εμφανιστεί δύο φορές σε μια προκύπτουσα γενιά ή να επανεμφανιστεί ξανά σε μελλοντική γενιά. Μετά τη δημιουργία της κάθε γενιάς ο αλγόριθμος προχωρά στην διαδικασία της Αξιολόγησης μέχρι να βρεθεί η βέλτιστη λύση ή έως να τερματιστεί ο αλγόριθμος βάσει άλλων κριτηρίων (προκαθορισμένος αριθμός επαναλήψεων).



Υπάρχουν αρκετοί μέθοδοι βελτιστοποίησης ενός EA, ανάλογα με την περίπτωση, το είδος του προβλήματος, τους διαθέσιμους υπολογιστικούς πόρους κ.α. Για παράδειγμα, ένας EA

μπορεί να αρχικοποιηθεί με πληθυσμό που απαρτίζεται από μεγάλο αριθμό οντοτήτων και εν συνέχεια, ο πληθυσμός της επόμενης γενιάς να μειωθεί κατά το ήμισυ, έτσι ώστε να υπάρχει αρκετά μεγάλη ποικιλία επιλογών αρχικά αλλά μετά την αξιολόγηση να εστιάσει στις επικρατέστερες βάσει βαθμολογικών κριτηρίων. Οι Γενετικοί Τελεστές μπορούν επίσης να παραμετροποιηθούν σύμφωνα με τις ανάγκες του προβλήματος, χωρίς να είναι απαραίτητο να χρησιμοποιηθούν όλοι.

Επίσης η μεθοδολογία που ακολουθούν οι EA είναι απλή και ευέλικτη, καθιστώντας την αξιοποιήσιμη είτε ως έναν αυτόνομο τρόπο επίλυσης είτε χρησιμοποιούμενη σε συνδυασμό με άλλες μεθοδολογίες, δημιουργώντας έτσι αρκετά εξειδικευμένες υβριδικές μορφές EA. Ακόμα οι EA δεν χρησιμοποιούν τις μεταβλητές ενός προβλήματος αλλά μία κωδικοποιημένη μορφή του συνόλου τιμών που οι μεταβλητές αυτές μπορούν να πάρουν. Επιπλέον οι EA είναι εφαρμόσιμοι σε δυναμικά συστήματα όπου ο στόχος και οι περιορισμοί του κάθε προβλήματος ενδέχεται να μεταβάλλονται σε σχέση με το χρόνο. Επιπροσθέτως οι EA μπορούν να εφαρμοστούν με επιτυχία σε περιπτώσεις όπου ο χώρος αναζήτησης είναι ασυνεχής, χαοτικός και γενικότερα σε περιπτώσεις όπου οι πιο παραδοσιακές μέθοδοι επίλυσης δεν μπορούν να δώσουν αρκετά ικανοποιητικά αποτελέσματα [372].

Οι EA χρησιμοποιούνται κυρίως σε θέματα βελτιστοποίησης διατηρώντας ανά πάσα στιγμή της λειτουργίας τους, έναν πλήθος λύσεων για το εκάστοτε πρόβλημα, με τις οποίες εργάζονται ταυτόχρονα αναζητώντας τη βέλτιστη.

Ορισμένοι από τους τομείς όπου μπορεί να εφαρμοστούν οι EA είναι οι εξής:

- Στη Μετεωρολογία για την μοντελοποίηση των παγκόσμιων θερμοκρασιακών αλλαγών κ.α.
- Στη Βιολογία για την κατασκευή φυλογενετικών δέντρων, την ανάλυση προφίλ γονιδιακών εκφράσεων, την πρόβλεψη δομής RNA, πρωτεϊνών κ.α.
- Στις Οικονομικές Επιστήμες για τον σχεδιασμό αυτοματοποιημένων συστημάτων συναλλαγών, τον εντοπισμό απατών, την βελτιστοποίηση χαρτοφυλακίου κ.α.
- Στις Κοινωνικές Επιστήμες για τον σχεδιασμό αντιτρομοκρατικών συστημάτων, για γλωσσικές αναλύσεις κ.α.
- Στις Τηλεπικοινωνίες για τη βελτιστοποίηση της υποδομής των δικτύων των οικιακών(σταθερών) και κινητών επικοινωνιών κ.α.
- Στον Σχεδιασμό Συστημάτων για βελτιστοποίηση του σχεδιασμού των συστημάτων παροχής νερού, τον αυτοματοποιημένο σχεδιασμό βιομηχανικού εξοπλισμού, τον αυτοματοποιημένο σχεδιασμό σύνθετων υλικών κ.α.
- Στις Μεταφορές με σκοπό τη βελτιστοποίηση της φόρτωσης κοντέινερ, των δρομολογίων των οχημάτων κ.α.

2.11 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)

Με τον όρο Νευρωνικό Δίκτυο (Neural Network) στη Βιολογία ορίζουμε το δίκτυο που αποτελείται από ένα σύνολο διασυνδεδεμένων νευρώνων που αποτελούν τμήμα του νευρικού συστήματος. Στην Υπολογιστική Νοημοσύνη συναντώνται τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks) τα οποία μιμούνται την λειτουργία των νευρώνων μέσα από αλγορίθμικές τεχνικές με σκοπό την επίλυση σύνθετων υπολογιστικών προβλημάτων

[373,374]. Ο τεχνητός νευρώνας ενός ΤΝΔ ονομάζεται «Αντίληπτρο» ή «νευρώνας Perceptron» [375].

Τα ΤΝΔ υποστηρίζουν μη γραμμικότητα, παρέχοντας απλές υπολογιστικές μονάδες με μεγάλο αριθμό συνδέσεων (συνάψεις). Μερικοί από τους τύπους προβλημάτων που καλούνται να λύσουν είναι ή προσέγγιση συνάρτησης, η αναγνώριση προτύπων, η κατηγοριοποίηση και ομαδοποίηση δεδομένων, η αναγνώριση και πρόβλεψη σημάτων, η εξαγωγή χαρακτηριστικών και πολλά άλλα.

Ορισμένα από τα βασικά χαρακτηριστικά των ΤΝΔ είναι τα παρακάτω:

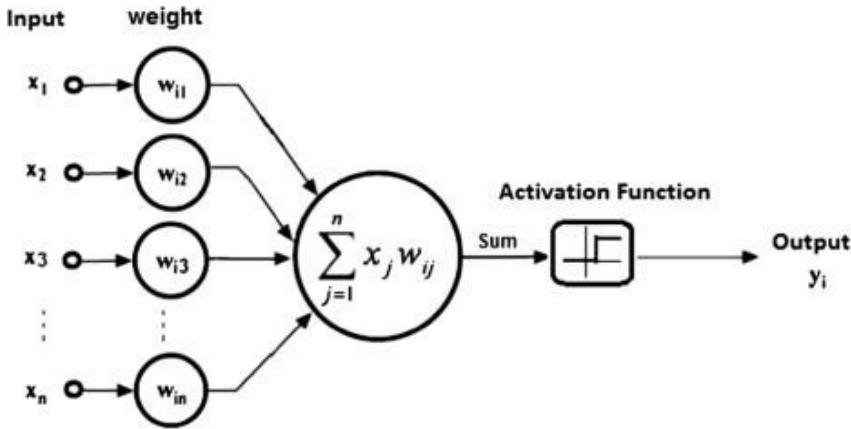
- ❖ η μη-γραμμικότητα
- ❖ η μαζική παραλληλία
- ❖ η κατανεμημένη αναπαράσταση
- ❖ η κατανεμημένη επεξεργασία
- ❖ η ικανότητα μάθησης
- ❖ η ικανότητα γενίκευσης
- ❖ η αντοχή στην αφαίρεση νευρώνων
- ❖ η προσαρμοστικότητα
- ❖ η ανοχή σε σφάλματα
- ❖ η πιθανότητα απομνημόνευσης οδηγώντας σε λάθος γενίκευση

Τα ΤΝΔ μπορούν να εξάγουν ασφαλή συμπεράσματα και με εξαιρετική ακρίβεια και αμελητέα περιθώρια σφάλματος, ακόμα και με ελάχιστες πληροφορίες για το πρόβλημα. Στόχος των ΤΝΔ είναι η ανάκληση ή η αντιστοίχιση, πληροφοριών ή συμπερασμάτων μετά από τη συλλογή, την αξιολόγηση και την επεξεργασία πληροφοριών (ερεθισμάτων). Στα ΤΝΔ όπως και στα αντίστοιχα βιολογικά δίκτυα νευρώνων, οι νευρώνες αλληλεπιδρούν μεταξύ από την ανταλλαγή σημάτων μέσω των συνάψεων που τους συνδέουν, με την κάθε σύναψη να χαρακτηρίζεται από το δικό της συναπτικό βάρος [373].

Βασικό ρόλο στην λειτουργία των ΤΝΔ είναι η κατάλληλη προ-επεξεργασία δεδομένων πριν από την είσοδο, που εξαρτάται από την κάθε εφαρμογή, έτσι ώστε να μειωθεί η διαστατικότητα των δεδομένων, να αυξηθούν ή να ενισχυθούν τα σχετικά χαρακτηριστικά γνωρίσματα και να φιλτραριστεί ή να κατασταλεί πλήρως ο θόρυβος.

Αρχές λειτουργίας και Αρχιτεκτονική

Η βασική μονάδα λειτουργίας ενός ΤΝΔ είναι ο νευρώνας ο οποίος έχει πολλές εισόδους αλλά μόνο μία έξοδο (Multiple Inputs Single Output, MISO) [373]. Το βασικό μοντέλο λειτουργίας του νευρώνα ενός ΤΝΔ απεικονίζεται στο παρακάτω σχήμα (**Εικόνα 31**):

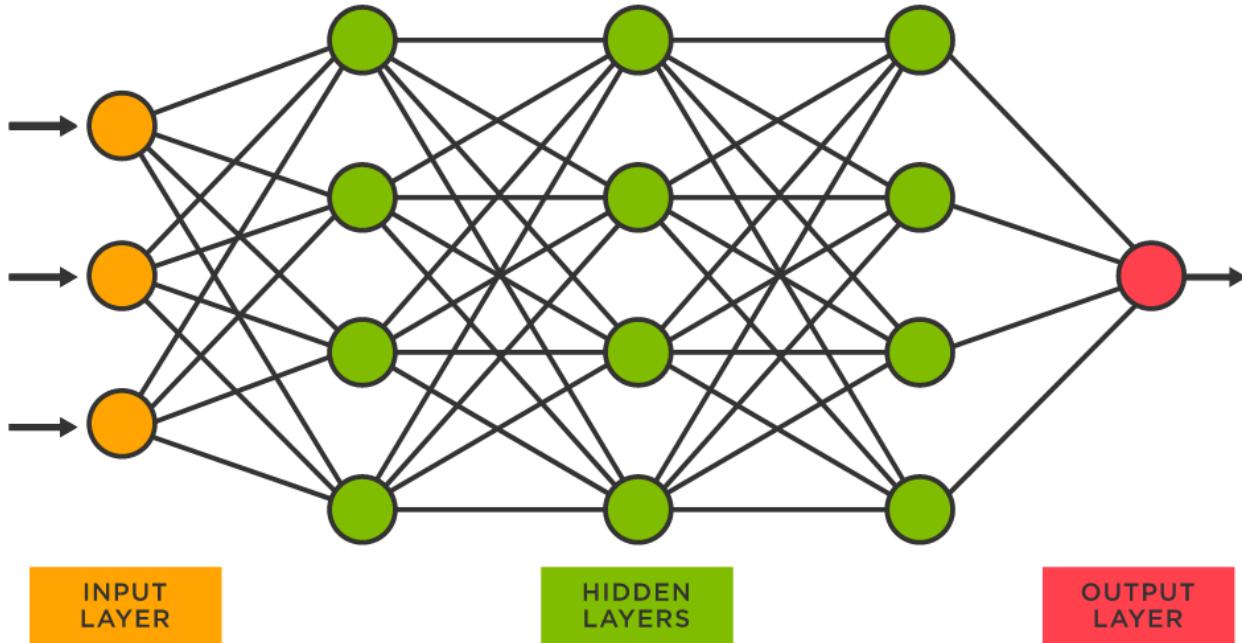


Εικόνα 31: Απεικόνιση βασικού μοντέλου λειτουργίας ενός τεχνητού νευρώνα ενός ΤΝΔ [376]

Έχοντας ως είσοδο το διάνυσμα $\mathbf{X}=(x_1, x_2, x_3, \dots, x_n)$ και βάρη $\mathbf{W}=(w_{i1}, w_{i2}, w_{i3}, \dots, w_{in})$, τα βασικά χαρακτηριστικά ενός νευρώνα είναι τα εξής:

- Τα Συναπτικά Βάρη (Synaptic Weights) τα οποία καθορίζουν τη σημαντικότητα της κάθε εισόδου του νευρώνα. Για παράδειγμα αν είσοδος του νευρώνα i στη σύναψη ij είναι το σήμα x_i , τότε αυτό πολλαπλασιάζεται με το συναπτικό βάρος w_{ij} . Για να μπορέσει να επιτευχθεί η μέγιστη δυνατή ποιότητα των αποτελεσμάτων ενός νευρώνα, είναι απαραίτητο όλα τα συναπτικά βάρη να καθοριστούν με ακρίβεια.
- Ο Αθροιστής (Adder) που δίνει στην έξοδό του το άθροισμα των σταθμισμένων εισόδων $\text{Sum} = \sum_{j=1}^n x_j w_{ij}$
- Η Συνάρτηση Ενεργοποίησης (Activation Function) από τη οποία περνά η έξοδος Sum του αθροιστή και δίνει στην έξοδο του νευρώνα το αποτέλεσμα y_i , ανάλογα με τον τύπο της συνάρτησης $f(\text{Sum})$ που επιλέχθηκε.
- Το κατώφλι (θ_k) ή αλλιώς μεροληψία (**bias**) που πρόκειται για δευτερεύουσα παράμετρο του συστήματος και επιλέγεται με στόχο την καλύτερη ευελιξία. Κατώφλι θεωρούμε συνήθως το συναπτικό βάρος w_{i0} που πολλαπλασιάζεται με το $x_0 (=1)$.

Κατά την λειτουργία του τεχνητού νευρώνα όλα τα σήματα εισόδου x_j πολλαπλασιάζονται με το αντίστοιχο συναπτικό τους βάρος w_{ij} το οποίο καθορίζει τη σημαντικότητα της σύνδεσης και κατόπιν προωθούνται σε έναν Αθροιστή. Το αποτέλεσμα του Αθροιστή (Sum), οδηγείται στη Συνάρτηση Ενεργοποίησης, η οποία θα καθορίσει την τελική έξοδο του νευρώνα και θα αποτελέσει είσοδο για τους νευρώνες του επόμενου στρώματος (layer). Οι νευρώνες μπορούν να διαμορφώσουν διάφορα στρώματα (layers) που το καθένα από αυτά οργανώνεται έτσι ώστε κάθε νευρώνας ενός στρώματος συνδέεται με κάθε νευρώνα του επόμενου.



Εικόνα 32: Απεικόνιση ενός απλού νευρωνικού δικτύου [377]

Οι νευρώνες ενός Νευρωνικού Δικτύου (**Εικόνα 32**) χωρίζονται σε 3 βασικές κατηγορίες:

- τους Νευρώνες εισόδου, που δέχονται τις πληροφορίες που θα υποστούν επεξεργασία.
- τους Ενδιάμεσους (Κρυφούς) νευρώνες.
- τους Νευρώνες εξόδου, οι οποίοι μεταφέρουν τα αποτελέσματα της επεξεργασίας του ΤΝΔ.

Τα ΤΝΔ διαχωρίζονται, ανάλογα με το πλήθος των στρωμάτων τους, σε μονοστρωματικά και πολυστρωματικά. Για τη συγκεκριμένη κατηγοριοποίηση, το επίπεδο εισόδων δεν προσμετράται στο πλήθος των επιπέδων ενός ΤΝΔ. Τα μονοστρωματικά ΤΝΔ δημιουργούν γραμμικά όρια και χρησιμοποιούνται σε πιο απλές εφαρμογές, όπως η ταξινόμηση (classification). Τα πολυστρωματικά ΤΝΔ μπορούν να συνδύασουν γραμμικά όρια και να επιλύσουν μη-γραμμικά προβλήματα ταξινόμησης [373].

Στα πολυστρωματικά ΤΝΔ, όλα τα επίπεδα νευρώνων μεταξύ του επιπέδου της εισόδου και αυτού της εξόδου, ονομάζονται κρυφά επίπεδα. Η αρχιτεκτονική ενός πολυστρωματικού δικτύου αποτελείται από ένα στρώμα εισόδου, ένα στρώμα εξόδου και ένα ή περισσότερα κρυφά στρώματα. Κάθε κόμβος σε κάθε στρώμα συνδέεται με κάθε νευρώνα του επόμενου στρώματος μέσω σταθμισμένων με συναπτικά βάρη συνδέσεων. Το στρώμα εισόδου του δικτύου λαμβάνει κάθε είσοδο από τα αρχικά δεδομένα, το κρυφό στρώμα παίζει το ρόλο του ανιχνευτή χαρακτηριστικών γνωρισμάτων, ενώ τα τελικά αποτελέσματα δίνονται από το στρώμα εξόδου. Ο σκοπός των κρυφών στρωμάτων είναι για να επιτρέψει στο δίκτυο να αντιμετωπίσει τα δεδομένα με μη γραμμικό τρόπο δίνοντας του έτσι την δυνατότητα να επιλύσει μη γραμμικά προβλήματα κατηγοριοποίησης, ταξινόμησης ή αναγνώρισης.

Όλες οι είσοδοι σταθμίζονται από τη μήτρα συναπτικών βαρών **W1** και διασυνδέονται πλήρως στα κρυφά στρώματα. Κάθε νευρώνας που ανήκει σε κρυφό στρώμα αθροίζει τα αποτελέσματα στάθμισης του στρώματος εισόδου και εφαρμόζει μια μη γραμμική συνάρτησης ενεργοποίησης. Αυτό το άθροισμα των σταθμισμένων εξόδων από το προηγούμενο στρώμα ονομάζεται ενεργοποίηση με τη μη γραμμικότητα να παίζει το ρόλο της συνάρτησης ενεργοποίησης. Έπειτα τα αποτελέσματα των κρυφών νευρώνων

σταθμίζονται από τη μήτρα συναπτικών βαρών **W2** και παρέχονται στους νευρώνες του επόμενου κρυφού στρώματος, εάν υπάρχει, ή αλλιώς στους νευρώνες στρώματος εξόδου. Οι κόμβοι εξόδου αθροίζουν τα αποτελέσματα του κρυφού στρώματος και εφαρμόζουν τη δικιά τους συνάρτηση ενεργοποίησης, η οποία μπορεί να είναι η ίδια με αυτή του κρυφού στρώματος ή οποιαδήποτε άλλη γραμμική ή μη γραμμική.

Τα ΤΝΔ σύμφωνα τη κατεύθυνση της πληροφορίας εντός του δικτύου χωρίζονται στα [373]:

- ΤΝΔ πρόσθιας τροφοδότησης (Feedforward ANN) στα οποία έχουμε τη σύνδεση των εξόδων των νευρώνων ενός επιπέδου με τις εισόδους των νευρώνων του επόμενου.
- ΤΝΔ οπισθοδιάδοσης (Feedback ANN) όπου έχουμε τη σύνδεση των εξόδων των νευρώνων ενός επιπέδου είτε με μία ή περισσότερες εισόδους νευρώνων ενός μη-επόμενου επιπέδου, έχοντας έτσι μερική οπισθοδιάδοση (partially connected ANN), είτε με όλες τις εισόδους νευρώνων κάθε μη-επόμενου επιπέδου έχοντας πλήρη οπισθοδιάδοση (fully connected ANN).

Εκπαίδευση ΤΝΔ

Τα νευρωνικά δίκτυα «μαθαίνουν» μέσω αλγορίθμων εκπαίδευσης. Είναι γνωστό πως τα Νευρωνικά δίκτυα χρησιμοποιούνται συνήθως μαζί με αλγόριθμους μάθησης στην εξόρυξη γνώσης. Σημαντικός περιορισμός των τεχνητών νευρωνικών δικτύων είναι πως σπάνια μπορεί να εξηγηθεί το πώς κατέληξαν σε ένα συμπέρασμα, εξαιτίας της μη γραμμικότητας, και για αυτό το λόγο αναφέρονται συχνά ως «μαύρα κουτιά».

Τα ΤΝΔ εκπαιδεύονται κυρίως με τρεις διαφορετικούς τρόπους:

1. **Εκπαίδευση με Εποπτεία** όπου το σύνολο των προτύπων (γνωστές τιμές) περιέχει τόσο διανύσματα εισόδου όσο και διανύσματα εξόδου. Ο στόχος της εκπαίδευσης είναι η ελαχιστοποίηση του καθολικού σφάλματος και η προσαρμογή των τιμών των σωστών βαρών με τέτοιο τρόπο ώστε όταν διαβιβάζουμε στο δίκτυο μια είσοδο αυτό να αναπαράγει τη σωστή έξοδο.
2. **Εκπαίδευση χωρίς Εποπτεία** όπου το σύνολο των προτύπων (γνωστές τιμές) περιέχει μόνο διανύσματα εισόδου τα οποία ομαδοποιούνται με βάση τα κοινά τους χαρακτηριστικά. Ο στόχος είναι και εδώ η ελαχιστοποίηση του καθολικού σφάλματος της προκύπτουσα συνάρτησης.
3. **Εκπαίδευση με Ανταγωνισμό** όπου οι νευρώνες ανταγωνίζονται για το ποιος θα συμμετέχει πιο ενεργά στην εκπαίδευση, έτσι ώστε να προκύψει ο επικρατών νευρώνας για την κάθε διαφορετική κλάση διανυσμάτων. Έτσι ο κάθε νευρώνας εξειδικεύεται στην ταυτοποίηση μιας μόνο κλάσης διανυσμάτων. Με τον τρόπο αυτό η κλάση του ενός διανύσματος ενεργοποιεί ένα μοναδικό νευρώνα κάθε φορά.

Τα συναπτικά βάρη αποτελούν την κωδικοποιημένη εμπειρική γνώση που αποκτά το ΤΝΔ, είτε μέσω της εκπαίδευσης είτε μέσω της αλληλεπίδρασής του με το περιβάλλον αφού τεθεί σε λειτουργία [373]. Αποτελούν το κύριο χαρακτηριστικό το οποίο δίνει στο ΤΝΔ την δυνατότητα να εξελίσσεται και να προσαρμόζεται συνεχώς ανάλογα με το περιβάλλον του. Η εκπαίδευση των ΤΝΔ γίνεται μέσω ειδικών αλγόριθμων που ονομάζονται «αλγόριθμοι εκπαίδευσης» ή «κανόνες μάθησης» και έχουν ως στόχο την βελτίωση της τελικής απόδοσης του ΤΝΔ. Η διαδικασία έχει επαναληπτικό χαρακτήρα. Μετά από κάθε επανάληψη, τα συναπτικά βάρη αναπροσαρμόζονται, ανάλογα με την έξοδο, ώστε να αποδώσουν τις επιθυμητές τιμές εξόδου. Έτσι το ΤΝΔ αποκτά περισσότερη γνώση την οποία

και αποθηκεύει εσωτερικά. Η εκπαίδευση τερματίζεται όταν δεν χρειάζεται να γίνει καμία άλλη αναπροσαρμογή των συναπτικών βαρών ή όταν έχει παρέλθει ένας συγκεκριμένος αριθμός επαναλήψεων. Η μέθοδος εκπαίδευσης που θα επιλέγει για ένα TNΔ εξαρτάται τόσο από την αρχιτεκτονική του όσο και από την εφαρμογή στην οποία θα κληθεί να δραστηριοποιηθεί.

Αφού ολοκληρωθεί η εκπαίδευση, το TNΔ είναι έτοιμο να τεθεί σε λειτουργία και να αποδώσει σωστές τιμές εξόδου χάρη στην ικανότητα ανάκλησης πληροφοριών από την μνήμη του [373]. Με τον όρο Ανάκληση (*recall*) αναφερόμαστε στη διαδικασία υπολογισμού ενός διανύσματος εξόδου για συγκεκριμένο διάνυσμα εισόδου. Ένα TNΔ ενδέχεται να γενικεύει σε περιπτώσεις όπου το διάνυσμα που θα λάβει στην είσοδο δεν το «γνωρίζει» ή «θυμάται» από την εκπαίδευσή την εμπειρία του. Γενίκευση καλείται η ικανότητα ενός TNΔ να αποδίδει με επιτυχία διανύσματα εξόδου, για διανύσματα εισόδου τα οποία δεν έχει συναντήσει στο παρελθόν. Η ικανότητα γενίκευσης ενός TNΔ, εξαρτάται κυρίως από την πολυπλοκότητα του προβλήματος, το μέγεθος και την αρχιτεκτονική του TNΔ, καθώς και από το μέγεθος τους συνόλου των διανυσμάτων που χρησιμοποιήθηκαν κατά την εκπαίδευση.

Πλεονεκτήματα και Μειονεκτήματα των TNΔ

Πλεονεκτήματα:

- Τα TNΔ βελτιώνονται από την μάθηση, πράγμα που μπορεί να συνεχιστεί ακόμη και μετά την εφαρμογή του συνόλου εκπαίδευσης
- Η χρήση των TNΔ μπορεί να παραλληλοποιηθεί ευκολότερα και με καλύτερη απόδοση.
- Μπορούμε να επιτύχουμε μεγαλύτερη ακρίβεια όταν εφαρμόζεται το κατάλληλο σύνολο εκπαίδευσης
- Τα TNΔ δεν επηρεάζονται τόσο πολύ από την ύπαρξη θορύβου

Μειονεκτήματα:

- Η δημιουργία κανόνων από TNΔ είναι πολύπλοκη διαδικασία.
- Οι τιμές των χαρακτηριστικών εισόδου πρέπει να είναι αριθμητικές.
- Τα TNΔ χρειάζονται Δοκιμές και Επαλήθευση.
- Μπορεί να εμφανιστούν φαινόμενα υπερπροσαρμογής.
- Η φάση της μάθησης μπορεί να αποτύχει να συγκλίνει.
- Τα νευρωνικά δίκτυα μπορεί να είναι αρκετά ακριβά από πλευράς χρόνου ή ενέργειας στη χρήση τους.

Εφαρμογές των TNΔ

Ενδεικτικά ορισμένοι από τους τομείς όπου βρίσκουν εφαρμογή τα TNΔ είναι:

- Στην Ιατρική για την ανάλυση συμπτωμάτων, την πρόβλεψη αντίδρασης οργανισμών σε νέες δραστικές ουσίες κ.α.

- Στην Φαρμακευτική για την ανάπτυξη εξατομικευμένων σχεδίων θεραπείας
- Στην Αεροδιαστημική για την ανίχνευση βλαβών, τον έλεγχο του αεροσκάφους, την προσομοίωση πτήσεων κ.α.
- Στη Βιομηχανία για τον έλεγχο διεργασιών, τον ποιοτικό έλεγχο, τη διάγνωση βλαβών κ.α.
- Στην Ηλεκτρονική για την τεχνητή όραση, τη σύνθεση ήχου κ.α.
- Στην Οικονομία για την πρόβλεψη δεικτών χρηματιστηρίου, την αξιολόγηση αιτήσεων τραπεζικών δανείων κ.α.
- Στην επιστήμη των Υπολογιστών μέσω της βελτίωσης της ακρίβειας των μηχανών αναζήτησης

Χρήση τεχνικών της Μηχανικής Μάθησης στην καταπολέμηση της πανδημίας COVID-19

Η έκρηξη του αριθμού των λοιμώξεων από την πανδημία της COVID-19 έχει οδηγήσει σε παγκόσμιες προσπάθειες για τον έλεγχο και τον περιορισμό της εξάπλωσης της. Μια σημαντική γραμμή έρευνας πραγματοποιείται με τη χρήση της Μηχανικής Μάθησης (ML) με σκοπό την κατανόηση και την καταπολέμηση της πανδημίας. Υπάρχει ήδη πληθώρα ερευνών. Εδώ θα εστιάσουμε στις μελέτες της ML που χρησιμοποιώντας τα μέχρι στιγμής διαθέσιμα κλινικά και εργαστηριακά δεδομένα, στοχεύουν στη διάγνωση, στην πρόβλεψη της σοβαρότητας της νόσησης και της θνησιμότητας της COVID-19, καθώς και σε έρευνες που εστιάζουν τόσο στην πρόβλεψη των μελλοντικών μεταλλάξεων του SARS-CoV-2, όσο και στην ταχύτερη ανάπτυξη θεραπευτικών στρατηγικών έναντι αυτού. Επίσης γίνεται αναφορά στις πτυχές των μελετών οι οποίες σχετίζονται με τους τύπους των αλγορίθμων, τα σύνολα δεδομένων εκπαίδευσης και την επιλογή των χαρακτηριστικών. Ο κύριος όγκος των αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται σε αυτές τις δύο εφαρμογές είναι αλγορίθμοι εποπτευόμενης μάθησης. Τα μοντέλα που αναπτύχθηκαν δεν έχουν ακόμα εφαρμοστεί στην αντιμετώπιση της πανδημίας, με μεγάλο μέρος των σχετικών ερευνών να είναι πειραματική. Τα διαγνωστικά και προγνωστικά χαρακτηριστικά που ανακαλύφθηκαν από τα μοντέλα ML είναι συνεπή με τα αποτελέσματα που παρουσιάζονται στην ιατρική βιβλιογραφία. Κύριο αντικείμενο του μέρους αυτού της εργασίας, είναι να τονιστούν οι συνεχιζόμενες προσπάθειες που πραγματοποιούνται μέσω της ML στο πλαίσιο της μάχης για την εξάλειψη της πανδημίας της COVID-19.

3.1 Διάγνωση της COVID-19

Τα πιο κοινά συμπτώματα του COVID-19 είναι ξηρός βήχας, πονόλαιμος και πυρετός τα οποία μπορεί να εξελιχθούν σε σοβαρή μορφή πνευμονίας με κρίσιμες επιπλοκές, όπως σηπτικό σοκ, πνευμονικό οίδημα, σύνδρομο οξείας αναπνευστικής δυσχέρειας και πολυοργανική ανεπάρκεια. Οπότε είναι απαραίτητο να υπάρξει έγκαιρη διάγνωση ώστε να αποφευχθούν τόσο τα βαρύτερα συμπτώματα όσο και η εξάπλωση του ιού. Πάνω σε αυτό το κομμάτι έχει γίνει πληθώρα ερευνών οι οποίες με τη χρήση μεθόδων της επιβλεπόμενης μηχανικής μάθησης και χρησιμοποιώντας δεδομένα ιατρικής απεικόνισης (ακτινογραφίες ή αξονικές τομογραφίες), μετρήσεις των ζωτικών λειτουργιών (σφυγμοί και αναπνευστικός ρυθμός) και δεδομένα των συμπτωμάτων μαζί με τα γενικά χαρακτηριστικά προσπαθούν να προβλέψουν το αν ο υπό εξέταση ασθενής πάσχει ή όχι από COVID-19.

3.1.1 Με τη Χρήση κλινικών δεικτών

Ορισμένοι κλινικοί δείκτες, με την βοήθεια της ML θα μπορούσαν να βοηθήσουν στην έγκαιρη διάγνωση της COVID-19. Η εξαγωγή χαρακτηριστικών από τους κλινικούς δείκτες (συμπτώματα, φύλο, ηλικία, κορεσμός οξυγόνου, καρδιακοί παλμοί κ.α.) αποτελεί ένα σημαντικό βήμα για την εκπαίδευση μοντέλων ML, καθώς η απόδοση του εκπαίδευόμενου μοντέλου εξαρτάται άμεσα από την ποιότητα των εξαγόμενων χαρακτηριστικών. Παρακάτω ακολουθούν ορισμένες μελέτες που κατασκευάστηκαν μοντέλα πρόβλεψης μέσω ML, χρησιμοποιώντας ως δεδομένα τους κλινικούς δείκτες των εξεταζόμενων για COVID-19.

Yazeed Zoabi et al. (January 2021)

Στη μελέτη του Yazeed Zoabi και των συνεργατών του [378], με σκοπό την πρόβλεψη ασθενών με COVID-19 χρησιμοποιήθηκε ένα μοντέλο μηχανικής μάθησης το οποίο επεξεργάζεται τις απαντήσεις σε 8 βασικές ερωτήσεις. Το μοντέλο εκπαιδεύτηκε από τα δεδομένα όλων των ατόμων στο Ισραήλ που έκαναν τεστ για SARS-CoV-2 τους πρώτους μήνες της πανδημίας COVID-19.

Το σύνολο των δεδομένων αποτελείται από τα στοιχεία που δημοσίευσε το Υπουργείο Υγείας του Ισραήλ τα οποία περιέχουν τα δεδομένα των ατόμων έκαναν τεστ RT-PCR ανάλυσης του ρινοφαρυγγικού επιχρίσματος για τον SARS-CoV-2 [379]. Στο σύνολο δεδομένων περιέχονται οι αρχικές εγγραφές, σε καθημερινή βάση, όλων των κατοίκων που εξετάστηκαν, όπως η ημερομηνία δοκιμής και το αποτέλεσμα του τεστ, τα κλινικά συμπτώματα, το φύλο και το αν το άτομο είναι ηλικίας 60 ετών και άνω. Το σετ εκπαίδευσης-επικύρωσης αποτελείται από αρχεία 51.831 ατόμων που έκαναν το τεστ την περίοδο 22-31 Μαρτίου 2020 και εκ των οποίων 4.769 επιβεβαιώθηκε ότι έχουν COVID-19. Το δοκιμαστικό σετ περιείχε δεδομένα από την επόμενη εβδομάδα, 1-7 Απριλίου, όπου από 47.401 άτομα τα οποία έκαναν το τεστ 3.624 επιβεβαιώθηκε ότι έχουν COVID-19. Το σετ εκπαίδευσης-επικύρωσης χωρίστηκε περαιτέρω σε σετ εκπαίδευσης και επικύρωσης με αναλογία 4 προς 1 (**Πίνακας 2**).

(#)	Feature	Total n = 99,232		COVID-19 negative n = 90,839		COVID-19 positive n = 8393	
		n	%	n	%	n	%
(1)	Sex						
	Male	50,350	50.74	45,545	50.1	4805	57.2
	Female	48,882	49.26	45,294	49.8	3588	42.7
(2)	Age 60+						
	True	15,279	15.4	13,619	14.9	1660	19.7
	False	83,953	84.6	77,220	85	6733	80.2
(3)	Cough						
	True	14,768	14.88	10,715	11.8	4053	48.2
	False	84,223	84.87	79,909	87.9	4314	51.4
(4)	Fever						
	True	8122	8.18	4387	4.83	3735	44.5
	False	90,868	91.5	86,237	94.9	4631	55.1
(5)	Sore throat						
	True	1273	1.28	96	0.11	1177	14
	False	95,062	95.8	88,059	96.9	7003	83.4
(6)	Shortness of breath						
	True	930	0.94	71	0.08	859	10.2
	False	95,405	96.14	88,084	96.9	7321	87.2
(7)	Headache						
	True	1799	1.81	68	0.07	1731	20.6
	False	94,536	95.27	88,087	96.9	6449	76.8
(8)	Known contact with an individual confirmed to have COVID-19						
	True	5507	5.55	1455	1.6	4052	48.2
	False	93,725	94.45	89,384	98.4	4341	51.8

Πίνακας 2: Χαρακτηριστικά του συνόλου δεδομένων και τα χαρακτηριστικά που χρησιμοποιούνται από το μοντέλο σε αυτή τη μελέτη. [380]

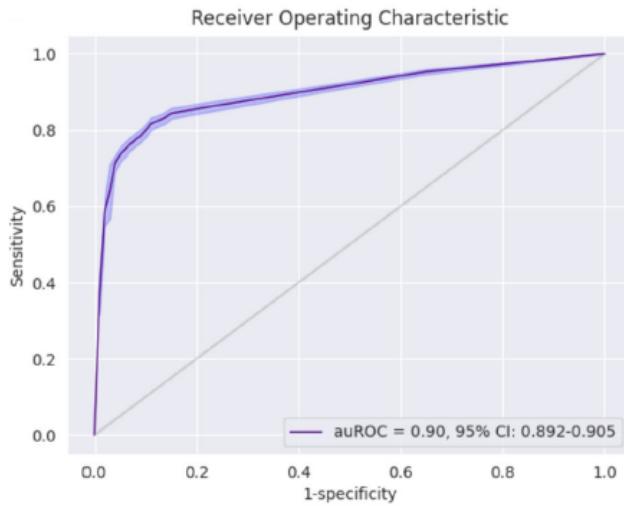
Με βάση αυτά τα δεδομένα, αναπτύχθηκε ένα μοντέλο το όποιο να μπορεί να προβλέψει το αν κάποιος πάσχει από COVID- 19 χρησιμοποιώντας οκτώ δυαδικά χαρακτηριστικά τα οποία είναι τα παρακάτω:

- Βασικές πληροφορίες:
 1. Φύλο (αρσενικό/θηλυκό)
 2. Ηλικία ≥ 60 ετών (Σωστό/Λάθος)
- Συμπτώματα:
 3. Βήχας (Σωστό/Λάθος)
 4. Πυρετός (Σωστό/Λάθος)
 5. Πονόλαιμος (Σωστό/Λάθος)
 6. Δύσπνοια (Σωστό/Λάθος)
 7. Πονοκέφαλος (Σωστό/Λάθος)
- Άλλες πληροφορίες:
 8. Γνωστή επαφή με άτομο που έχει επιβεβαιωθεί ότι έχει COVID-19 (Σωστό/Λάθος).

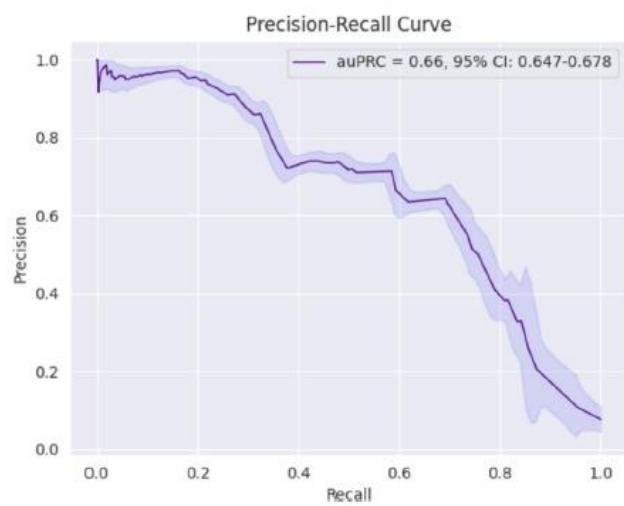
Οι προβλέψεις δημιουργήθηκαν χρησιμοποιώντας ένα μοντέλο ενίσχυσης κλίσης με τη χρήση δένδρων αποφάσεων [381]. Η ενίσχυση κλίσης θεωρείται ο καλύτερος τρόπος για την πρόβλεψη των δεδομένων με δομή πίνακα [382] και χρησιμοποιείται από πολλούς αλγόριθμους στον τομέα της ML [383]. Οι τιμές που έλειπαν διαχειρίστηκαν από το προγνωστικό ενίσχυσης κλίσης [384] το οποίο και εκπαιδεύτηκε με το πακέτο LightGBM [385] της Python. Το σετ επικύρωσης χρησιμοποιήθηκε για πρώτη διακοπή [386], με μετρητή απόδοσης την περιοχή κάτω από την ανάλυση καμπύλης του λειτουργικού χαρακτηριστικού του δέκτη (auROC).

Για να προσδιοριστούν τα κύρια χαρακτηριστικά με βάσει τα οποία θα πραγματοποιηθούν οι προβλέψεις του μοντέλου οδήγησης, υπολογίστηκαν οι τιμές SHAP [387] οι οποίες είναι κατάλληλες για πολύπλοκα μοντέλα όπως τα τεχνητά νευρωνικά δίκτυα και οι μηχανές ενίσχυσης κλίσης [388]. Οι τιμές SHAP διαχωρίζουν το αποτέλεσμα πρόβλεψης για το κάθε δείγμα ανάλογα με τη συνεισφορά της κάθε τιμής του χαρακτηριστικού, μέσω της εκτίμησης των διαφορών μεταξύ των υποσυνόλων που αποτελούν το σύνολο των τιμών των χαρακτηριστικών. Έτσι χρησιμοποιώντας τον μέσο όρο Με τον μέσο όρο τιμών των δειγμάτων, οι τιμές SHAP υπολογίζουν τη συνεισφορά του κάθε χαρακτηριστικού στις τελικές προβλέψεις.

Για το μελλοντικό σύνολο δοκιμών, το μοντέλο προέβλεψε με auROC ίσο με 0,90 με όριο εμπιστοσύνης 95% μεταξύ 0,892–0,905 (**Εικόνα 33**). Χρησιμοποιώντας προβλέψεις από το σετ δοκιμής, τα 2 σημεία με την καλύτερη απόδοση ευαισθησίας/ειδικότητας είναι αυτά με 87,30%/71,98% και με 85,76% /79,18%. Η PPV (θετική προγνωστική τιμή) μιας διάγνωσης με COVID-19 έναντι ευαισθησίας, υπολογίστηκε auPRC (περιοχή κάτω η καμπύλη ακριβείας-ανάκλησης) ίση με 0,66 με όριο εμπιστοσύνης 95% μεταξύ 0,647–0,678 (**Εικόνα 34**).



Εικόνα 33: Καμπύλη Receiver Operating Characteristic [389]



Εικόνα 34: Καμπύλη Precision-Recall [390]

Η κατάταξη των πιο σημαντικών χαρακτηριστικών του μοντέλου έδειξε πως η εμφάνιση πυρετού, βήχα και η στενή επαφή με ένα άτομο που επιβεβαιώθηκε ότι έχει COVID-19 ήταν τα σημαντικότερα χαρακτηριστικά, επιβεβαιώνοντας ότι η σημασία της κοινωνικής αποστασιοποίησης. Επιπλέον το ανδρικό φύλο αποκαλύφθηκε ως προγνωστικός παράγοντας θετικού αποτελέσματος από το μοντέλο. Έτσι σύμφωνα με τα παραπάνω δεδομένα το μοντέλο μπορεί να εφαρμοστεί πιαγκοσμίως για αποτελεσματικό έλεγχο και ιεράρχηση των δοκιμών για την ιού στον γενικό πληθυσμό.

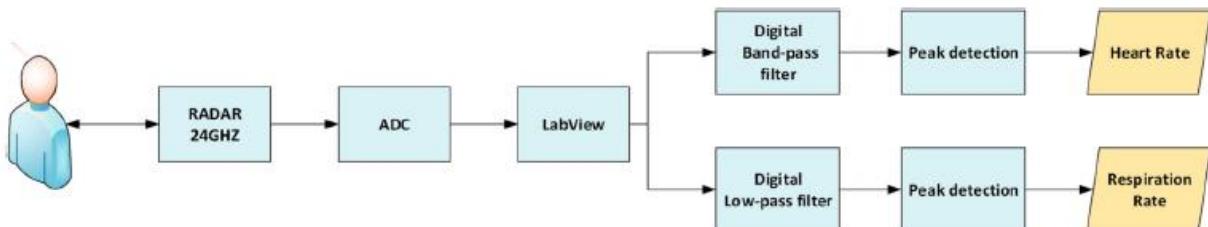
Trong Thanh Hana et al. (May 2021)

Στη μελέτη της Trong Thanh Hana και των συνεργατών της [392], με σκοπό την πρόβλεψη ασθενών με COVID-19 με τη χρήση πέντε αλγορίθμων μηχανικής μάθησης που λαμβάνουν δεδομένα μέσω ενός συστήματος μέτρησης το οποίο χρησιμοποιεί ιατρικό ραντάρ για τη λήψη δεδομένων, όπως οι καρδιακοί παλμοί και ο ρυθμός της αναπνοής. Πιο συγκεκριμένα χρησιμοποιήθηκε το Ιατρικό Ραντάρ NJR4262J (24 GHz, Ιαπωνία) της **Εικόνας 35** για τη συλλογή 500 δειγμάτων (**Πίνακας 3**) των ζωτικών σημείων των Καρδιακών Παλμών (Heart Rate, HR) και των Αναπνευστικών Ρυθμών (Respiratory Rates, RR), από τα οποία τα 140 ανήκαν σε υγιή άτομα που ήταν φοιτητές/τριες του Πανεπιστημίου Επιστήμης και Τεχνολογίας του Ανόι (HUST), ηλικίας 18–22 ετών και χωρίς σημάδια ασθένειας κατά την περίοδο μέτρησης και τα 360 ανήκαν σε άτομα που είχαν μολυνθεί από τον SARS-CoV-2 και ήταν ασθενείς στο Εθνικό Νοσοκομείο Τροπικών Ασθενειών (NHTD).

No.	Age	Sex	Reference			Infection Screening System			Diagnosis
			T-Body []	HR [bpm]	SpO2 [%]	RR [bpm]	HR [bpm]	SpO2 [%]	
1	21	M	37.4	85	90	10.795	83.057	95	Normal
2	21	M	37	74	99	27.608	74.621	96	Normal
3	21	M	37.1	87	99	32.117	84.914	98	Normal
4	21	M	36.7	76	98	11.309	72.081	97	Normal
5	19	M	37	72	99	28.778	75.025	97	Normal
6	22	M	36.7	84	99	30.908	81.918	98	Normal
7	20	M	36.9	69	90	26.793	73.364	96	Normal
8	29	M	36.9	86	98	12.497	88.072	98	Normal
9	21	F	37.7	83	99	24.547	68.069	98	Normal
10	20	F	37.2	94	99	14.144	89.115	98	Normal
11	29	F	40.2	125	99	35.157	124	98	Fever
12	30	F	37.5	122	90	25.292	116	96	Fever
13	27	F	37.7	105	98	25.437	98	97	Fever
14	29	M	39	124	97	29.866	118	95	Fever
15	28	F	38.5	121	98	21.008	120	95	Fever
16	22	M	39	97	97	24.751	92	96	Fever
17	38	M	37.7	102	98	22.595	99	96	Fever
18	29	M	39.6	98	98	28.5	95	98	Fever
19	39	M	38.1	110	98	24.025	106	97	Fever
20	25	F	39.7	112	96	19.831	117	97	Fever
...

Πίνακας 3: Μερικά από τα δείγματα του σύνολο δεδομένων [393]

Οι Αλγόριθμοι ML που χρησιμοποιήθηκαν ήταν ο Ταξινομητής Naïve Bayes (NBC), οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), τα Δέντρα Απόφασης (DT), ή Λογιστική Παλινδρόμηση (LR) και ο αλγόριθμος Βαθιάς Μάθησης (Deep Learning) Μακρο-Βραχυπρόθεσμης Μνήμης (LSTM). Τέλος όλα τα ληφθέντα αποτελέσματα συγκρίθηκαν ώστε να αξιολογηθεί η αποτελεσματικότητα της κάθε μεθόδου.



Εικόνα 35: Το Σύστημα NJR4262J που χρησιμοποιήθηκε για τη μέτρηση των ζωτικών σημείων [394]

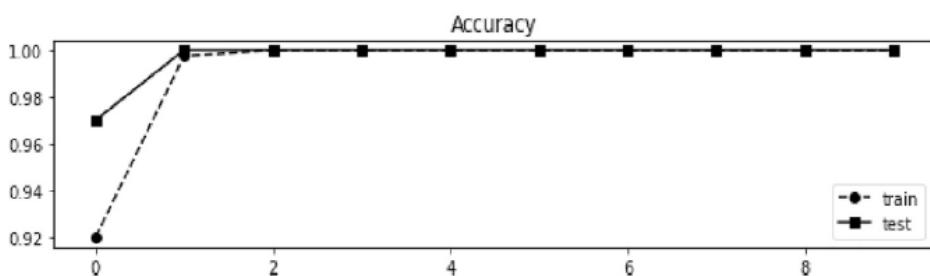
Από τα αποτελέσματα (Πίνακας 4) παρατηρούμε ότι ο αλγόριθμος των Δέντρων Απόφασης (DT) εμφανίζει τα υψηλότερα αποτελέσματα στα βήματα εκπαίδευσης, πρόβλεψης και ανάκλησης, με ποσοστά 88,07%, 85,29% και 85,15% αντίστοιχα, μετά έχουμε την Λογιστική Παλινδρόμηση (LR), έπειτα τον ταξινομητή Naive Bayes (NBC), και τελευταίο τον αλγόριθμο των Μηχανών Διανυσμάτων Υποστήριξης (SVM). Στην ίδια γραμμή κινούνται και τα αποτελέσματα για το F1 σκορ με τα DTs να πετυχαίνουν το υψηλότερο αποτέλεσμα 85,15% (έως 85,8%) και να ακολουθούν η LR με 84,71%, τον NBC με 83,12% και τελευταίο πάλι τον αλγόριθμο SVM με 80,62%.

Με βάση την διάρκεια της εκπαίδευσης και της δοκιμής (test), ο NBC είχε το χαμηλότερο χρόνο εκπαίδευσης στα 0,66ms και ο αλγόριθμος της LR είχε το χαμηλότερο χρόνο δοκιμής στα 0,16ms. Τα DTs πέτυχαν τον 2^o καλύτερο χρόνο εκπαίδευσης στα 1,22ms και τον 4^o χρόνο δοκιμής στα 0,22ms αλλά οι χρόνοι δοκιμής είχαν πολύ μικρή διακύμανση (0,16ms-0,22ms). Επομένως από τους 4 αλγόριθμους που χρησιμοποιήθηκαν ο αλγόριθμος των DTs αποτελεί την βέλτιστη μέθοδο με την μεγαλύτερη ακρίβεια.

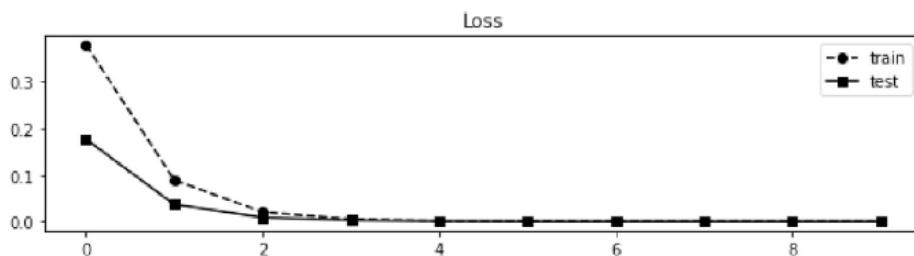
	Train accuracy	Training times	Test prediction	Test times	Test Recall	F1 score
Naïve Bayes	83.42%	0.66 ms	84.25%	0.20 ms	83.12%	83.57%
SVM	80.16%	4.16 ms	80.26%	0.21 ms	80.62%	80.4%
Logistic Regression	85.4%	2.02 ms	84.73%	0.16 ms	84.71%	84.6%
Decision Tree	88.07%	1.22 ms	85.29%	0.22 ms	85.15%	85.0%

Πίνακας 4: Αποτελέσματα των 4 αλγορίθμων NBC, SVM, LR και DTs [395]

Το μοντέλο του αλγόριθμου LSTM, υλοποιείται σε 10 εποχές (epochs), με την κάθε εποχή να ορίζεται ως το χρονικό διάστημα κατά το οποίο το σύνολο των δεδομένων μεταβιβάζεται προς τους μπροστινούς και προς τους πίσω νευρώνες του δικτύου μόνο μία φορά και αξιολογείται από δύο συντελεστές, την ακρίβεια και την απώλεια. Τα αποτελέσματα της μοντελοποίησης του LSTM για την ακρίβεια απεικονίζεται στην **Εικόνα 36** με το διάγραμμα για την απώλεια να απεικονίζεται στην **Εικόνα 37**, επίσης στον **Πίνακα 5** παρατηρούμε πως ο αλγόριθμος LSTM πέτυχε ακρίβεια σχεδόν 98% με 2% απώλεια στην πρώτη epoch.



Εικόνα 36: Καμπύλη ακρίβειας του μοντέλου LSTM [396]



Εικόνα 37: Καμπύλη απώλειας του μοντέλου LSTM [397]

	Loss	Accuracy
Train	5%	95.5%
Test	2%	98%

Πίνακας 5: Αποτελέσματα του αλγορίθμου LSTM [398]

Με την αξιολόγηση τεσσάρων παραγόντων, όπως η ακρίβεια εκπαίδευσης, η πρόβλεψη δοκιμής, η ανάκληση δοκιμής και το F1 σκορ μεταξύ και των 5 αλγορίθμων, ο LSTM έχει τις καλύτερες επιδόσεις εμφανίζοντας ακρίβεια μέχρι και 98%. Λόγω του πιο περίπλοκου υπολογισμού της διαδικασίας μάθησης και δοκιμής, τα αποτελέσματα ταξινόμησης του αλγορίθμου LSTM είναι πολύ πιο ακριβή ενώ τα σφάλματα είναι μικρότερα σε σχέση με τους υπόλοιπους αλγόριθμους. Εάν τα σύνολα δεδομένων είναι μεγάλα με περισσότερα χαρακτηριστικά, ο LSTM εμφανίζει καλύτερη απόδοση σε σχέση με τους άλλους 4, καθώς αυτοί είναι πιο επιρρεπείς στην υπερπροσαρμογή (overfitting).

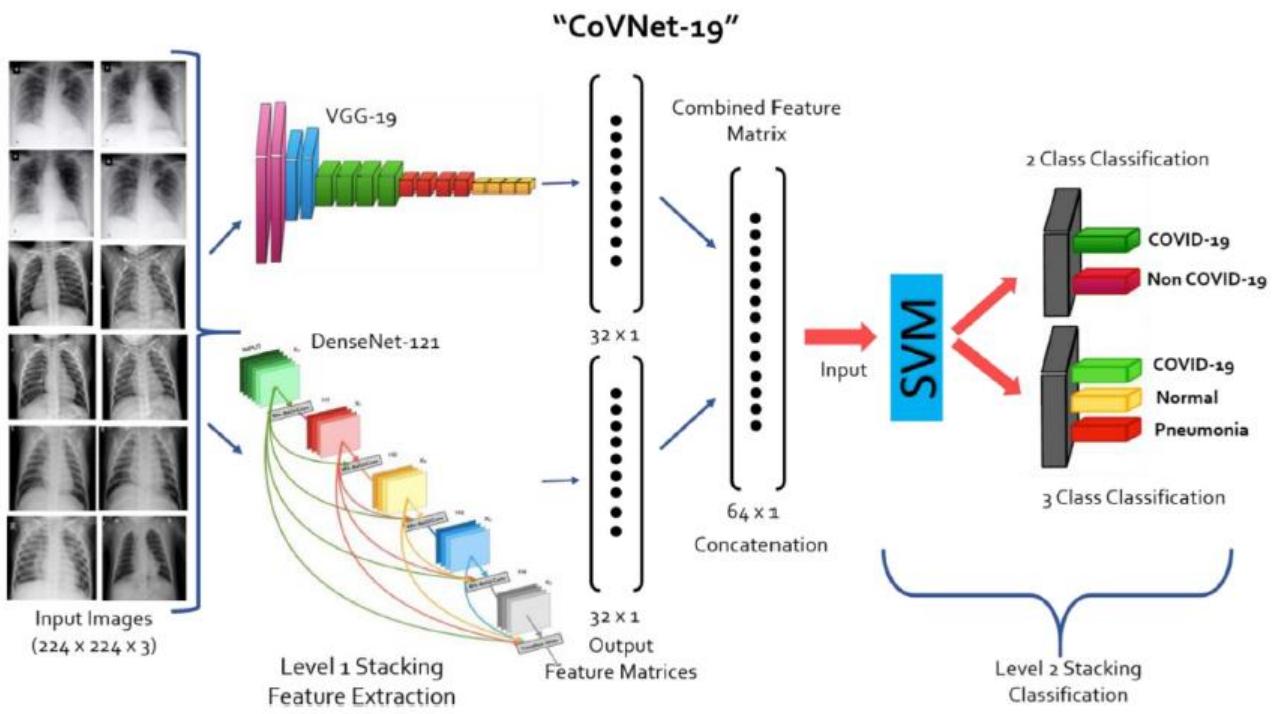
3.1.2 Με τη Χρήση Ακτινολογικών Δεδομένων

Τα συστήματα ιατρικής απεικόνισης όπως οι ακτινογραφίες (X-rays) και οι αξονικές τομογραφίες (CTs), μαζί με την βοήθεια της ML θα μπορούσαν να βοηθήσουν στην έγκαιρη διάγνωση των ασθενών με COVID-19. Οι διαγνωστικές απεικονιστικές μέθοδοι, όπως η ακτινογραφία θώρακος και η αξονική τομογραφία διαδραματίζουν σημαντικό ρόλο στην επιβεβαίωση της πρωτογενούς διάγνωσης. Η εξαγωγή χαρακτηριστικών από τις μεθόδους ακτινολογίας είναι ένα ουσιαστικό βήμα στην εκπαίδευση μοντέλων μηχανικής εκμάθησης, καθώς η απόδοση του μοντέλου εξαρτάται άμεσα από την ποιότητα των εξαγόμενων χαρακτηριστικών. Παρακάτω ακολουθούν ορισμένες μελέτες που κατασκευάστηκαν μοντέλα πρόβλεψης μέσω ML, χρησιμοποιώντας ως δεδομένα εικόνες από ακτινογραφίες ή αξονικές τομογραφίες.

Priyansh Kedia et al. (February 2021)

Η μελέτη του *Priyansh Kedia* και των συνεργατών του [399] είχε ως στόχο τη διαμόρφωση μιας ακριβούς, γρήγορης και αποτελεσματικής μεθόδου διάγνωσης της COVID-19 με τη χρήση των ακτινογραφιών (X-rays). Σε αυτή τη μελέτη χρησιμοποιήθηκε ένα σύνολο Βαθιά Συνελικτικών Νευρωνικών Δικτύων (ΒΣΝΔ) με την ονομασία **CoVNet-19**. Ο κύριος στόχος της χρήσης ΒΣΝΔ ήταν η εξαγωγή βασικών και πολύτιμων χαρακτηριστικών από δεδομένα εικόνας μέσω μοντέλου Μηχανικής Μάθησης Συνόλων (Ensemble ML).

Το **CoVNet-19** (Εικόνα 38) αποτελείται από 2 επίπεδα και συνδυάζει πολλαπλά μοντέλα ταξινόμησης ώστε να μπορεί να ερμηνεύσει τα ίδια δεδομένα με διαφορετικούς τρόπους. Στο 1^o επίπεδο του μοντέλου CoVNet-19, συνδυάστηκαν δύο μοντέλα ΒΣΝΔ, τα VGG19 [400] και DenseNet121 [401], τα οποία εισήχθησαν με προ-εκπαίδευμένους πίνακες βαρών. Η επιλογή των 2 αυτών μοντέλων έγινε λόγω της εξαιρετικά καλής απόδοσής στην ταξινόμηση ακτινογραφιών. Η εκπαίδευση των 2 ΒΣΝΔ έγινε χρησιμοποιώντας ως δεδομένα εισόδου εικόνες ακτινογραφιών θώρακος με διαστάσεις 224x224x3 της ΒΔ ImageNet, η οποία περιέχει περισσότερες από 20 χιλιάδες κατηγορίες, με περισσότερες από 14 εκατομμύρια εικόνες και που αποδειγμένα δίνει μοντέλα με εξαιρετικά επιτυχημένα αποτελέσματα στην ταξινόμηση εικόνων [402]. Στο 2^o επίπεδο του CoVNet-19, χρησιμοποιήθηκε ένα μοντέλο Μηχανών Διανυσμάτων Υποστήριξης (SVM), το οποίο και εκπαιδεύτηκε πάνω στα χαρακτηριστικά τα οποία εξήχθησαν από το μοντέλο του 1^{ου} επιπέδου. στο να εκτελεί πολυταξική και δυαδική ταξινόμηση. Εκπαιδεύτηκε στο χαρακτηριστικά που εξάγονται από το βασικό μοντέλο. Το κίνητρο για τη χρήση του μοντέλου της Μηχανικής Μάθησης Συνόλων (Ensemble ML) ήταν το να επιτραπεί καλύτερη απόδοση στην πρόγνωση, σε σύγκριση με ένα μεμονωμένο μοντέλο ταξινόμησης ΒΣΝΔ, με τα δύο ΒΣΝΔ να εκπαιδεύονται ξεχωριστά στα δεδομένα της ImageNet.



Εικόνα 38: Η Αρχιτεκτονική του μοντέλου CoVNet-19 [403]

Τα αποτελέσματα έδειξαν ότι το μοντέλο ταξινόμησης CoVNet-19 μπορούσε να διακρίνει και να ταξινομήσει τις ακτινογραφίες ασθενών με COVID-19, ασθενών με πνευμονία και υγειών ατόμων, με μεγάλη ακρίβεια. Τα αποτελέσματα υποδηλώνουν ότι τα ΒΣΝΔ VGG19 και DenseNet121 μαζί βοήθησαν να επιτύχουμε την καλύτερη ακρίβεια ταξινόμησης εξάγοντας πολύτιμα εξελιγμένα χαρακτηριστικά. Ως εκ τούτου, μια αυτοματοποιημένη ένταξη Deep Learning για τη διάγνωση του COVID-19 με χρήση μόνο ακτινογραφίας θώρακα μπορεί να είναι επωφελής για τα ιατρικά ιδρύματα. Το μοντέλο CoVNet-19 πέτυχε υψηλή συνολική ακρίβεια **98,28%** στην ταξινόμηση **3 κατηγοριών** (COVID-19/πνευμονία/υγείες) και συνολική ακρίβεια **99,71%** στην ταξινόμηση **2 κατηγοριών** (COVID-19/non COVID-19), έχοντας 3 ψευδώς αρνητικά (Σφάλμα τύπου 2) και 3 ψευδώς θετικά (σφάλμα τύπου 1) στην τριμερή ταξινόμηση και εμφανίζοντας 2 ψευδώς αρνητικά και 1 ψευδώς θετικό στη δυαδική ταξινόμηση η οποία περιείχε **276** αληθινά θετικά περιστατικά ασθενών COVID-19. Το F_1 σκορ του και στις 2 περιπτώσεις ήταν **99%**. Παρακάτω απεικονίζονται οι Πίνακες Σύγχυσης των 2 (Εικόνα 39) και των 3^{ων} κατηγοριών (Εικόνα 40)

		Predicted Class		True Class
		COVID-19	Non COVID-19	
COVID-19	COVID-19	274	2	
	Non COVID-19	1	775	

Εικόνα 39: Πίνακας σύγχυσης 2 κατηγοριών [404]

		Predicted Class			True Class
		COVID-19	1	2	
COVID-19	COVID-19	273	1	2	
	NORMAL	0	992	8	
	PNEUMONIA	3	25	972	

Εικόνα 40: Πίνακας σύγχυσης 3 κατηγοριών [405]

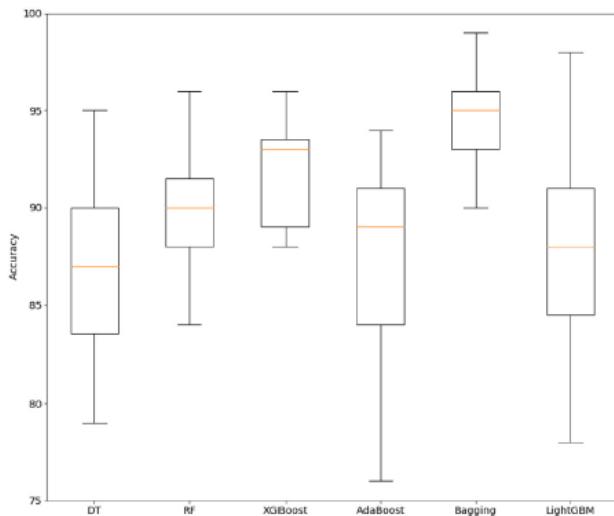
Με εκατοντάδες χιλιάδες ανθρώπους να μολύνονται καθημερινά ταυτόχρονα περιορισμένη διαθεσιμότητα κιν δοκιμών και άλλων ιατρικών πόρων, όπως συμβαίνει σε απομακρυσμένα μέρη, το CoVNet-19 θα μπορούσε να παρέχει σημαντική βοήθεια και υποστήριξη στους ιατρούς και το νοσηλευτικό προσωπικό για την αντιμετώπιση εκθετικά αυξανόμενων περιπτώσεων COVID-19 μέσω της αυτόματης και ταχύτερης διάγνωσης ασθενών.

Sara Hosseinzadeh Kassania et al. (June 2021)

Η μελέτη της Sara Hosseinzadeh Kassania και των συνεργατών της [406] είχε ως σκοπό την αυτόματη ανίχνευση της COVID-19 σε ακτινογραφίες (X-rays) και Αξονικές Τομογραφίες (CTs) με τη χρήση της ML και πιο συγκεκριμένα των Βαθιά Συνελικτικών Νευρωνικών Δικτύων (ΒΣΝΔ) MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, NASNet καθώς και ταξινομητών όπως τα Δέντρα Απόφασης (ΔΑ) [407], Random Forest (RF) [408], XGBoost [384], AdaBoost [409], Bagging classifier [410] και LightGBM [411]. Η διαδικασία που ακολούθησαν ήταν η εξής. Αρχικά συλλέχθηκαν διαθέσιμες εικόνες από τις ακτινογραφίες θώρακος και τις αξονικές τομογραφίες. Έπειτα το παρεχόμενο σύνολο των δεδομένων προεπεξεργάστηκε χρησιμοποιώντας τυπικές τεχνικές κανονικοποίησης εικόνας με σκοπό τη βελτίωση της ποιότητας των οπτικών πληροφοριών των δεδομένων εισόδου. Κατόπιν οι προεπεξεργασμένες εικόνες χρησιμοποιήθηκαν ως είσοδοι στα ΒΣΝΔ ώστε να εξαχθούν τα βαθέα χαρακτηριστικά τις κάθε εικόνας. Στη συνέχεια τα χαρακτηριστικά που εξήχθησαν από την προηγούμενη φάση τροφοδοτήθηκαν στους ταξινομητές. Τέλος χρησιμοποιήθηκε μια Διασταυρούμενη Επικύρωση 10 τμημάτων για την αξιολόγηση της μέσης απόδοσης γενίκευσης των ταξινομητών κάθε πειράματος αφού ταξινομήθηκαν δοκιμαστικές (άγνωστες) εικόνες απεικονίσεων.

Για όλα τα ΒΣΝΔ τα βάρη δικτύου αρχικοποιήθηκαν από τα βάρη που εκπαιδεύτηκαν στο ImageNet. Το σύστημα υπολογιστή βασισμένο στα Windows που χρησιμοποιήθηκε για αυτήν την εργασία είχε επεξεργαστή Intel Core i7-8700 K 3,7 GHz με 32 GB RAM. Η διαδικασία εκπαίδευσης και δοκιμής της προτεινόμενης αρχιτεκτονικής για αυτό το πείραμα εφαρμόστηκε στην Python χρησιμοποιώντας το πακέτο Keras με το backend Tensorflow ως backend πλαισίου βαθιάς εκμάθησης και εκτελείται σε Nvidia GeForce GTX 1080 Ti GPU με 11 GB RAM.

Ο **Πίνακας 6** και η **Εικόνα 41** συνοψίζουν την απόδοση ακρίβειας έξι αλγορίθμων μηχανικής εκμάθησης, συγκεκριμένα, DT, RF, XGBoost, AdaBoost, Bagging classifier και LightGBM στο χαρακτηριστικό που εξάγεται από ΒΣΝΔ. Κάθε καταχώρηση στον **Πίνακα 6**, έχει τη μορφή (μ±σ) όπου μ είναι η μέση ακρίβεια ταξινόμησης και σ είναι τυπική απόκλιση. Σύμφωνα με τον πίνακα το κορυφαίο αποτέλεσμα λήφθηκε από τον ταξινομητή **Bagging** με μέγιστη ακρίβεια **99,00% ± 0,09** σε χαρακτηριστικά που εξήχθησαν από την αρχιτεκτονική **DenseNet121**, με χρόνο εξαγωγής χαρακτηριστικών **9,306** δευτερόλεπτα και χρόνο εκπαίδευσης **30,748** δευτερόλεπτα, που είναι το υψηλότερο αποτέλεσμα για αυτό το συνόλου δεδομένων. Από τον **Πίνακα 6** συμπεραίνεται επίσης ότι το δεύτερο καλύτερο αποτέλεσμα που ελήφθη από τον εξαγωγέα χαρακτηριστικών **ResNet50** και τον ταξινομητή **LightGBM**, με χρόνο εξαγωγής χαρακτηριστικών **10,206** δευτερόλεπτα, χρόνο εκπαίδευσης **0,960** δευτερόλεπτα και συνολική ακρίβεια **98,00 ± 0,09**. Συγκρίνοντας 2 καλύτερους συνδυασμούς, παρατηρείται πως ενώ η ακρίβεια ταξινόμησης του DenseNet121 με Bagging είναι ελαφρώς καλύτερη (1%) από το ResNet50 με LightGBM, ο χρόνος εκπαίδευσης του δεύτερου είναι αρκετά δελεαστικός, σχεδόν 30 φορές καλύτερος από τον πρώτο νικητή από άποψη της ακρίβειας. Αν και το Bagging μαθαίνει αργά, έχει τη χαμηλότερη τυπική απόκλιση και ως εκ τούτου είναι πιο σταθερό από τους άλλους ταξινομητές.



Εικόνα 41: Η Απόδοση της ακρίβειας (accuracy) των διαφορετικών ταξινομητών ML στην δοκιμασία ανίχνευσης COVID-19 [412]

	Decision Tree	Random Forest	XGBoost	AdaBoost	Bagging	LightGBM
MobileNet	83.00 ± 0.26	93.00 ± 0.23	95.00 ± 0.16	80.00 ± 0.17	96.00 ± 0.11	82.00 ± 0.28
DenseNet121	92.00 ± 0.15	90.00 ± 0.21	94.00 ± 0.16	92.00 ± 0.19	99.00 ± 0.07	96.00 ± 0.11
DenseNet201	84.00 ± 0.26	90.00 ± 0.24	90.00 ± 0.18	87.00 ± 0.25	96.00 ± 0.11	87.00 ± 0.17
Xception	95.00 ± 0.17	90.00 ± 0.19	96.00 ± 0.11	93.00 ± 0.20	96.00 ± 0.11	96.00 ± 0.11
InceptionV3	82.00 ± 0.22	84.00 ± 0.29	88.00 ± 0.15	80.00 ± 0.12	95.00 ± 0.12	84.00 ± 0.16
InceptionResNetV2	84.00 ± 0.31	93.00 ± 0.16	93.00 ± 0.19	87.00 ± 0.33	94.00 ± 0.12	88.00 ± 0.21
ResNet50	89.00 ± 0.17	90.00 ± 0.15	93.00 ± 0.16	94.00 ± 0.12	93.00 ± 0.16	98.00 ± 0.09
ResNet152	93.00 ± 0.12	92.00 ± 0.16	93.00 ± 0.16	94.00 ± 0.17	91.00 ± 0.22	93.00 ± 0.20
VGG16	90.00 ± 0.19	91.00 ± 0.19	88.00 ± 0.19	90.00 ± 0.19	90.00 ± 0.19	85.00 ± 0.19
VGG19	90.00 ± 0.19	87.00 ± 0.21	88.00 ± 0.19	90.00 ± 0.19	90.00 ± 0.19	85.00 ± 0.25
NASNetLarge	82.00 ± 0.23	88.00 ± 0.19	89.00 ± 0.17	81.00 ± 0.23	93.00 ± 0.19	82.00 ± 0.26
NASNetMobile	87.00 ± 0.17	88.00 ± 0.22	94.00 ± 0.19	87.00 ± 0.17	93.00 ± 0.19	89.00 ± 0.17
ResNet50V2	87.00 ± 0.12	96.00 ± 0.11	92.00 ± 0.19	90.00 ± 0.18	95.00 ± 0.12	88.00 ± 0.10
ResNet101V2	79.00 ± 0.32	89.00 ± 0.24	89.00 ± 0.28	76.00 ± 0.32	95.00 ± 0.12	78.00 ± 0.26
ResNet152V2	90.00 ± 0.27	86.00 ± 0.26	93.00 ± 0.16	89.00 ± 0.20	96.00 ± 0.11	88.00 ± 0.28

Πίνακας 6: Σύγκριση απόδοσης ταξινόμησης (μ \pm σ) των διαφορετικών μοντέλων μηχανικής εκμάθησης που μετράται με ακρίβεια. Η έντονη τιμή υποδηλώνει το καλύτερο αποτέλεσμα. Η υπογραμμισμένη τιμή αντιπροσωπεύει το δεύτερο καλύτερο αποτέλεσμα της αντίστοιχης κατηγορίας. [413]

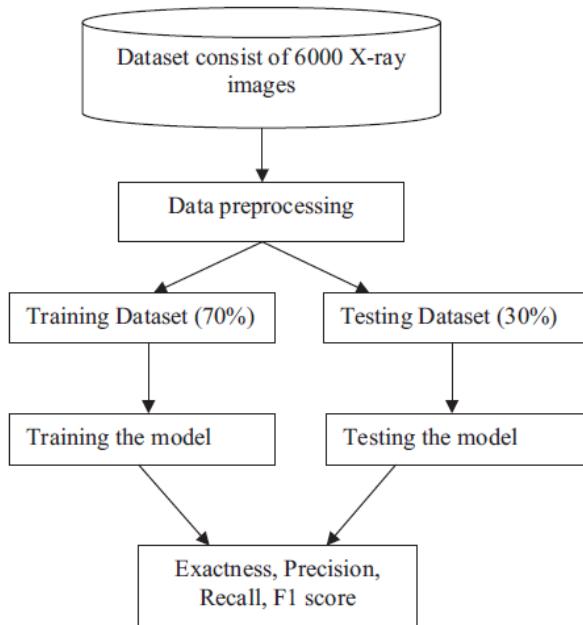
Αυτή η μελέτη παρουσιάζει μια προσέγγιση ταξινόμησης βαθιάς μηχανικής μάθησης βασισμένη σε εξαγωγείς χαρακτηριστικών χρησιμοποιώντας ως δεδομένα ακτινογραφίες θώρακος και αξονικές τομογραφίες, για τη διάγνωση της COVID-19. Λαμβάνοντας υπόψη την υψηλή οπτική πολυπλοκότητα των δεδομένων εικόνας, η σωστή εξαγωγή χαρακτηριστικών σε βάθος θεωρείται κρίσιμο βήμα για την ανάπτυξη μοντέλων ΒΣΝΔ. Τα πειραματικά αποτελέσματα καταδεικνύουν ότι τα χαρακτηριστικά που εξάγονται από το ΒΣΝΔ **DenseNet121** και εκπαιδεύονται από έναν ταξινομητή **Bagging** παράγουν μια πολύ ακριβή πρόβλεψη **99,00%** όσον αφορά την ακρίβεια ταξινόμησης.

Mohit Kumar et al. (December 2021)

Στη μελέτη του Mohit Kumar και των συνεργατών του [414], όπου και εδώ ο σκοπός ήταν η αυτόματη ανίχνευση της COVID-19 σε ακτινογραφίες θώρακος (X-rays) μέσω της ML, εδώ χρησιμοποιήθηκε μια νέα προσέγγιση, όπως το Υβριδικό Συνελικτικό Νευρωνικό Δίκτυο (Hybrid Convolutional Neural Network, HDCNN), το οποίο ενσωματώνει το Συνελικτικό Νευρωνικό Δίκτυο (CNN) και την Αρχιτεκτονική Επαναλαμβανόμενου Νευρωνικού Δικτύου (Recurrent Neural Network, RNN) για την ταυτοποίηση των ασθενών με COVID-19 [415].

Για την βελτίωση της ακρίβειας του HDCNN χρησιμοποιήθηκε η προσέγγιση μεταβίβασης της μάθησης της Χαρτογράφησης Ενεργοποίησης Κλάσεων-με στάθμιση κλίσης (Grad-CAMs), για την εμφάνιση εικόνων που είναι υπεύθυνες για τη λήψη αποφάσεων.

Το σύνολο δεδομένων των εικόνων των ακτινογραφιών θώρακος ασθενών COVID-19 που χρησιμοποιήθηκε για την εκπαίδευση και τη δοκιμή του HDCNN συλλέχθηκε μέσω διάφορων πηγών όπως το Github, το Kaggle, το Mendeley. Συνολικά συγκεντρώθηκαν 18.000 δείγματα COVID-19, πνευμονίας και υγειών ατόμων, τα οποία χωρίστηκαν σε 3 σετ από 6.000 δείγματα το καθένα. Για την εκπαίδευση χρησιμοποιήθηκαν το 70% των δειγμάτων του κάθε σετ, με το υπόλοιπο 30% να χρησιμοποιείται ως σύνολο δεδομένων δοκιμής όπως απεικονίζεται στην **Εικόνα 42**.



Εικόνα 42: Γενικό διάγραμμα ροής της χρήσης των datasets των ακτινογραφιών στο HDCNN [416]

Περιληπτικά στο μοντέλο του HDCNN, το CNN είναι υπεύθυνο για εξαγωγή των χαρακτηριστικών και για την δειγματοληψία, με τα εξαγόμενα δεδομένα τροφοδοτούνται στο RNN και τέλος χρησιμοποιείται η προσέγγιση μεταβίβασης μάθησης Grad- CAM για την πρόβλεψη της τάξης της εκάστοτε εικόνας.

Το Συνελικτικό Νευρωνικό Δίκτυο (CNN) δέχεται ως είσοδο το σύνολο δεδομένων με τη μορφή εικόνων ακτινογραφιών θώρακα, την οποία διαχειρίζονται 2 στρώματα συνέλιξης, 2 στρώματα συγκέντρωσης (pooling layers), δύο συσχετιζόμενα στρώματα και μια συνάρτηση Softmax. Το CNN αποτελείται από τρία κύρια επίπεδα, το επίπεδο που εκτελείται η συνέλιξη, ένα ενδιάμεσο επίπεδο που εκτελούνται οι διαδικασίες συγκέντρωσης (pooling), και τα πυκνά στρώματα (dense layers).

Το RNN σχεδιάστηκε για να ταξινομεί τα πρωτογενή δεδομένα των σημάτων εισόδου, χωρίζεται σε 700 βήματα και 64 νευρώνες ανά βήμα. Οι έξοδοι του RNN στα τελευταία 20 βήματα έχουν ενσωματωθεί σε πλήρως συνδεδεμένα στρώματα ώστε να ταξινομηθούν. Στο Επαναληπτικό Νευρωνικό Δίκτυο (RNN) οι νευρώνες μεταφέρουν «κριτική» σε ένα παρόμοιο κρυψό επίπεδο, το οποίο λαμβάνει υπόψιν την απόδοση από προηγούμενες εκτελέσεις, παίζοντας έτσι το ρόλο της μνήμης του RNN. Το RNN περιλαμβάνει επίσης ένα επίπεδο Gated Recurrent Units (GRU), για την εγγραφή σημάτων εισόδου από προηγούμενα δεδομένα, με τη μορφή χρονοσειρών.

Η Χαρτογράφηση Ενεργοποίηση Κλάσεων-με στάθμιση κλίσης (Grad-CAMs) επεξεργάζεται τις εξόδους του RNN για τη δημιουργία ενός πρόχειρου χάρτη εντοπισμού των βασικών

περιοχών της εκάστοτε εικόνας. Η Grad-CAM δεν απαιτεί επανεκπαίδευση και γενικά ισχύει για όλες τις αρχιτεκτονικές που βασίζονται σε CNN και μπορεί να συνδυάσει υπάρχουσες απεικονίσεις με τα rīxel της εικόνας για τη δημιουργία μιας απεικόνισης υψηλών στόχων με κατηγορικές διακρίσεις (Καθοδηγούμενη Grad-CAM). Τέλος πραγματοποιήθηκαν συνεντεύξεις για να ποσοτικοποιηθεί το αν οι διευκρινίσεις CAM με καθοδηγούμενο βαθμό βοηθούν τους ειδικούς στο να βασίζονται στις προβλέψεις του HDCNN.

Η υβριδική προσέγγιση HDCNN υλοποιήθηκε χρησιμοποιώντας το Anaconda σε περιβάλλον Python 3. Τα όρια που χρησιμοποιήθηκαν για την αξιολόγηση του προτεινόμενου μοντέλου είναι, η ακρίβεια (accuracy & precision), η ανάκληση (recall) και το F1 σκορ. Στη σύγκριση των υπαρχόντων μοντέλων CNN με το HDCC παρατηρείται ότι το HDCNN έχει ξεπεράσει τα άλλα υπάρχοντα μοντέλα (**Πίνακας 7**). Πιο συγκεκριμένα το HDCNN συγκρινόμενο με άλλα CNN όπως το Inception-v3, το ShuffleNet, το SqueezeNet, το VGG-19 και το DenseNet πέτυχε accuracy 98,20%, precision 97,31%, ανάκληση 97,1% και F1 σκορ 0,97, πετυχαίνοντας έτσι καλύτερα αποτελέσματα δείχνοντας έτσι πως μπορεί να χρησιμοποιηθεί για την διάγνωση μετά και από τις κατάλληλες εγκρίσεις.

Model	Accuracy	Precision	Recall	F1 score
Inception-v3	93.62	96.20	90.71	0.94
ShuffleNet	95.97	95.44	96.57	0.96
SqueezeNet	87.52	86.84	88.29	0.88
VGG-19	90.16	87.34	93.33	0.90
DenseNet	96.20	95.78	96.67	0.96
HDCNN	98.20	97.31	97.1	0.97

Πίνακας 7: Σύγκριση άλλων μοντέλων CNN με HDCNN [417]

3.2 Πρόβλεψη της Σοβαρότητας της Νόσησης και της Θνησιμότητας ασθενών με COVID-19

Τα κλινικά χαρακτηριστικά των ασθενών μαζί με την βοήθεια της ML θα μπορούσαν να βοηθήσουν στην πρόγνωση της εξέλιξης της νόσου με τον έγκαιρο εντοπισμό των ασθενών COVID-19 οι οποίοι να χρειαστούν εισαγωγή ή ακόμα και εισαγωγή σε ΜΕΘ μειώνοντας ενδεχομένως τα ποσοστά θνησιμότητας. Παρακάτω θα αναλύσουμε ορισμένες μελέτες που χρησιμοποίησαν βιοχημικές ενδείξεις και τα κλινικά χαρακτηριστικά ασθενών με COVID-19 με σκοπό την ακριβή πρόβλεψη της εξέλιξης της COVID-19

Alexandre de Fatima Cobre et al. (May 2021)

Αυτή η μελέτη του Alexandre de Fatima Cobre και των συνεργατών του [418], είχε ως στόχο να εφαρμόσει και να αξιολογήσει μοντέλα που βασίζονται σε μηχανική μάθηση για την πρόβλεψη του COVID-19. διάγνωση και σοβαρότητα της νόσου.

Το σύνολο των δεδομένων λήφθηκαν από τη δημόσια πλατφόρμα Kaggle [419] και περιλαμβάνει άτομα που εξετάστηκαν μέσω RT-PCR τεστ για το αν έχουν μολυνθεί ή όχι από τον SARS-CoV-2, στο νοσοκομείο «Αλμπερτ Αϊνστάιν» του Σάο Πάολο της Βραζιλίας.

Οι ασθενείς ανεξάρτητα με το αν βρέθηκαν υγείες ή ασθενείς με COVID-19 υποβλήθηκαν σε βιοχημικές, αιματολογικές και ουρολογικές εξετάσεις (**Πίνακας 8**). Τα δείγματα ομαδοποιήθηκαν με 2 διαφορετικούς τρόπους σύμφωνα με τα αποτελέσματα των τεστ RT-PCR. Στην 1^η ομαδοποίηση συμπεριλήφθηκαν όλα τα 5.643 δείγματα τα οποία χωρίστηκαν με βάση το τεστ που υποβλήθηκαν τα υπό εξέταση άτομα αν ήταν αρνητικό (n = 5.086) ή θετικό (n = 557). Στην 2^η ομαδοποίηση συμπεριλήφθηκαν μόνο τα 557 θετικά με COVID-19 δείγματα τα οποία χωρίστηκαν σε ασυμπτωματικούς ή με ήπια συμπτώματα που πήραν εξιτήριο και στους ασθενείς που εμφάνισαν σοβαρά συμπτώματα της COVID-19 και οι οποίοι χρειάστηκε απλά να νοσηλευτούν ή να εισέλθουν σε ΜΕΘ. Για την ελαχιστοποίηση της πιθανότητα ψευδώς θετικών δειγμάτων, καθώς η λοίμωξη SARS-CoV-2 μοιάζει με άλλες ασθένειες του αναπνευστικού, εξαιρέθηκαν από τις αναλύσεις οι ασθενείς που βρέθηκαν θετικοί σε τουλάχιστον έναν άλλο ιό ή βακτήριο του αναπνευστικού.

Biochemical tests

Glucose serum, urea, C-reactive protein, creatinine, potassium, sodium, alanine transaminase, aspartate transaminase, gamma-glutamyltransferase, total bilirubin, direct bilirubin, indirect bilirubin, alkaline phosphatase, ionised pH, blood, magnesium analysis, HCO₃ (venous blood gas analysis), lactate dehydrogenase, creatine phosphokinase, ferritin, arterial lactic acid, lipase dosage, HCO₃ (arterial blood gas analysis), phosphorus, pCO₂ (venous blood gas analysis), Hb saturation (venous blood gas analysis), base excess (venous blood gas analysis), pO₂ (venous blood gas analysis), total CO₂ (venous blood gas analysis), Hb saturation (arterial blood gases), pCO₂ (arterial blood gas analysis), base excess (arterial blood gas analysis), pH (arterial blood gas analysis), total CO₂ (arterial blood gas analysis), pO₂ (arterial blood gas analysis), arterial FiO₂, and ctO₂ (arterial blood gas analysis).

Haematological tests

Hematocrit, Hemoglobin, Platelets, Mean platelet volume, Red blood Cells, Lymphocytes, Mean corpuscular hemoglobin concentration, Leukocytes, Basophils, Mean corpuscular hemoglobin, Eosinophils, Mean corpuscular volume, Monocytes, Red blood cell distribution width

Urine tests

Urine pH, segmented neutrophil, promyelocytes, metamyelocytes and myeloblasts, and the international normalised ratio (INR).

Virological tests

Respiratory syncytial virus, influenza A, influenza B, parainfluenza 1, coronavirus NL63, rhinovirus/enterovirus, coronavirus HKU1, parainfluenza 3, adenovirus, parainfluenza 4, coronavirus 229E, coronavirus OC43, influenza A H1N1, influenza H1N1 test, and influenza A rapid test.

Bacteriological tests

Mycoplasma pneumoniae, *Chlamydophila pneumoniae*, and *Streptococcus A*.

Πίνακας 8: Οι Βιοχημικές, αναλύσεις ούρων, αιματολογικές, ιολογικές και βακτηριολογικές εξετάσεις των υπό εξέταση ατόμων που συμπεριλήφθηκαν στη μελέτη. [420]

Συνολικά χρησιμοποιήθηκαν τέσσερις αλγόριθμοι, τα Τεχνητά Νευρωνικά Δίκτυα (ANN), τα Δέντρα Απόφασης (DT), η Διακριτική Ανάλυση Μερικών Ελαχίστων Τετραγώνων (PLS-DA), η μέθοδος των Κ-Πλησιέστερων Γειτόνων (KNN). Για την εφαρμογή αυτών των μοντέλων με αλγόριθμους για τη διάγνωση και πρόβλεψη της σοβαρότητας της COVID-19, χρησιμοποιήθηκε το 70% των δειγμάτων ως σετ προπόνησης και 30% ως σετ δοκιμών, με την επιλογή των δείγματα για το κάθε σετ να πραγματοποιείται μέσω της μεθόδου Kennard-Stone [421].

Οι λανθάνουσες μεταβλητές (Latent Variables, LVs) που χρησιμοποιήθηκαν για τα μοντέλα των αλγορίθμων επιλέχθηκαν με τη μέθοδο της διασταυρούμενης επικύρωσης «άφησε ένα έξω» (Leave One Out Cross Validation, LOOCV). Η προγνωστική ικανότητα του μοντέλου αξιολογήθηκε χρησιμοποιώντας την μέθοδο της Τετραγωνικής Ρίζας του Μέσου Τετραγωνικού Σφάλματος της Πρόβλεψης (Root Mean Squared Error on Prediction, RMSEP), σύμφωνα με την οποία τα μοντέλα βελτιστοποιήθηκαν με στόχο το χαμηλότερο RMSEP. Οι αναλυτικές επιδόσεις των μοντέλων μετρήθηκαν χρησιμοποιώντας ως μετρικά την ευαισθησία, την ειδικότητα (περιοχή κάτω από την καμπύλη ROC [**Εικόνα 43**]), την ακρίβεια, το χρόνος εκπαίδευσης, το σφάλμα εκπαίδευσης μοντέλου και το σφάλμα

διασταυρούμενης επικύρωσης (**Πίνακας 9**). Στη μελέτη επίσης συμπεριλήφθηκαν τα δεδομένα από βιοχημικές, ουροποιητικές, βακτηριολογικές και ιολογικές εξετάσεις των εξεταζόμενων (ασθενών και υγειών) με στόχο τον εντοπισμό βιοδεικτών που σχετίζονται με την μόλυνση και την σοβαρή νόσηση με COVID-19.

Metric	Diagnostic model				Disease severity model			
	RNA	DT	PLS-DA	K-NN	RNA	DT	PLS-DA	K-NN
Training time	21 min. 43 s	27 min. 11 s	31 min. 19 s	22 min. 15 s	7 min. 1 s	10 min. 19 s	18 min. 3 s	09 min. 53 s
Calibration error	1.0%	0.5%	1.2%	0.5%	1.0%	0.4%	6.0%	0.4%
Cross validation error	0.8%	1.0%	0.9%	0.6%	0.5%	1.8%	4.0%	0.7%
Sensibility	0.93	0.89	0.88	0.84	0.99	0.90	0.87	0.82
Specificity	0.94	0.89	0.90	0.83	0.97	0.94	0.88	0.88
Accuracy ^a	0.94	0.90	0.90	0.84	0.98	0.92	0.88	0.86

^a Area under the ROC curve.

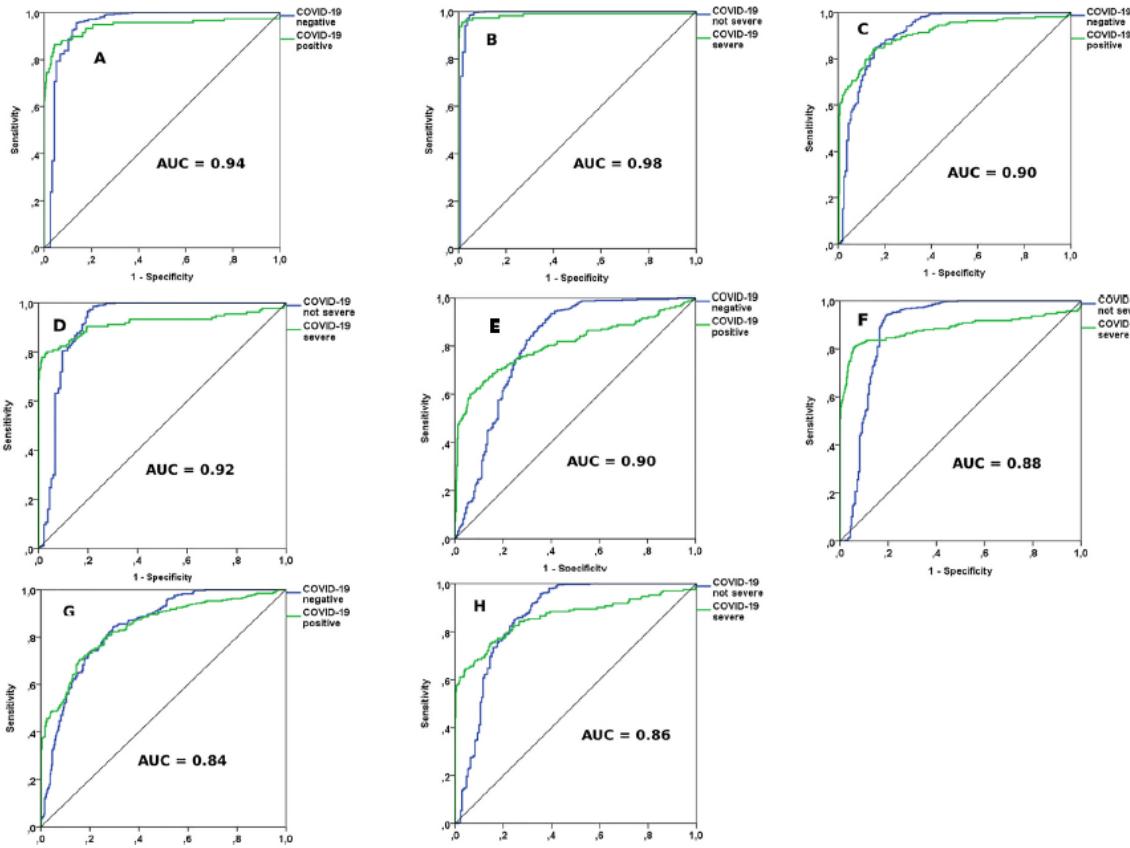
Πίνακας 9: Σύγκριση της απόδοσης των μοντέλων των αλγορίθμων που χρησιμοποιήθηκαν. [422]

Biomarkers	Diagnostic model				Disease severity model			
	ANN	DT	PLSDA	KNN	ANN	DT	PLSDA	KNN
Ferritin	++	++	++	++	++	++	++	++
Gamma-glutamyltransferase	+	-	-	-	-	-	-	-
HCO ₃ (arterial)	+	-	-	+	-	-	+	+
Base excess (arterial)	+	-	-	-	-	-	+	+
Base excess (venous)	-	-	-	-	-	-	+	-
Sodium	+	-	-	-	-	-	-	-
Total O ₂ (arterial)	-	-	-	-	+	-	-	-
pO ₂ (arterial)	-	-	-	+	-	-	-	-
Total CO ₂	-	+	+	+	-	-	-	+
pCO ₂ (arterial)	+	-	-	-	-	-	+	+
pCO ₂ (venous)	+	-	+	-	-	-	+	-
Indirect bilirubin	+	-	-	-	-	-	-	-
Alkaline phosphatase	+	-	-	-	+	-	-	-
Urine pH	-	-	+	+	+	-	+	-
pH (venous)	-	-	-	-	-	-	+	-
pH (arterial)	-	-	-	-	-	-	-	+
FiO ₂ (arterial)	-	-	-	-	+	-	-	-
ctO ₂ (arterial)	-	-	-	-	-	-	-	+
Total bilirubin	-	-	-	-	+	-	-	-
Red blood cell distribution width	-	-	-	-	+	-	-	-
Platelets	-	-	-	-	+	-	-	-
C-reactive protein	-	-	-	-	+	-	+	-
Calcium ionised	-	-	+	+	-	-	-	-
Urine-density	-	-	+	+	-	-	-	-
Lactate dehydrogenase	-	-	-	-	-	-	+	-
Arterial lactic acid	-	-	-	+	-	-	-	+
Haemoglobin saturation (arterial)	-	-	-	-	-	-	-	+
Phosphorous	-	-	+	+	-	-	-	-
Lipase dosage	-	-	-	-	-	-	+	-
Rods	-	-	-	+	-	-	+	-

Less important variable (-); Important variable (+); Critical variable (++)

Πίνακας 10: Απεικόνιση της σημαντικότητας των βιοδεικτών στα μοντέλα των αλγορίθμων, για τη διάγνωση και την ταξινόμηση της σοβαρότητας της COVID-19. [423]

Η ακρίβεια των μοντέλων κυμάνθηκε μεταξύ 84%-98% (**Εικόνα 43**), που είναι παρόμοια με τα αποτελέσματα που λαμβάνονται με RT-PCR και αποτελούν το ελάχιστο συνιστώμενο όριο για διαγνωστικές εξετάσεις, με το μοντέλο των ANN να εμφανίζει τις υψηλότερες επιδόσεις καθώς χρειάστηκε λιγότερο χρόνο εκπαίδευσης και με τα λιγότερα σφάλματα πρόβλεψης, σημειώνοντας ακρίβεια 94% (AuROC) για τη διάγνωση και 98% (AuROC) για την πρόβλεψη της σοβαρότητας και ως εκ τούτου, θα μπορούσε να χρησιμοποιηθεί ως υποστηρικτικό εργαλείο λήψης αποφάσεων για επαγγελματίες υγείας στην πράξη.



**Εικόνα 43-
Καμπύλες ROC
της ακρίβειας
των μοντέλων
[424]:**

Τεχνητά
Νευρωνικά Δίκτυα
(ANN): Διάγνωση
(Α) και
Σοβαρότητα (Β),

Δέντρα
Αποφάσεων (DT):
Διάγνωση (C) και
Σοβαρότητας (D),

Διακριτική
ανάλυση με
μερικά ελάχιστα
τετράγωνα (PLS-
DA): διάγνωση
(Ε) και
σοβαρότητα (F),

Κ-πλησιέστεροι
γείτονες (KNN):
διάγνωση (G) και
σοβαρότητα (H)

Σύμφωνα με την μελέτη οι βιοδείκτες που συνέβαλαν περισσότερο στην πρόβλεψη της διάγνωσης και της σοβαρότητας της COVID-19 ήταν η υπερφερριτιναιμία, η υπασβεστιαιμία (χαμηλά επίπεδα ιονισμένου ασβεστίου), η υποξαιμία (χαμηλή αρτηριακή πίεση οξυγόνου, pO_2), η πνευμονική υποξία (κλάσμα χαμηλής εισπνοής αρτηριακού οξυγόνου, FiO_2), η αναπνευστική οξέωση (υψηλά επίπεδα ολικού CO_2 και pCO_2), η μεταβολική οξέωση (υψηλά επίπεδα γαλακτικού οξέος και χαμηλό φλεβικό pH), το χαμηλό pH ούρων και τα υψηλά επίπεδα γαλακτικής αφυδρογονάσης (LDH) [Πίνακας 10], με τη φερριτίνη να κατατάσσεται ως ο πιο σημαντικός από τους βιοδείκτες για όλα τα μοντέλα.

Με βάση τα αποτελέσματα της μελέτης παρατηρήθηκε ότι τόσο οι ασθενείς με ασυμπτωματική ή ήπια νόσηση όσο και οι ασθενείς με σοβαρή νόσηση παρουσίασαν χαμηλά επίπεδα αιμοσφαιρίνης και υψηλά επίπεδα φερριτίνης. Όλα τα μοντέλα που αναπτύχθηκαν (ANN, PLS-DA, KNN και DT) έδειξαν ότι οι διαφορές στα επίπεδα αυτών των δύο βιοδεικτών ήταν κρίσιμες, τόσο για την πρόβλεψη της διάγνωσης COVID-19 όσο και την πρόβλεψη της σοβαρότητας της νόσησης. Αυτό υποστηρίζεται και από άλλες πρόσφατες μελέτες που έχουν δείξει πως αναιμία και υπερφερριτιναιμία, αποτελούν ισχυρούς βιοδείκτες για την πρόγνωση της θνησιμότητας ασθενών με COVID-19 [425-430]. Επίσης παρατηρήθηκαν υψηλά επίπεδα ολικού CO_2 , pCO_2 (αναπνευστική οξέωση), υψηλά επίπεδα γαλακτικού οξέος, χαμηλό φλεβικό pH (μεταβολική οξέωση), καθώς και αυξημένα επίπεδα χολερυθρίνης, τρανσαμινάσης της αλανίνης, και τρανσαμινάσης του ασπαρτικού οξέος, σε ασθενείς με ασυμπτωματική ή ήπια νόσηση και σε ασθενείς με σοβαρή νόσηση, τα οποία συνάδουν μελέτες άλλων χώρων [431-433]. Τέλος, εκτός από τις βιοχημικές παραμέτρους παρατηρήθηκε πως τα χαμηλά επίπεδα ασβεστίου αποτελούν προγνωστικό παράγοντα για τη διάγνωση και τη σοβαρότητα της νόσου. Αυτοί οι βιοδείκτες αποτελούν πιθανούς θεραπευτικούς στόχους που θα πρέπει να διερευνηθούν αποτελεσματικότερα σε περαιτέρω κλινικές δοκιμές.

Sumayh S. Aljameel et al. (April 2021)

Η μελέτη του Sumayh S. Aljameel και των συνεργατών του [434], στοχεύει στην παροχή μιας μεθόδου πρόβλεψης για την έγκαιρη αναγνώριση της έκβασης του κάθε ασθενούς με COVID-19 με τη χρήση της τριών αλγορίθμων ταξινόμησης με δεδομένα τα κλινικά χαρακτηριστικά των ασθενών οι οποίοι παρακολουθούνται στο σπίτι, ενώ βρίσκονται σε καραντίνα.

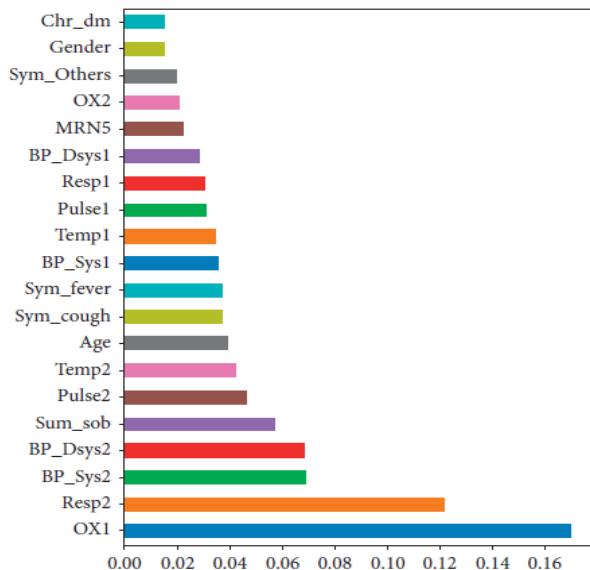
Η μελέτη πραγματοποιήθηκε χρησιμοποιώντας δείγματα ασθενών με COVID-19, από το Πανεπιστημιακό Νοσοκομείο King Fahad, Σαουδική Αραβία. Ειδικότερα το σύνολο δεδομένων περιέχει δημογραφικά και κλινικά στοιχεία δεδομένα ασθενών με COVID-19 την περίοδο από 30 Απρίλιος 2020 έως 24 Ιουλίου 2020. Το σύνολο δεδομένων περιέχει 287 αρχεία ασθενών με COVID-19, που εισήχθησαν στο Πανεπιστημιακό Νοσοκομείο King Fahad της Σαουδικής Αραβίας κατά την παραπάνω περίοδο (243 επιζήσαντες και 44 νεκροί). Ο **Πίνακας 11** απεικονίζει το σύνολο των χαρακτηριστικών των δεδομένων των ασθενών. Επιπλέον το σύνολο δεδομένων περιέχει πολύ μικρό αριθμό στοιχείων που να μην έχουν καταχωρηθεί.

No.	Feature name	UOM	Data type	Missing values
1	Age	Years	Numeric	0
2	Gender	Male/female	Nominal	0
3	BodyTemp (1&2)	Celsius (°C)	Numeric	1%-11%
4	Pulse rate (1&2)	Beats per minute (BPM)	Numeric	7%-5%
5	Resp (1&2)	Breaths per minute (BPM)	Numeric	3%-3%
6	BP_Sys (1&2)	mm Hg	Numeric	10%-7%
7	BP_Dsys (1&2)	mm Hg	Numeric	5%-5%
8	OX (1&2)	mm Hg	Numeric	4%-5%
9	Fever	Yes/no	Nominal	0
10	SOB	Yes/no	Nominal	0
11	Cough	Yes/no	Nominal	0
12	Symptoms_Others	—	Nominal	0
13	chr_dm	Yes/no	Nominal	0
14	chr_htn	Yes/no	Nominal	0
15	chr_cardiac	Yes/no	Nominal	0
16	chr_dlp	Yes/no	Nominal	0
17	Chr_ckd	Yes/no	Nominal	0
18	Chr disease_others	—	Nominal	0

Πίνακας 11-Περιγραφή του συνόλου δεδομένων [435]: Το πεδίο BodyTemp 1 στο Ο πίνακας δείχνει την θερμοκρασία του ασθενούς τη στιγμή της εισαγωγής του στο νοσοκομείο, με το πεδίο το BodyTemp 2 να είναι η θερμοκρασία του ασθενούς λίγο πριν το εξιτήριο το. Το πεδίο SOB (Shortness of Breath) υποδηλώνει δύσπνοια, το chr_dm υποδηλώνει χρόνια νόσο σακχαρώδη διαβήτη, Το chr_htn υποδηλώνει χρόνια υπέρταση, το chr_cardiac αντιπροσωπεύει τα χρόνια καρδιαγγειακά νοσήματα, το chr_dlp αντιπροσωπεύει δυσλιπιδαιμία, και το chr_ckd υποδηλώνει χρόνια νεφρική νόσο.

Στη μελέτη, χρησιμοποιήθηκαν 3 αλγόριθμοι ταξινόμησης, ο αλγόριθμος της λογιστικής παλινδρόμησης (Logistic Regression, LR), ο αλγόριθμος του τυχαίου δάσους (Random Forest, RF), και εκείνος της ακραίας ενίσχυσης κλίσης (eXtreme Gradient Boosting, XGB). Αρχικά, τα δεδομένα υποβλήθηκαν σε προεπεξεργασία χρησιμοποιώντας διάφορες τεχνικές προεπεξεργασίας. Επιπλέον, εφαρμόστηκε η μέθοδος της Διασταυρούμενης Επικύρωσης 10 (k) τμημάτων για την κατάτμηση των δεδομένων και η μέθοδος SMOTE (Synthetic Minority Oversampling Technique) [436] ώστε να μετριαστεί η ανισορροπία στα δεδομένα.

Τα μοντέλα υλοποιήθηκαν σε γλώσσα Python χρησιμοποιώντας το Jupyter (6.1.4) και τη βιβλιοθήκη sklearn (0.23.2). Δημιουργήθηκαν 4 σετ χαρακτηριστικών μέσω των ταξινομητών ExtraTree. Το 1^ο σετ περιελάμβανε όλα τα χαρακτηριστικά (25), το 2^ο τα κορυφαία 20 χαρακτηριστικά (Εικόνα 44), το 3^ο τα κορυφαία 15 χαρακτηριστικά και το 4^ο τα κορυφαία 10 χαρακτηριστικά, αντίστοιχα.



Εικόνα 44: Οι βαθμοί συσχέτισης των 20 κορυφαίων χαρακτηριστικών του σύνολο των δεδομένων [437]

Τα πειραματικά αποτελέσματα έδειξαν ότι ο αλγόριθμος του Τυχαίου Δάσους (RF) είχε την καλύτερη απόδοση όταν χρησιμοποιούνταν τα κορυφαία 20 χαρακτηριστικά, τόσο στα αρχικά δεδομένα, όσο και μετά την επεξεργασία με SMOTE, πετυχαίνοντας ακρίβεια 0,952, ευαισθησία 0,949, ειδικότητα 0,956 και σκορ F 0,955. Επίσης πέτυχε την υψηλότερη ειδικότητα 1,00 χρησιμοποιώντας τα κορυφαία 15 χαρακτηριστικά με SMOTE, σημείωσε επίσης AUC 0,99 στην καμπύλη ROC, με βάση τα κορυφαία 20 χαρακτηριστικά.

Η Λογιστική Παλινδρόμηση (LR), από την άλλη, ήταν τελευταία σε σχέση με άλλους ταξινομητές για τα κορυφαία 20, 15 και 10 χαρακτηριστικά που χρησιμοποιούν Δεδομένα SMOTE με ακρίβεια 0,863, 0,822 και 0,849, αντίστοιχα. Το LR σημείωσε AUC 0,91 στην καμπύλη ROC, με βάση τα κορυφαία 20 χαρακτηριστικά. Τα AU ROC σκορ των RF, XGB και LR απεικονίζονται στις **Εικόνες 45-47**.

Σε αυτή τη μελέτη, η ηλικία κατατάχθηκε μεταξύ των 10 κορυφαίων χαρακτηριστικών από τα 25 χαρακτηριστικά που χρησιμοποιήθηκαν στο μοντέλο πρόβλεψής μας. Παρακάτω στους **Πίνακες 12-15** απεικονίζονται οι αποδόσεις των ταξινομητών RF, XGB και LR για τα 4 διαφορετικά σετ χαρακτηριστικών (25,20,15,10).

Classifier	Sampling technique	Accuracy	Sensitivity	Specificity	F-score
LR	Without SMOTE	0.874	0.538	0.932	0.56
	With SMOTE	0.753	0.766	0.739	0.766
RF	Without SMOTE	0.908	0.75	0.924	0.6
	With SMOTE	0.938	0.947	0.929	0.941
XGB	Without SMOTE	0.885	0.6	0.922	0.545
	With SMOTE	0.925	0.923	0.926	0.929

Πίνακας 12 [438]

Classifier	Sampling technique	Accuracy	Sensitivity	Specificity	F-score
LR	Without SMOTE	0.874	0.538	0.932	0.56
	With SMOTE	0.863	0.82	0.93	0.88
RF	Without SMOTE	0.908	0.7	0.935	0.636
	With SMOTE	0.952	0.949	0.956	0.955
XGB	Without SMOTE	0.862	0.5	0.909	0.455
	With SMOTE	0.897	0.878	0.922	0.906

Πίνακας 13 [439]

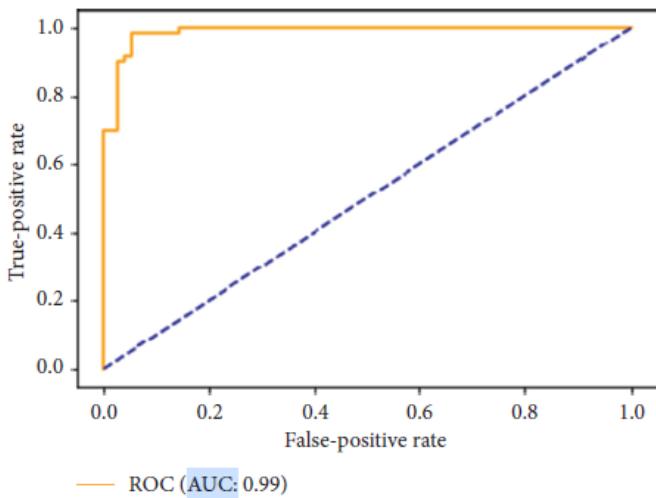
Classifier	Sampling technique	Accuracy	Sensitivity	Specificity	F-score
LR	Without SMOTE	0.874	0.583	0.932	0.56
	With SMOTE	0.822	0.793	0.864	0.841
RF	Without SMOTE	0.908	0.7	0.935	0.636
	With SMOTE	0.911	0.856	1	0.922
XGB	Without SMOTE	0.851	0.455	0.908	0.435
	With SMOTE	0.932	0.894	0.984	0.938

Πίνακας 14 [440]

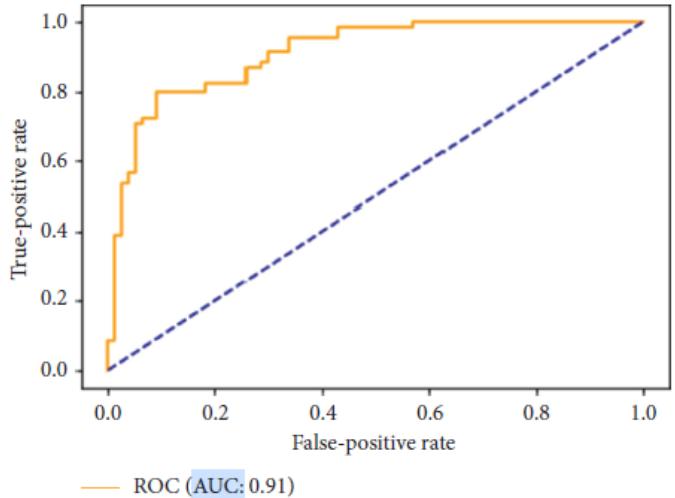
Επεξήγηση Πινάκων: Σύγκριση της απόδοσης των ταξινομητών με χρήση όλων(25) των χαρακτηριστικών (Πίνακας 12), των κορυφαίων 20 χαρακτηριστικών (Πίνακας 13), των κορυφαίων 15 χαρακτηριστικών (Πίνακας 14), των κορυφαίων 10 χαρακτηριστικών (Πίνακας 15), χρησιμοποιώντας πρωτότυπα δεδομένα και δεδομένα μετά από επεξεργασία με SMOTE.

Classifier	Sampling technique	Accuracy	Sensitivity	Specificity	F-score
LR	Without SMOTE	0.862	0.5	0.909	0.45
	With SMOTE	0.849	0.867	0.831	0.855
RF	Without SMOTE	0.89	0.63	0.934	0.609
	With SMOTE	0.925	0.884	0.983	0.933
XGB	Without SMOTE	0.851	0.455	0.908	0.43
	With SMOTE	0.89	0.843	0.965	0.904

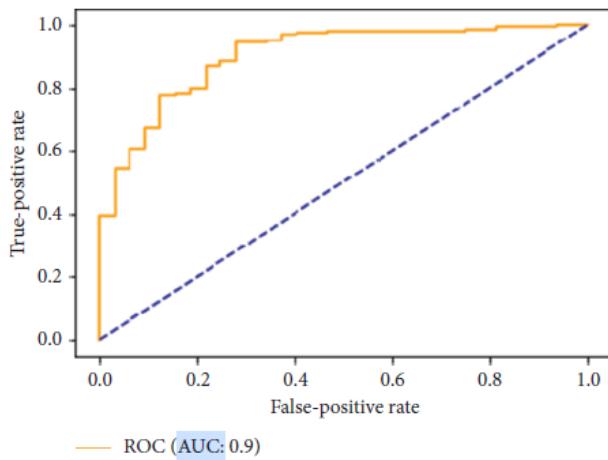
Πίνακας 15 [441]



Εικόνα 45: Η καμπύλη ROC της RF που χρησιμοποιούνται τα κορυφαία 20 χαρακτηριστικά [442]



Εικόνα 46: Η καμπύλη ROC της LR που χρησιμοποιούνται τα κορυφαία 20 χαρακτηριστικά [443]



Εικόνα 47: Η καμπύλη ROC της XGB που χρησιμοποιούνται τα κορυφαία 20 χαρακτηριστικά [444]

Σύμφωνα με τα παραπάνω αποτελέσματα είναι εμφανές πως το μοντέλο του αλγόριθμου Τυχαίου Δάσους (RF) ξεπέρασε τις επιδόσεις των άλλων 2 μοντέλων σε δεδομένα που εφαρμόστηκε η μέθοδος της Διασταυρούμενης Επικύρωσης 10 τμημάτων και η τεχνική αναζήτησης πλέγματος για τη βελτιστοποίηση των παραμέτρων. Η μελέτη πέτυχε ακρίβεια 0,952 και AUC 0,99 με το μοντέλο του Τυχαίου Δάσους (RF).

Πάρα την σημαντικά αποτελέσματα που επιτεύχθηκαν από αυτό το προτεινόμενο μοντέλο, υπάρχει ακόμα ανάγκη βελτίωσης. Επιπλέον θα πρέπει τα μοντέλα να επικυρωθούν χρησιμοποιώντας πολλαπλά σύνολα δεδομένων, και να ενσωματωθούν σε αυτά και άλλα κλινικά χαρακτηριστικά και εργαστηριακά αποτελέσματα (όπως η γαλακτική αφυδρογονάση LDH, τα επίπεδα ουδετερόφιλων και της ευαίσθητη C-αντιδρώσας πρωτεΐνης) τα οποία προσδιορίστηκαν ως σημαντικά σε άλλες μελέτες.

3.3 Πρόβλεψη μεταλλάξεων και εντοπισμός των επιτόπων του SARS-CoV-2

Ο ίος SARS-CoV-2 όπως και πολλοί άλλοι ιοί έχει καταφέρει να «μεταμορφωθεί» με συνεχή τρόπο δίνοντας έτσι νέες παραλλαγές μέσω μεταλλάξεων οι οποίες του προσδίδουν πλεονέκτημα επιβίωσης καθιστώντας τις παραλλαγές αρκετά επικίνδυνες. Ως εκ τούτου, απαιτούνται πολύπλοκες και χρονοβόρες αναλύσεις προκειμένου να εμβαθύνουμε στον ακριβή αντίκτυπο της κάθε νέας μετάλλαξης και παραλλαγής του ιού, ο χρόνος όμως που απαιτείται για αυτές τις αναλύσεις καθιστά δύσκολη την κατανόηση των νέων Παραλλαγών Ανησυχίας (VOCs) περιορίζοντας έτσι την προληπτική δράση που μπορεί να ληφθεί ενάντια στην ταχεία εξάπλωση τους. Στην επιτάχυνση αυτής της διαδικασίας στοχεύουν οι προσπάθειες ανάπτυξης μοντέλων πρόβλεψης των μεταλλάξεων και εντοπισμού αντιγονικών επιτόπων του SARS-CoV-2, ώστε να εξοικονομηθεί σημαντικός χρόνος και να επιταχυνθούν οι διαδικασίες σχεδιασμού και ανάπτυξης εμβολίων και άλλων θεραπευτικών στρατηγικών. Παρακάτω θα εξετάσουμε την ανάπτυξη 2 τέτοιων μοντέλων με το ένα να αποσκοπεί στην πρόβλεψη των μεταλλάξεων του ιού και το άλλο να έχει ως στόχο τον εντοπισμό των επιτόπων με την πιο επιτυχημένη αντιγονικότητα οι οποίοι μπορούν να χρησιμοποιηθούν ως βάση για την ανάπτυξη εμβολίων.

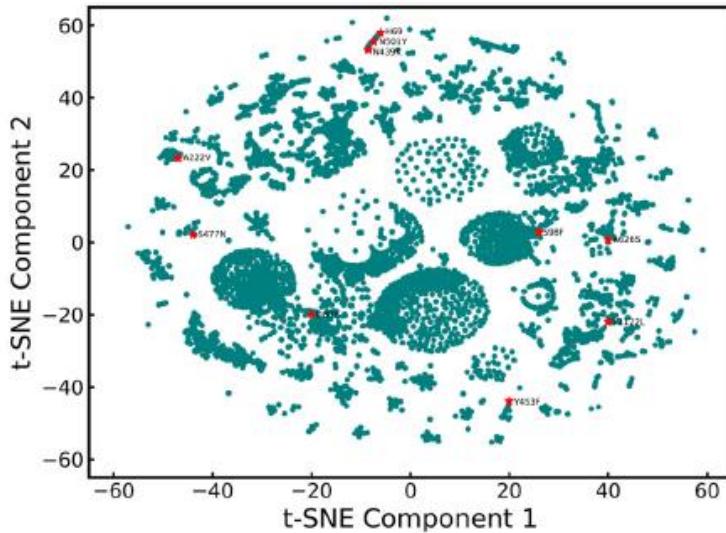
Baishali Mullick et al. (October 2021)

Η μελέτη της *Baishali Mullick* και των συνεργατών της [445], προτείνει μια υπολογιστική μεθοδολογία για τον εντοπισμό των πιθανών σημείων μετάλλαξης (*hotspots*) της πρωτεΐνης τη ακίδας (*Spike protein*) του SARS-CoV-2, τα οποία ενδέχεται να εμφανιστούν στο μακρινό μέλλον. Επίσης η μελέτη αυτή στοχεύει στον εντοπισμό και στην κατανόηση των ομοιοτήτων και των μοτίβων των διαφορετικών μεταλλάξεων με τη χρήση της ανάλυσης συστάδων, έτσι ώστε να κατανοηθούν καλύτερα οι βιολογικοί συσχετισμοί των μεταλλάξεων του ιού.

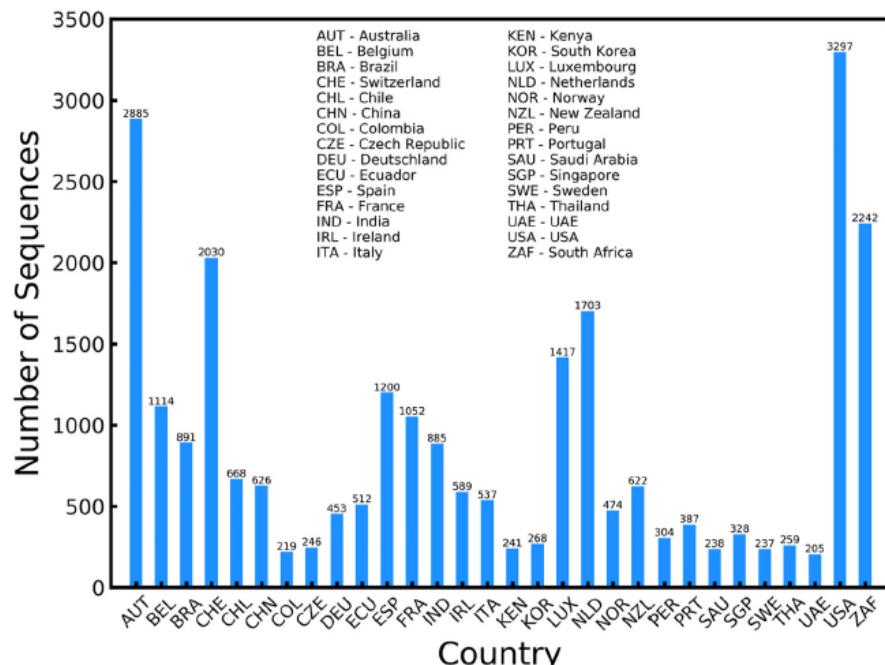
Ο εντοπισμός των θέσεων της αμινοξικής αλληλουχίας της ακίδας που είναι πιο επιρρεπείς σε μεταλλάξεις έγινε μέσω της στατιστικής τεχνικής της εντροπίας *Shannon*, η οποία αποτελεί μέτρο υπολογισμού της τυχαιότητας εμφάνισης ενός γεγονότος (μετάλλαξης) σε μια δεδομένη θέση μιας ακολουθίας [446], στη συνέχεια, χρησιμοποιούμε τις τεχνικές της ανάλυσης συστάδων ώστε να ομαδοποιήσουμε γνωστές και επικίνδυνες μεταλλάξεις με βάση τις ομοιότητες των ιδιοτήτων τους. Ο εντοπισμός μεταλλάξεων παρόμοιας φύσης και η επιλογή περιοχών ενδιαφέροντος, της πρωτεΐνης, με βάση την τάση για μετάλλαξη, γίνεται με τη βοήθεια του αλγορίθμου μοντελοποίησης της αμινοξικής αλληλουχίας των πρωτεϊνών *ProtBERT*.

Τα δεδομένα που χρησιμοποιήθηκαν στην μελέτη αποτελούνται από 311.256 διαφορετικές αλληλουχίες τη πρωτεΐνης ακίδας από τον ιστότοπο της **GISAID** (<https://www.gisaid.org/>) στις 3 Ιανουαρίου 2020 [93,447]. Οι μεταλλάξεις όλων των αλληλουχιών αναλύθηκαν ως προς την αρχική αλληλουχία της πρωτεΐνης της ακίδας που εντοπίστηκε στη Wuhan και αναλύθηκε από τους Zhao et al. [448]. Μετά από την προεπεξεργασία των δεδομένων απέμειναν 53.898 μοναδικές αλληλουχίες που ανήκουν σε κύριες Παραλλαγές

Ενδιαφέροντος (VOIs). Στη συνέχεια, αυτό το σύνολο δεδομένων χρησιμοποιήθηκε για τη δημιουργία embeddings (διαγράμματα ενός χώρου σχετικά χαμηλής διάστασης στο οποίο μπορεί να μεταφραστούν τα διανύσματα υψηλών διαστάσεων) μέσω του μοντέλου ProtBERT. Αυτά τα embeddings χρησιμοποιήθηκαν περαιτέρω για την πραγματοποίηση της μηχανικής μάθησης χωρίς επιβίβλεψη. Για τη καλύτερη κατανόηση της διασποράς των δεδομένων και την οπτικοποίησή τους, κατασκευάστηκε το αντίστοιχο διάγραμμα χρησιμοποιώντας το t-SNE (t-Κατανεμημένη Στοχαστική Ενσωμάτωση γειτόνων, t-Distributed Stochastic Neighbor Embedding) [449] του **Πίνακα 16**. Επίσης περαιτέρω, ανάλυση των γεωγραφικών τοποθεσιών και των χωρών της γενικής κατανομής της αλληλουχίες της πρωτεΐνης Spike βρέθηκε πως στο Ηνωμένο Βασίλειο και τη Δανία αντιστοιχούν πάνω από το 50% των μεταλλαγμένων αλληλουχιών στο σύνολο δεδομένων, με 140.458 ακολουθίες να προέρχονται από το Ηνωμένο Βασίλειο και 20.346 από τη Δανία (**Πίνακα 17**).



Πίνακας 16: Διάγραμμα t-SNE που αποτυπώνει την κατανομή των δεδομένων που συλλέγονται από το Διακομιστή GISAID. Ορισμένες από τις παραλλαγές που προκαλούν ανησυχία, όπως οι N439K, N501Y, συγκεντρώνονται κοντά το ένα στο άλλο, μέσω του οποίου μπορούμε να συμπεράνουμε πως οι μεταλλάξεις του SARS-CoV-2 έχουν μοναδικά χαρακτηριστικά. [450]



Πίνακας 17: Διάγραμμα που απεικονίζει την κατανομή των διαφορετικών ακολουθιών στα δεδομένα του GISAID. [451]

Μέσω της εντροπίας Shannon υπολογίστηκε η εντροπία της κάθε θέσης της αμινοξικής αλληλουχίας της πρωτεΐνης Spike και εκτιμήθηκαν οι περιοχές (hotspots) μεταλλάξεων

(Πίνακες 18 & 19). Από αυτή την ανάλυση της εντροπίας εντοπίστηκαν τα hotspots A222V, N439K, Y453F, S477 N, N501, D614G, V1122L [94], E484K, T478K και L452R που είναι επιρρεπή για μεταλλάξεις. Στις παραπάνω μεταλλάξεις συμπεριλαμβάνονται η μετάλλαξη D614G, που είναι μια από τις πιο κυρίαρχες μεταλλάξεις και βρέθηκε ότι ενισχύει την αντιγραφή του SARS-CoV-2 στα κύτταρα του πνεύμονα [452], η μετάλλαξη L452R που αναγνωρίστηκε για πρώτη φορά από το Τμήμα Δημόσιας Υγείας της Καλιφόρνια στις 17 Ιανουαρίου 2021[453] και αργότερα βρέθηκε ότι ήταν η κυρίαρχη μετάλλαξη σε τους μήνες Απρίλιο και Μάιο του 2021 παγκοσμίως, καθώς και η E484K που ανήκει στην οικογένεια B.1.25 αναγνωρίστηκε ως VOC στη Νότια Αφρική τον Απρίλιο του 2021 [454].

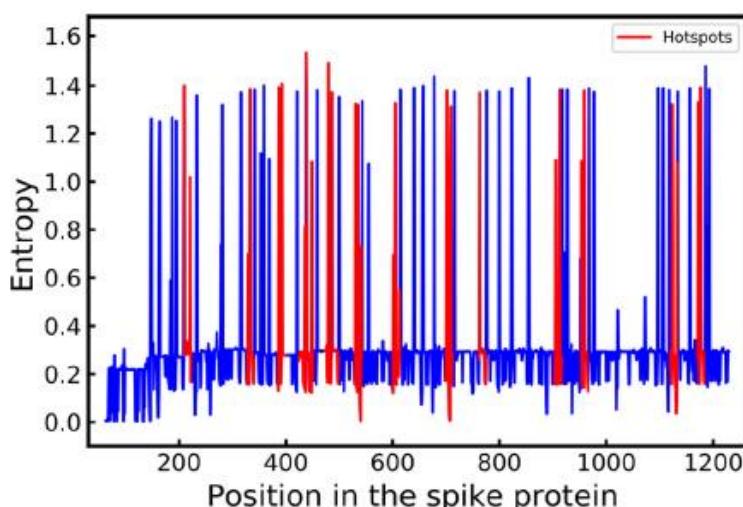
Για περαιτέρω επεξήγηση των τοπικών hotspot εντροπίας, σχεδιάστηκε η εντροπία θέσης για ολόκληρη την αλληλουχία της πρωτεΐνης ακίδας του SARS-CoV-2 στην Εικόνα 48, όπου βρέθηκαν άλλα 9 hotspot, που σύμφωνα με την ανάλυση του Πίνακα 18 είναι πιθανό να εμφανιστεί μια νέα ανησυχητική μετάλλαξη. Η θέση που βρίσκονται τα hotspot των μεταλλάξεων, στην πρωτεΐνη της ακίδας απεικονίζεται στον Πίνακα 19.

Hotspots	Mutation
211–225	A222V
439–453	N439K, L452R, Y453F
473–487	S477 N, T478K, E484K
487–501	N501Y
602–616	D614G
1121–1135	V1122L

Πίνακας 18: Τα Hotspots των μεταλλάξεων που βρέθηκαν αναλύοντας την εντροπία θέσης. [455]

Spike Protein Region	Mutation
N – Terminal domain	A222V
Receptor-Binding Domain	N439K, L452R, Y453F, S477 N, T478K, E484K, N501Y
Heptapeptide repeat sequence	V1122L

Πίνακας 19: Η θέση των hotspot των μεταλλάξεων στην πρωτεΐνη ακίδας του SARS-CoV-2. [456]



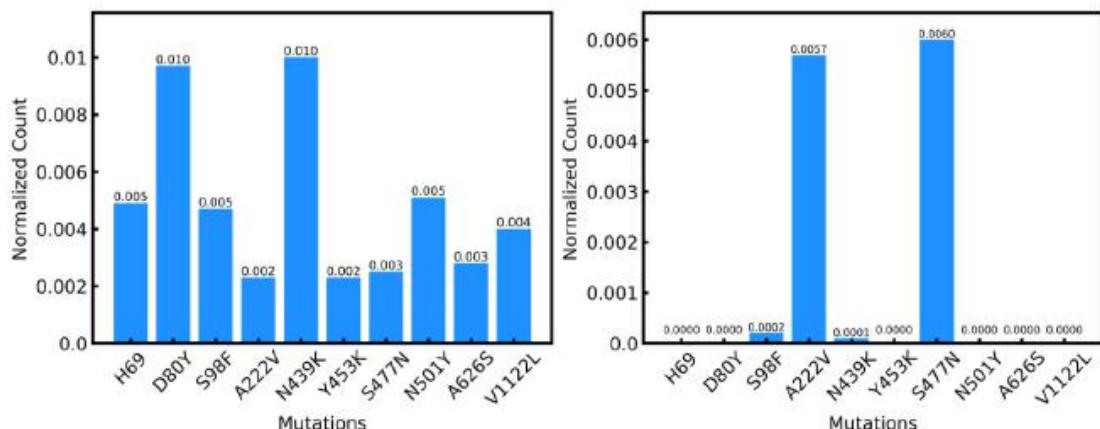
Εικόνα 48: Διάγραμμα διακύμανσης της εντροπίας και της θέσης στην πρωτεΐνη ακίδας. Τα Hotspots με υψηλότερη πιθανότητα μεταλλαξιγένεσης και υψηλή εντροπία έχουν σημειωθεί με κόκκινο χρώμα. Οι μπλε περιοχές στο διάγραμμα υποδεικνύουν τις περιοχές του σχετικά χαμηλότερη μέση εντροπία. Σύμφωνα με ανάλυση εντροπίας θέσης οι επικίνδυνες μεταλλάξεις πρωτεΐνης ακίδας είναι πιθανότερο να προκύψουν από τα κόκκινα hotspots. [457]

Επίσης χρησιμοποιήθηκε το μοντέλο Prot-BERT που εκπαιδεύτηκε στο σύνολο δεδομένων UniRef100 ώστε να κατασκευάζει embeddings ακολουθιών [458] τα οποία έπειτα χρησιμοποιούνται για τη εύρεση ομοιοτήτων και των κατανομών μεταξύ των μεταλλάξεων μέσω της μη ιεραρχικής συσταδοποίησης K-Means [459]. Για την εκτέλεση της k-means χρησιμοποιήθηκε η βιβλιοθήκη scikit-learn [460,461], επιλέγοντας τον αριθμό των συστάδων να είναι 10, το μέγιστο αριθμό επαναλήψεων να είναι 1000 και το συνολικό αριθμό

των αρχικοποιήσεων να είναι 50. Οι συγκεκριμένοι παράμετροι προέκυψαν μετά από πολλαπλές δοκιμές προκειμένου να σταθεροποιηθεί ο σχηματισμός συστάδων. Τα κέντρα των συστάδων που προέκυψαν από την k-means αντιστοιχούν στους διαφορετικούς τύπους μετάλλαξης, επαληθεύοντας έτσι την υπόθεση ότι οι διαφορετικοί τύποι συμπλέγματος ομαδοποιούνται χωριστά, με τις 7 από τις 10 διαφορετικές μεταλλάξεις να αναγνωρίζονται ως κέντρα συστάδων μετά από λίγες επαναλήψεις. Αναλύοντας τις αλληλουχίες της ακίδας που σχηματίζουν την κάθε συστάδα καθώς και την αντιπροσωπευτική ακολουθία του κέντρου κάθε συστάδας, διαπιστώθηκε ότι στις περισσότερες περιπτώσεις η πλειοψηφία των ακολουθιών τις κάθε συστάδας αναγνωρίζεται ως ίδιου τύπου με την αλληλουχία του κέντρου, ενώ στις άλλες περιπτώσεις ο τύπος αλληλουχίας του κέντρου της συστάδας είναι μεταξύ των 3 κορυφαίων τύπων αλληλουχιών της συστάδας με τους άλλους 2 τύπους να φέρουν παρόμοια χαρακτηριστικά (**Πίνακας 20**). Για παράδειγμα τα διαγράμματα της **Εικόνας 49** δείχνουν ότι οι αλληλουχίες που ανήκουν στις συστάδες με κέντρα τις αλληλουχίες S477 N και N439K, έχουν στην πλειοψηφία τους χαρακτηριστικά όμοια των κέντρων τους. Επιπλέον η αλληλουχία A222V εμφανίζεται να έχει τη 2^η υψηλότερη μέτρηση στη συστάδα με κέντρο την S477 N, υποδεικνύοντας ομοιότητες μεταξύ τους, το ίδιο συμβαίνει και με τη D80Y στη συστάδα με κέντρο την N439K, υποδηλώνοντας έτσι πως έχουν παρόμοια χαρακτηριστικά.

Cluster Centers	Dominant Mutations in the Cluster
S477 N	S477N, A222V, S98F, N439K
N439K	N439K, D80Y, N501Y, H69-70
N501Y	D80Y, N439K, N501Y, H69
A222V	V1122L, A222V, N501Y, S477 N

Πίνακας 20: Συστάδες όπου οι 3 κορυφαίες κυριαρχείς μεταλλάξεις στο σύμπλεγμα συμφωνούν με τη μετάλλαξη του κέντρου συστάδας. Οι κορυφαίες 3 κυριαρχείς μεταλλάξεις είναι πιο πιθανό να έχουν παρόμοια χαρακτηριστικά με τις μεταλλάξεις της κάθε συστάδας. [462]



Εικόνα 49 [463]:

- Α) Ανάλυση συσταδοποίησης για τη μετάλλαξη N439K στην πρωτεΐνη ακίδας του SARS-CoV-2. Μετά από ανάλυση της συστάδας με κέντρο την N439K μπορούμε να συμπεράνουμε ότι η D80Y μπορεί να έχει παρόμοια χαρακτηριστικά με αυτήν
- Β) Ανάλυση συσταδοποίησης για τη μετάλλαξη S477 N στην πρωτεΐνη ακίδας του SARS-CoV-2. Η πλειοψηφία των ακολουθιών σε αυτήν τη συστάδα ανήκει στη μετάλλαξη S477 N με την 2^η επικρατέστερη μετάλλαξη να είναι η A222V, δείχνοντας έτσι την μεταξύ τους ομοιότητα.

Σε αυτή τη μελέτη, βρέθηκαν 15 περιοχές ενδιαφέροντος στην αμινοξική αλληλουχία της πρωτεΐνης Spike ως πιθανά hotspot για νέες μεταλλάξεις στον SARS-CoV-2, με 6 από αυτά

τα hotspots να περιέχουν 10 μεταλλάξεις που έχουν ήδη σημανθεί ως πιο μεταδοτικές. Επιπλέον μερικές από τις νέες παραλλαγές όπως αυτές της Ινδίας (Delta) και της Νότιας Αφρικής (Beta) που χαρακτηρίστηκαν επικίνδυνες (VOCs) τον Απρίλιο και τον Μάιο του 2021 αντίστοιχα προσδιορίστηκαν από τη μεθοδολογία της συγκεκριμένης μελέτης, αν και χρησιμοποιήθηκαν δεδομένα ακολουθιών της GISaid πριν από τον Δεκέμβριο 2020. Ένα ακόμα αποτέλεσμα της έρευνας ήταν η ανάλυση της ομοιότητας μεταξύ των μεταλλάξεων S477N και A222V, δείχνοντας έτσι ότι αυτές οι μεταλλάξεις μοιράζονται κοινά χαρακτηριστικά και μπορεί να υποβληθούν σε παρόμοιες θεραπευτικές στρατηγικές. Καθώς η πανδημία εξελίσσεται ραγδαία, είναι πολύ πιθανό ότι τα υπόλοιπα 9 hotspot μπορεί να περιέχονται σε παραλλαγές του ιού που μπορεί να εμφανιστούν στο μέλλον. Η ανάλυσης της ομοιότητας μεταξύ των μεταλλάξεων είναι σημαντική, γιατί κατά το σχεδιασμό θεραπευτικών στρατηγικών αντιμετώπισης παραλλαγών με νέες μεταλλάξεις, μπορεί να εξοικονομηθεί σημαντικός χρόνος και επιταχυνθεί η διαδικασία ανάπτυξης θεραπείας.

Zeynep Banu Ozger et al. (December 2021)

Η μελέτη της *Zeynep Banu Ozger* και των συνεργατών της [464], έχει ως στόχο να προτείνει μια νέα προσέγγιση για την επιτυχή πρόβλεψη των επιτόπων των B-λεμφοκυττάρων για τον SARS-CoV-2, συμβάλλοντας έτσι στην ανάπτυξη γρήγορων και αποτελεσματικών εμβολίων κατά του ιού.

Το σύνολο δεδομένων [465], των επιτόπων των B-λεμφοκυττάρων, που χρησιμοποιήθηκε στη μελέτη, ελήφθη από το IEDB και το UniProt. Τα δεδομένα αυτά χωρίζονται σε 3 κατηγορίες, στα δεδομένα των επιτόπων των B-λεμφοκυττάρων, τα δεδομένα των επιτόπων του SARS-CoV και τα δεδομένα των πεπτιδίων του SARS-CoV-2. Από αυτές τις κατηγορίες οι 2 πρώτες είναι επισημασμένα δεδομένα ενώ η 3^η (πεπτίδια του SARS-CoV-2) δεν είναι επισημασμένα. Το σύνολο δεδομένων του SARS-CoV περιέχουν συνολικά 520 δείγματα, με τα 140 να ανήκουν στη θετική κατηγορία, που σημαίνει αυτό το αντίστοιχο πεπτίδιο είναι ο επίτοπος, και τα υπόλοιπα 380 δείγματα να ανήκουν στην αρνητική, δηλαδή δεν είναι επίτοποι. Το μακρύτερο πεπτίδιο έχει μήκος 393 αμινοξέα και το μικρότερο πεπτίδιο αποτελείται από μόνο 5 αμινοξέα. Το σύνολο δεδομένων του SARS-CoV-2 περιλαμβάνει 20.312 δείγματα με το μήκος των πεπτιδίων να κυμαίνεται από 5-20 αμινοξέα. Οι πρωτεΐνες των συνόλων δεδομένων είναι οι πρωτεΐνες των αντισωμάτων ανοσοσφαιρίνης, του πιο κοινού τύπου αντισωμάτων που βρίσκεται στον ορό του αίματος. Επίσης το σύνολο των δεδομένων περιλαμβάνει τις αλληλουχίες των πρωτεϊνών και των πεπτιδίων, τα αναγνωριστικά των πρωτεϊνών, τις θέσεις έναρξης και λήξης των πεπτιδίων, καθώς και τα χαρακτηριστικά των πρωτεϊνών και των πεπτιδίων. Η πρόβλεψη των επιτόπων των B-λεμφοκυττάρων γίνεται με βάση την αντιγονικότητα, την υδροφοβικότητα, την επιφανειακή προσβασιμότητα, βήτα στροφές και ιδιότητες ευελιξίας των επιτόπων. Τα χαρακτηριστικά του συνόλου δεδομένων δίνονται στον **Πίνακα 21**.

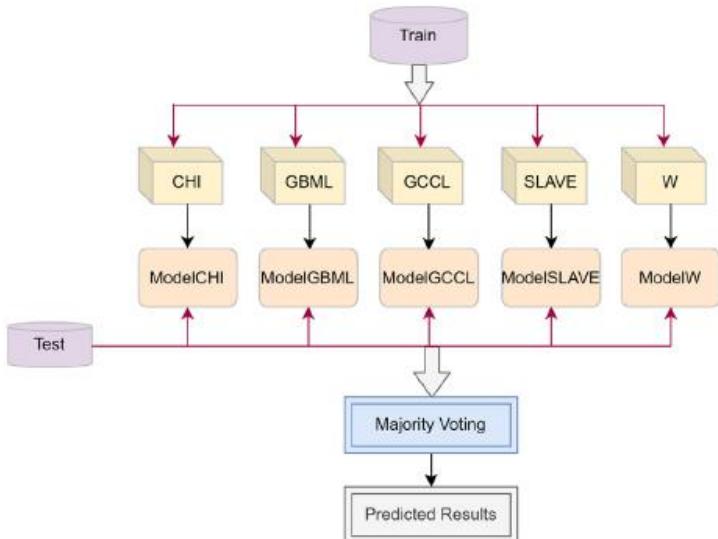
Feature	Description
Chou-Fasman	Peptide feat. Relative frequency analysis on the basis of amino acids for tertiary structural elements. Given here for B-Turn.
Emini	Peptide feat, relative surface accessibility, a measure of residue solvent exposure.
Kolaskar-Tongaonkar	Peptide feat. Antigenicity, antigenic propensity of residues.
Parker	Peptide feat. A measure of hydrophobicity of peptide.
Isoelectric-point	Protein feat. pH value of the amino acid in an electric field.
Aromaticity	Protein feat. A factor for protein fragment solubility.
Hydrophobicity	Protein feat. A measure of the degree of affinity between water and the side chain of an amino acid.
Stability	Protein feature.

Πίνακας 21: Τα χαρακτηριστικά των πεπτιδίων και πρωτεΐνών των συνόλων δεδομένων. [466]

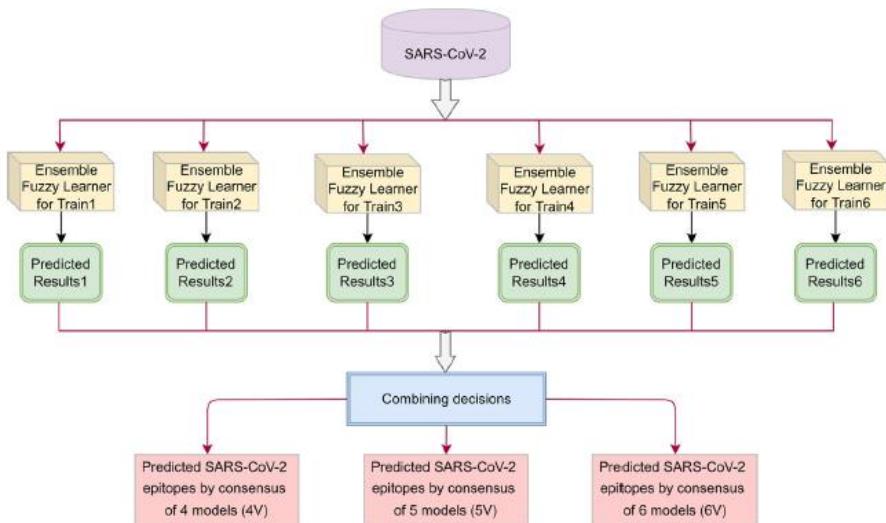
Στη μελέτη συγκρίθηκαν οι αποδόσεις 6 διαφορετικών μοντέλων ασαφών ταξινομητών, στο να προβλεφθούν οι κατάλληλοι επίτοποι των Β-λεμφοκυττάρων του SARS-CoV:

1. το μοντέλο της Γενετικής Συνεργατικής/Ανταγωνιστικής Μάθησης(Genetic Cooperative Competitive Learning, GCCL) [467]
2. το μοντέλο των Ασαφών Γενετικών αλγόριθμων Μηχανικής Μάθησης(Based-Machine Learning GBML) [468]
3. το ασαφές μοντέλο της μεθόδου του Chi (CHI) [469]
4. το μοντέλο της μεθόδου του Ishibuchi με συντελεστή βάρους (W) [470]
5. το μοντέλο του αλγορίθμου δομικής μάθησης σε ασαφές περιβάλλον (Structural Learning Algorithm on Vague Environment, SLAVE) [471]
6. το προτεινόμενο μοντέλου ασαφούς ταξινόμησης συνάθροισης(ensemble fuzzy classification model) [464]

Το μοντέλο της ασαφούς ταξινόμησης συνάθροισης (ensemble fuzzy classification model) συνδυάζει τα άλλα 5 μοντέλα (GCCL, GBML, CHI, W και SLAVE) χρησιμοποιώντας ένα πλειοψηφικό σύστημα λήψης απόφασης, ώστε να καθορίσει την κλάση του κάθε δείγματος (**Εικόνα 50**). Τα σετ εκπαίδευσης και η δοκιμής του μοντέλου δημιουργήθηκαν μέσω τυχαίας επιλογής δειγμάτων, με τη διαδικασία δημιουργίας των σετ να επαναλαμβάνεται 6 φορές, δημιουργώντας έτσι 6 ζεύγη σετ εκπαίδευσης-δοκιμής. Η απόδοση του μοντέλου υπολογίστηκε ως ο μέσος όρος από τις επιδόσεις του μοντέλου για κάθε ζεύγος σετ δεδομένων εκπαίδευσης-δοκιμής. Για την δημιουργία των σετ εκπαίδευσης χρησιμοποιήθηκαν τα δεδομένα του SARS-CoV, ενώ για τα σετ δοκιμής (προβλέψεων) χρησιμοποιήθηκαν τα μη επισημασμένα δεδομένα του SARS-CoV-2. Οι αποφάσεις που λαμβάνονται από τους 6 κύκλους του μοντέλου συνδυάζονται με διαφορετικούς βαθμούς της ακρίβειας, με τις τελικές αποφάσεις για την επιλογή των επιτόπων να ονομάζονται 4V (τουλάχιστον 4 φόρες εμφάνισης του επιτόπου στους 6 κύκλους), 5V (τουλάχιστον 5 φόρες εμφάνισης του επιτόπου στους 6 κύκλους) και 6V (εμφάνιση του επιτόπου και στους 6 κύκλους). Η διαδικασία πρόβλεψης επιτόπου για τον SARS-CoV-2 είναι απεικονίζεται στην **Εικόνα 51**.



Εικόνα 50: Το προτεινόμενο μοντέλο ασαφούς ταξινόμησης συνάθροισης(ensemble fuzzy classification model). [472]



Εικόνα 51: Διαδικασία πρόβλεψης για SARS-CoV-2 μέσω της πλειοψηφικής λήψης απόφασης μετά από 6 κύκλους εκτέλεσης του προτεινόμενου μοντέλου ασαφούς ταξινόμησης συνάθροισης(ensemble fuzzy classification model).[473]

Έπειτα επιλέχθηκε το προτεινόμενο μοντέλο της ασαφούς ταξινόμησης συνάθροισης, το οποίο εμφάνισε το χαμηλότερο σφάλμα (μέσος όρος ποσοστών σφαλμάτων μοντέλου ασαφούς ταξινόμησης συνάθροισης=8,33, GCCL=30,42, GBML=23,82, CHI=29,17, W=46,25 και SLAVE=20,42), για την πρόβλεψη των επιτόπων του SARS-CoV-2 (**Πίνακας 22**).

Το κύριο πλεονέκτημα της προτεινόμενης μεθόδου είναι ότι οι αποφάσεις που γίνονται μέσω των επιμέρους ασαφών μεθόδων συνδυάζονται σε μια δομή συνόλου. Οι ασαφείς μέθοδοι που χρησιμοποιούνται περιλαμβάνουν τρόπους προσεγγίσεις όσον αφορά την εκπαίδευση και τη λήψη αποφάσεων (προβλέψεις) με την κάθε μια από αυτές να μαθαίνει διαφορετικά χαρακτηριστικά του συνόλου δεδομένων. Έτσι μειώνοντας τη συμβολής του κάθε μοντέλου στην τελική απόφαση, βελτιώνεται η μέση απόδοση πρόβλεψης του μοντέλου της συνάθροισης (**Πίνακας 22**) εμφανίζοντας μέση ακρίβεια 91,7%.

Method	Test1	Test2	Test3	Test4	Test5	Test6	Avg.	Sig
CHI	17.5	37.5	25	27.5	22.5	20	29.17	+
GBML	17.5	15	32.5	15	20	22.5	23.82	+
GCCL	22.5	55	30	12.5	30	32.5	30.42	+
SLAVE	27.5	17.5	7.5	25	22.5	22.5	20.42	+
W	52.5	52.5	52.5	27.5	45	47.5	46.25	+
Ensemble fuzzy	7.5	12.5	7.5	10	7.5	5	8.33	

Πίνακας 22: Σφάλματα των 6 μοντέλων στην πρόβλεψη των επιτόπων για τον SARS-CoV. [474]

Στη συνέχεια οι προβλέψεις των αλληλουχιών των επιτόπων για το SARS-CoV-2 που προέκυψαν, συγκρίθηκαν με τις προβλέψεις των αλληλουχιών των επιτόπων του **BepiPred server** [475] που έγιναν με βάσει τα ίδια δεδομένα, όπως και τα **Vaxijen 2.0** σκορ μελετών ανοσοπληροφορικής [476]. Ο BepiPred πραγματοποιεί προβλέψεις μέσω ενός μοντέλου, που χρησιμοποιεί το αλγόριθμο Random Forest (RF) και που έχει εκπαιδευτεί σε ένα σύνολο δεδομένων 649 κρυσταλλικών δομών αντιγόνου-αντισώματος. Δεδομένου ότι οι αμινοξικές αλληλουχίες (πεπτίδια) του SARS-CoV-2 δεν είναι σημασμένες, οι επίτοποι που επιλέχθηκαν με το προτεινόμενο μοντέλο συνάθροισης, συγκρίθηκαν με τους επιτόπους που εντόπισε ο BepiPred server για την ίδια πρωτεΐνη. Ο BepiPred για την πρωτεΐνη Spike, εντόπισε 44 πεπτίδια, μήκους 1–36 αμινοξέων, ως υποψήφιους επιτόπους για κατασκευή εμβολίων. Συγκρίνοντας τα αποτελέσματα αυτά, με εκείνα του προτεινόμενου μοντέλου, διαπιστώθηκε πως από τα 44 πεπτίδια, με τη μέθοδο 4V εντοπίστηκαν τα 21 από αυτά, με τη μέθοδο 5V εντοπίστηκαν τα 20 από αυτά και με τη μέθοδο 6V εντοπίστηκαν τα 18 από αυτά. Όταν πεπτίδια με διαφορετικά μήκη αλληλουχίας συγκρίθηκαν με βάσει τις βαθμολογίες αντιγονικότητας τους, 6 πεπτίδια που βρέθηκαν με τη προτεινόμενη μέθοδο (ensemble) είχαν υψηλότερα βαθμολογίες αντιγονικότητας, με την 5V να δίνει το καλύτερο αποτέλεσμα, ενώ το BepiPred είχε πιο επιτυχημένη αντιγονικότητας για 5 πεπτίδια.

Τέλος τα παραπάνω αποτελέσματα, συγκρίθηκαν με επιτόπους που βρέθηκαν, σε άλλες βιβλιογραφικές μελέτες με διάφορα εργαλεία βιοπληροφορικής, ή *in vitro* μεθόδους, όπου φάνηκε ότι οι περισσότεροι από τους επιτόπους που προτείνονται για την κατασκευή εμβολίων από τις βιβλιογραφικές μελέτες μπορούν επίσης να ανιχνευθούν και από το μοντέλο ασαφούς ταξινόμησης συνάθροισης (ensemble fuzzy classification model), αποδεικνύοντας έτσι την ευρωστία του προτεινόμενου μοντέλου, με την μέθοδο 5V να εμφανίζεται ως η πιο επιτυχημένη όσον αφορά τον μέσο όρο βαθμολογία αντιγονικότητας.

Το προτεινόμενο μοντέλο πρόβλεψης (ensemble fuzzy classification model) επιτόπων που αναπτύχθηκε έχει σκοπό να βοηθήσει στο σχεδιασμό αποτελεσματικών εμβολίων και φαρμάκων κατά των μελλοντικών εξάρσεων τόσο του SARS-CoV-2 και των παραλλαγών του όσο και από εξάρσεις από άλλα είδη της οικογενείας των κορονοϊών.

ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Εργαλεία και Μέθοδοι

4.1 Σκοπός

Η παρούσα εργασία στοχεύει στην σύγκριση **6** διαφορετικών αλγορίθμων μηχανικής μάθησης, με σκοπό την εύρεση του βέλτιστου μοντέλου πρόβλεψης της θνησιμότητας ασθενών με COVID-19 χρησιμοποιώντας ως δεδομένα τα κλινικά χαρακτηριστικά τους.

4.2 Δεδομένα

Το σύνολο των δεδομένων απαρτίζεται από **12.425.179** άτομα ύποπτα για COVID-19, τα οποία προσήλθαν σε υγειονομικές μονάδες του Μεξικού. Τα δεδομένα είναι υπό μορφή αρχείου csv το ανακτήθηκε μέσω συνδέσμου που παρέχεται από την κυβέρνηση του Μεξικού [[477](#)].

4.3 Προεπεξεργασία

Καθαρισμός Δεδομένων

Αρχικά τα ονόματα των χαρακτηριστικών μεταφράστηκαν από τα Ισπανικά στα Αγγλικά. Έπειτα αφαιρέθηκαν τα δείγματα εκείνων που δεν βρέθηκαν θετικοί με COVID-19 καταλήγοντας έτσι σε **3.993.464** ασθενείς. Τα **3.993.464** δείγματα κρίθηκαν θετικά με κριτήριο την τιμή που είχε το καθένα για το χαρακτηριστικό FINAL CLASSIFICATION, με τις τιμές 1, 2 και 3 να δηλώνουν τη θετικότητα στον SARS-CoV-2. Κατόπιν από τα 40 χαρακτηριστικά(στήλες) των δεδομένων αφαιρέθηκαν εκείνα που δεν σχετίζονται με την πρόβλεψη της θνησιμότητας από COVID-19 καταλήγοντας στη χρήση των 24 από αυτά. Στη συνέχεια το χαρακτηριστικό «Ημερομηνία Θανάτου» (DATE OF DEATH) μετασχηματίστηκε σε κατηγορικό (δυαδικό) με νέο όνομα «Επιβίωσε» (SURVIVED) καθώς οι ασθενείς που επιβίωσαν είχαν την τιμή 9999-99-99. Τέλος τα χαρακτηριστικά «Ημερομηνία Έναρξης Συμπτωμάτων» (SYMPTOM ONSET DATE) και «Ημερομηνία Εισαγωγής» (HOSPITAL ADMISSION DATE), συνδυάστηκαν σε ένα χαρακτηριστικό στο οποίο δόθηκε το όνομα «Ημέρες από την Εκδήλωση των Συμπτωμάτων έως την Εισαγωγή» (DAYS FROM SYMPTOM TO HOSPITALIZATION) καταλήγοντας έτσι τελικά σε 23 χαρακτηριστικά (στήλες), με τη μια εξ αυτών να αποτελεί τον «Κωδικό Ταυτοποίησης» (REGISTRATION_ID). Τα 22 χαρακτηριστικά που χρησιμοποιήθηκαν συνοψίζονται στον **Πίνακα 23**.

Πίνακας 23: Τα 22 χαρακτηριστικά που χρησιμοποιήθηκαν για τον κάθε ασθενή.

01	SEX	Το φύλο του ασθενή [1:Γυναίκα, 2:Άντρας]
02	TYPE OF PATIENT	Το είδος φροντίδας που έλαβε ο ασθενής, πιο συγκεκριμένα το αν έκανε ή όχι εισαγωγή στο νοσοκομείο. [1:Δεν έκανε Εισαγωγή, 2:Έκανε Εισαγωγή]
03	INTUBATED	Το αν ο ασθενής διασωληνώθηκε ή όχι. [1:Nαι, 2:Όχι, 97:Δεν μπορεί να εφαρμοστεί το κριτήριο, καθώς ο ασθενής δεν διασωληνώθηκε]
04	PNEUMONIA	Το αν ο ασθενής διαγνώστηκε με πνευμόνια. [1:Nαι, 2:Όχι]
05	AGE	Η ηλικία του ασθενή
06	PREGNANCY	Το αν ο ασθενής είναι έγκυος. [1:Nαι, 2:Όχι, 97:Δεν μπορεί να εφαρμοστεί το κριτήριο, καθώς ο ασθενής είναι άντρας]
07	DIABETIC	Το αν ο ασθενής έχει διαγνωστεί με Διαβήτη. [1:Nαι, 2:Όχι]
08	COPD	Το αν ο ασθενής έχει διαγνωστεί με ΧΑΠ. [1:Nαι, 2:Όχι]
09	ASTHMA	Το αν ο ασθενής έχει διαγνωστεί με Άσθμα. [1:Nαι, 2:Όχι]
10	IMMUNOSUPPRESSED	Το αν ο ασθενής λαμβάνει Ανοσοκατασταλτικά. [1:Nαι, 2:Όχι]
11	HYPERTENSION	Το αν ο ασθενής έχει διαγνωστεί με Υπέρταση. [1:Nαι, 2:Όχι]
12	OTHER CHRONIC DISEASE	Το αν ο ασθενής έχει διαγνωστεί με άλλο Χρόνιο νόσημα. [1:Nαι, 2:Όχι]
13	CARDIOVASCULAR	Το αν ο ασθενής έχει διαγνωστεί με ασθένεια του Καρδιαγγειακού. [1:Nαι, 2:Όχι]
14	OBESITY	Το αν ο ασθενής είναι Υπέρβαρος. [1:Nαι, 2:Όχι]
15	CHRONIC KIDNEY FAILURE	Το αν ο ασθενής έχει διαγνωστεί με Χρόνια Νεφρική Ανεπάρκεια. [1:Nαι, 2:Όχι]
16	SMOKER	Το αν ο ασθενής είναι Καπνιστής. [1:Nαι, 2:Όχι]
17	CONTACT WITH COVID-19 CASE	Το αν ο ασθενής είχε επαφή επιβεβαιωμένο κρούσμα SARS CoV-2. [1:Nαι, 2:Όχι]
18	LAB RESULT	Το αποτέλεσμα της ανάλυσης του δείγματος μετά από τον εργαστηριακό έλεγχο [1:Θετικό στον SARS-CoV-2, 2:Αρνητικό στον SARS-CoV-2, 3: Ev Αναμονή Αποτελέσματός, 4:Το Αποτέλεσμα δεν είναι Σαφές]. Η στήλη σε συνδυασμό με την FINAL CLASSIFICATION χρησιμοποιήθηκαν για την επιλογή μόνο των πραγματικών κρουσμάτων

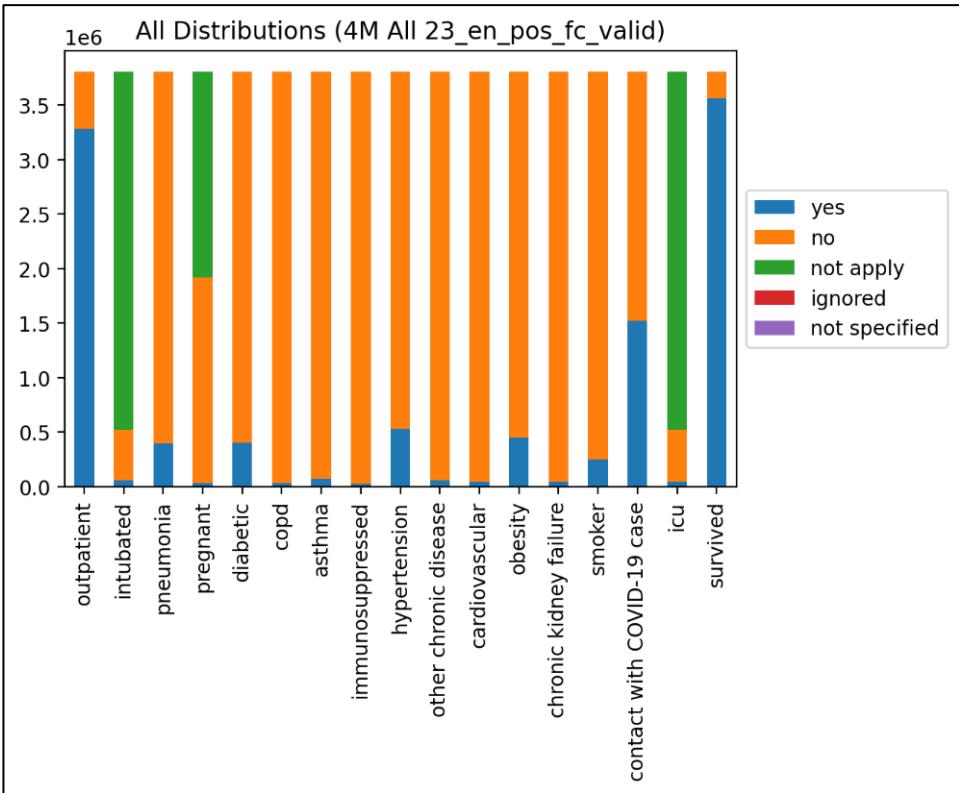
19	FINAL CLASSIFICATION	Το αν ο ασθενής ανήκει σε κρούσμα COVID-19. [1: Επιβεβαιωμένο κρούσμα σύμφωνα με τον Επιδημιολογικό Σύλλογο του Μεξικού, 2: Επιβεβαιωμένο κρούσμα σύμφωνα με την Επιτροπή Ιατρικών Αποφάσεων του Μεξικού, 3: Επιβεβαιωμένο κρούσμα, 4: Μη έγκυρα ταυτοποιημένο κρούσμα, 5: Δείγμα που ο έλεγχος του δεν πραγματοποιήθηκε από εργαστήριο, 6: 'Υποπτο μη επιβεβαιωμένο κρούσμα, 7: Δείγμα Αρνητικό για την COVID-19] Η στήλη σε συνδυασμό με την LAB RESULT χρησιμοποιήθηκαν για την επιλογή μόνο των πραγματικών κρουσμάτων
20	ICU	Το αν ο ασθενής εισήχθη σε ΜΕΘ. [1:Ναι, 2:Όχι, 97:Δεν μπορεί να εφαρμοστεί το κριτήριο, καθώς ο ασθενής δεν εισήχθη σε ΜΕΘ]
21	DAYS FROM SYMPTOM TO HOSPITALIZATION (Δημιουργήθηκε)	Χαρακτηριστικό(στήλη) που δημιουργήθηκε μετά από την αφαίρεση της ημερομηνίας έναρξής των συμπτωμάτων (SYMPTOM ONSET DATE) από την ημερομηνία εισαγωγής στο νοσοκομείο (HOSPITAL ADMISSION DATE).
22	SURVIVED (Δημιουργήθηκε)	Χαρακτηριστικό(στήλη) που δημιουργήθηκε με βάση τη στήλη DATE_OF_DEATH, με τους ασθενείς που επιβίωσαν [1:Ναι] να έχουν ημερομηνία της μορφής 9999-99-99 και εκείνους του απεβίωσαν [2:Όχι] να έχουν ημερομηνία τις μορφής Έτος(YYYY)-Μήνας(MM)-Ημέρα(DD).

Στη συνέχεια αφαιρέθηκαν **184.345** δείγματα τα οποία περιείχαν μη έγκυρες(invalid) τιμές για 1 ή περισσότερα χαρακτηριστικά τους. Ως μη έγκυρες τιμές θεωρούνται οι παρακάτω:

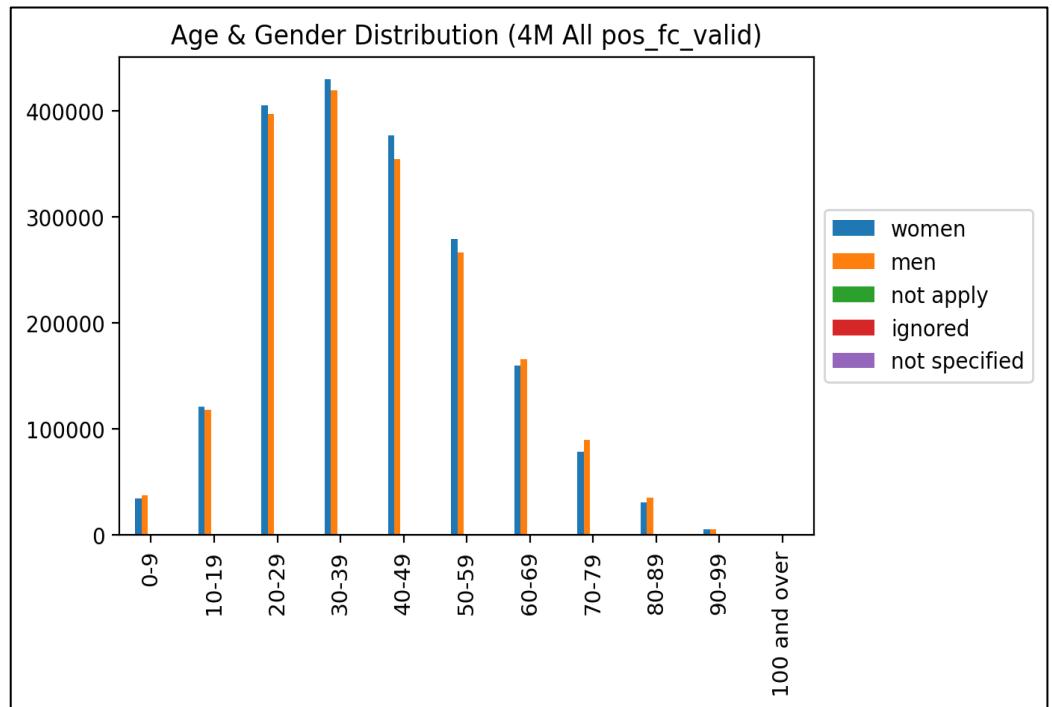
- 98: Αγνοείται ή τιμή (Ignored)
- 99: Αδιευκρίνιστο (Not Specified)
- οι κενές, χωρίς τιμή (NaN)

Έτσι ο τελικός αριθμός των δειγμάτων (dataset) κατέληξε να είναι τα **3.809.119** δείγματα, τα οποία και αποθηκεύτηκαν και αυτά με τη μορφή αρχείου csv το οποίο ονομάστηκε «Covid19MPD_7_23_en_pos_fc_valid.csv»

Ορισμένα από τα χαρακτηριστικά του συνόλου των δεδομένων όπως η κατανομή των τιμών για τα διάφορα χαρακτηριστικά και η κατανομή των ηλικιακών ομάδων απεικονίζονται στις παρακάτω εικόνες (**Εικόνα 52**, **Εικόνα 53**).



Εικόνα 52: Διάγραμμα κατανομής των διαφορετικών τιμών των 17 κατηγορικών χαρακτηριστικών του συνόλου δεδομένων df_23_en_pos_fc_valid. (Εξαιρέθηκαν τα χαρακτηριστικά: Φύλο [SEX], Ηλικία [AGE], LAB RESULT, FINAL CLASSIFICATION και DAYS FROM SYMPTOM TO HOSPITALIZATION)



Εικόνα 53: Διάγραμμα κατανομής των ηλικιακών ομάδων και του φύλου για το df_23_en_pos_fc_valid. Παρατηρούμε πως οι περισσότεροι ασθενείς είναι μεταξύ 20-59 ετών, επίσης τα Φύλα (Ανδρες, Γυναίκες) των ασθενών είναι σχεδόν ίσα για τις περισσότερες ηλικιακές ομάδες.

4.4 Μετασχηματισμός-Κωδικοποίηση

Στο συγκεκριμένο στάδιο τα δεδομένα, του δημιουργημένου αρχείου «**Covid19MPD_7_23_en_pos_fc_valid.csv**», πέρασαν από περαιτέρω επεξεργασία με την χρήση στατιστικών μεθόδων από την βιβλιοθήκη scikit-learn της python. Οι μέθοδοι που χρησιμοποιήθηκαν είναι οι παρακάτω:

- Η μέθοδος **LabelEncoder** της **sklearn**, η οποία χρησιμοποιείται σε κατηγορικά χαρακτηριστικά, για την κωδικοποίηση των τιμών-στόχων με τιμή μεταξύ **0** και **n-1**, όπου **n** είναι ο αριθμός των διαφορετικών τιμών για το συγκεκριμένο χαρακτηριστικό. Χρησιμοποιήθηκε στα χαρακτηριστικά (στήλες) **SEX**, **TYPE_OF_PATIENT**, **PNEUMONIA**, **DIABETIC**, **COPD**, **ASTHMA**, **IMMUNOSUPPRESSED**, **HYPERTENSION**, **OTHER_CHRONIC_DISEASE**, **CARDIOVASCULAR**, **OBESITY**, **CHRONIC_KIDNEY_FAILURE**, **SMOKER**, **CONTACT_WITH_COVID-19_CASE** και **SURVIVED**.
- Η μέθοδος **get_dummies** της **pandas**, η οποία χρησιμοποιείται για τη μετατροπή κατηγορικών χαρακτηριστικών σε εικονικές μεταβλητές ή μεταβλητές δείκτες. Χρησιμοποιήθηκε στα χαρακτηριστικά (στήλες) **INTUBATED**, **PREGNANCY** και **ICU** «σπάζοντας» έτσι στις αντίστοιχες **INTUBATED_2** (δεν διασωληνώθηκε), **INTUBATED_97** (δεν νοσηλεύτηκε, οπότε δεν μπορεί να εφαρμοστεί το κριτήριο για το αν διασωληνώθηκε), **PREGNANCY_2** (δεν είναι έγκυος), **PREGNANCY_97** (είναι άντρας, οπότε δεν μπορεί να εφαρμοστεί το κριτήριο για το αν είναι έγκυος), **ICU_2** (δεν εισήχθη σε Μ.Ε.Θ.) και **ICU_97** (δεν νοσηλεύτηκε, οπότε δεν μπορεί να εφαρμοστεί το κριτήριο για το αν εισήχθη σε Μ.Ε.Θ.), με την κάθε νέα στήλη να παίρνει τιμές 0 ή 1 με την διπλή άρνηση π.χ. **ICU_2=0** και **ICU_97=0** να υποδηλώνει πως εισήχθη στην εντατική.
- Η μέθοδος **StandardScaler** της **sklearn**, η οποία χρησιμοποιείται για τυποποιήσει τα χαρακτηριστικά αφαιρώντας τη μέση τιμή και κλιμακώνοντας τη διακύμανση μονάδας. Αυτό το επιτυγχάνει αντικαθιστώντας την τιμή του κάθε δείγματος με τη τυπική βαθμολογία (z), η οποία υπολογίζεται ως εξής: $z = (x - m) / std$ όπου m είναι ο μέσος όρος των δειγμάτων εκπαίδευσης(means) και std η τυπική απόκλιση των δειγμάτων εκπαίδευσης. Χρησιμοποιήθηκε ως μια από τις 2 εναλλακτικές στα χαρακτηριστικά (στήλες) **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION**.
- Η μέθοδος **MinMaxScaler** της **sklearn**, η οποία χρησιμοποιείται για το μετασχηματισμό τα αριθμητικών χαρακτηριστικών κλιμακώνοντας τις τιμές του κάθε χαρακτηριστικού σε ένα δεδομένο εύρος. Πιο συγκεκριμένα κλιμακώνει και μεταφράζει κάθε χαρακτηριστικό ξεχωριστά έτσι ώστε να βρίσκεται στο δεδομένο εύρος στο σετ εκπαίδευσης, π.χ. μεταξύ **0** και **1**. Η νέα τιμή (**X_scaled**) για το κάθε δείγμα υπολογίζεται από τον τύπο: $X_{scaled} = X_{std} * (max - min) + min$. Χρησιμοποιήθηκε ως μια από τις 2 εναλλακτικές στα χαρακτηριστικά (στήλες) **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** για 4 διαφορετικά min-max εύρη, **0-1,0-10,0-100** και **0-1000**.

Μετά και από την ολοκλήρωση της παραπάνω επεξεργασίας δημιουργήθηκαν **6** διαφορετικά αρχεία csv, ανάλογα με την μέθοδο μετασχηματισμού των χαρακτηριστικών AGE και DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION, τα ονόματα των οποίων είναι τα:

1. **Covid19MPD_8_23_en_pos_fc_valid_lb_none.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** δεν έχουν υποστεί κάποιο μετασχηματισμό.
2. **Covid19MPD_8_23_en_pos_fc_valid_lb_std.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** κωδικοποιήθηκαν με τη χρήση της μεθόδου **StandardScaler** της **sklearn**.
3. **Covid19MPD_8_23_en_pos_fc_valid_lb_mm_0-1.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** κωδικοποιήθηκαν με τη χρήση της μεθόδου **MinMaxScaler** της **sklearn**, χρησιμοποιώντας εύρη μεταξύ **0** και **1**.
4. **Covid19MPD_8_23_en_pos_fc_valid_lb_mm_0-10.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** κωδικοποιήθηκαν με τη χρήση της μεθόδου **MinMaxScaler** της **sklearn**, χρησιμοποιώντας εύρη μεταξύ **0** και **10**.
5. **Covid19MPD_8_23_en_pos_fc_valid_lb_mm_0-100.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** κωδικοποιήθηκαν με τη χρήση της μεθόδου **MinMaxScaler** της **sklearn**, χρησιμοποιώντας εύρη μεταξύ **0** και **100**.
6. **Covid19MPD_8_23_en_pos_fc_valid_lb_mm_0-1000.csv**: όπου οι στήλες **AGE** και **DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION** κωδικοποιήθηκαν με τη χρήση της μεθόδου **MinMaxScaler** της **sklearn**, χρησιμοποιώντας εύρη μεταξύ **0** και **1000**.

Η τελική Προεπεξεργασία των δεδομένων που εφαρμόστηκε σε κάθε ένα από τα 6 παραπάνω αρχεία για τη κάθε επανάληψη (**iteration**) του κάθε αλγόριθμου περιελάμβανε τα παρακάτω βήματα:

- **Βήμα 1º**: Αρχικά συντέθηκε ένα νέο σύνολο δεδομένων (**dataframe**) μέσω της τυχαίας επιλογής του 20% των δειγμάτων του αρχείου με την χρήση της συνάρτησης **sample()** της **pandas**.
- **Βήμα 2º**: Στο νεοσυσταθέν **dataframe** εφαρμόστηκε η μέθοδος της τυχαίας υπερδειγματοληψίας **SMOTE** (**Synthetic Minority Oversampling Technique**) [436] από την βιβλιοθήκη **imblearn** της **python**, ώστε να μετριαστεί η ανισορροπία στα δεδομένα καταλήγοντας σε **dataframe** με την τελική αναλογία **Αποθανόντων** προς τους **Επιζώντες** από COVID-19 να είναι **1:10**. Η **SMOTE** είναι μια από τις πιο συχνά χρησιμοποιούμενες μεθόδους υπερδειγματοληψίας για την επίλυση του προβλήματος της ανισορροπίας και στόχος της είναι η εξισορρόπηση της κατανομής 2 ή περισσοτέρων κλάσεων για 1 συγκεκριμένο χαρακτηριστικό, αυξάνοντας τυχαία τα δείγματα της μειοψηφούσας κλάσης με την αναπαραγωγή αντιγράφων από τα υπάρχοντα μειοψηφικά δείγματα.
- **Βήμα 3º**: Στο προκύπτον **dataframe** εφαρμόστηκε η μέθοδος τυχαίας υποδειγματοληψίας **RandomUnderSampler** από την βιβλιοθήκη **imblearn**, η οποία περιλαμβάνει την τυχαία διαγραφή, δειγμάτων της πλειοψηφικής κλάσης, από το

σύνολο δεδομένων, μέχρι να επιτευχθεί τελική αναλογία **1:2** μεταξύ **Θανόντων** προς **Επιζώντες**.

- **Βήμα 4^o:** Τέλος το παραπάνω dataframe χωρίστηκε σε **2** υποσύνολα το υποσύνολο εκπαίδευσης (**train**) και αυτό της ελέγχου- δοκιμής (**test**), με το κλάσμα των δεδομένων εκπαίδευσης (**train**) του δείγματος να είναι **0.7 (70%)** και άρα το κλάσμα των δεδομένων ελέγχου- δοκιμής (**test**) του δείγματος να είναι **0.3 (30%)**. Η διαδικασία έγινε με τη χρήση της συνάρτησης **train_test_split()** της **sklearn**.

4.5 Μοντέλα και Αλγόριθμοι

Στη μελέτη, χρησιμοποιήθηκαν 6 αλγόριθμοι ταξινόμησης των βιβλιοθηκών της Python:

1. **Η Λογιστική Παλινδρόμηση** (Logistic Regression, **LR**).
2. **Τα Δέντρα Απόφασης** (Decision Trees, **DT**).
3. Ο αλγόριθμος του **Τυχαίου Δάσους** (Random Forest, **RF**), που είναι αλγόριθμος Συνόλου (**Ensemble**).
4. Ο αλγόριθμος της **Ακραίας Ενίσχυσης Κλίσης** (eXtreme Gradiend Boosting, **XGB**), που είναι και αυτός αλγόριθμος Συνόλου (**Ensemble**).
5. **Τα Τεχνητά Νευρωνικά Δίκτυα (TNΔ) Πολυστρωματικών Αντιλήπτρων** (Multi-Layer Perceptrons, **MLPs**).
6. Ο αλγόριθμος των **K Εγγύτερων Γειτόνων** (K-Nearest Neighbors, **KNN**).

Τα μοντέλα υλοποιήθηκαν σε γλώσσα **Python (3.7)** χρησιμοποιώντας το λογισμικό (IDE) **Spyder (5.1.5.)** και τη βιβλιοθήκη **pandas (1.3.5)**.

Δημιουργήθηκαν **3** ξεχωριστά σετ χαρακτηριστικών για τον κάθε αλγόριθμο, ανάλογα με την βαθμολογία της σημαντικότητας του χαρακτηριστικού:

1. Το σετ που περιελάμβανε όλα τα **22** χαρακτηριστικά.
2. Το σετ που είχε μόνο τα κορυφαία **15** χαρακτηριστικά.
3. Το σετ που είχε μόνο τα κορυφαία **10** χαρακτηριστικά.

Η βαθμολόγηση της σημαντικότητας για το κάθε χαρακτηριστικό, υπολογίστηκε ως το σύνολο των επιμέρους βαθμολογίων που αυτό έλαβε για κάθε ένα από τους **6** διαφορετικούς τύπους κωδικοποίησης με την κατάταξη των χαρακτηριστικών να διαφέρει μεταξύ των αλγορίθμων.

Επίσης για τον κάθε αλγόριθμο χρησιμοποιήθηκαν **3** διαφορετικά σετ υπερπαραμέτρων:

1. Οι **προκαθορισμένες** (default) τιμές των υπερπαραμέτρων
2. Το **1^o** σετ **βέλτιστοποιημένων** (optimal_01) τιμών των υπερπαραμέτρων.
3. Το **2^o** σετ **βέλτιστοποιημένων** (optimal_02) τιμών των υπερπαραμέτρων.

Για την εύρεση των **2** σετ βέλτιστοποιημένων υπερπαραμέτρων εφαρμόστηκε η μέθοδος της αναζήτησης πλέγματος **GridSearchCV()** της **sklearn** για την εύρεση των βέλτιστων τιμών για τις περισσότερες υπερπαραμέτρους. Η μέθοδος αυτή χρησιμεύει στην αναζήτηση της βέλτιστης τιμής της κάθε υπερπαραμέτρου, μέσα από ένα δεδομένο πλέγμα (**grid**), το οποίο

περιέχει όλα τα σύνολα των τιμών των διαφορετικών υπερπαραμέτρων. Αποτελεί μια μέθοδο διασταυρούμενης επικύρωσης **10 τμημάτων** (10-fold cross validation), στην συγκεκριμένη περίπτωση, η οποία δέχεται ως είσοδο τον εκάστοτε αλγόριθμο (μοντέλο) και το σύνολο των τιμών των διαφορετικών υπερπαραμέτρων και εξάγει την βέλτιστη τιμή για την κάθε υπερπαραμέτρο.

Έτσι καταλήξαμε σε **9** διαφορετικούς συνδυασμούς (3 σετ χαρακτηριστικών x **3** σετ υπερπαραμέτρων) που εκτελέστηκαν **10** φορές (iterations) για καθένα από τα **6** αρχεία csv, καταλήγοντας έτσι σε **540** διαφορετικές εκτελέσεις (iterations) για τον κάθε αλγόριθμο.

4.5.1 Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση (Logistic Regression, **LR**) αναπτύχθηκε από τον David Cox το 1958 [478] και αποτελεί μια πιο γενικευμένη μορφή της απλής Γραμμικής Παλινδρόμησης. Η LR είναι παρόμοια με τη Γραμμική Παλινδρόμηση εκτός από τον τρόπο χρήσης τους, καθώς η Γραμμική Παλινδρόμηση στοχεύει στην επίλυση προβλημάτων Παλινδρόμησης π.χ. την πρόβλεψη της αυριανής θερμοκρασίας, ενώ η LR στοχεύει στην επίλυση προβλημάτων ταξινόμησης. Στην LR, εν αντιθέσει με την Γραμμική Παλινδρόμηση πέραν από τις αριθμητικές, μπορούν να χρησιμοποιηθούν και κατηγορικές μεταβλητές, ως εξαρτημένες μεταβλητές.

Χρησιμοποιείται κυρίως όταν χρειάζεται να προβλέψουμε την εμφάνιση ή όχι ενός χαρακτηριστικού ή την πραγματοποίηση ή όχι ενός γεγονότος. Πιο συγκεκριμένα βασική της χρήση είναι η πρόβλεψη της τιμή μιας δυαδικής ή κατηγορικής εξαρτημένης μεταβλητής, ως συνάρτηση του ποσοστού συνεισφοράς της κάθε εξαρτημένης μεταβλητής στην διαμόρφωση της τελικής τιμής της εξαρτημένης, αναδεικνύοντας έτσι τις ανεξάρτητες μεταβλητές που εμφανίζουν την μεγαλύτερη συσχέτιση με την εξαρτημένη.

Η πιο διαδεδομένη, έκφραση της εξίσωσης της LR είναι [478]:

$$\ln(\text{odds}) = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

με **odds** = $p/(1-p)$

όπου:

- **a** είναι το ύψος της κλίσης της γραμμής παλινδρόμησης
- **b_1, \dots, b_k** είναι συντελεστές (coefficients) παλινδρόμησης με τον καθένα να εκφράζει το μέγεθος της συμβολής (σημαντικότητα) της κάθε μεταβλητής, με την θετική τιμή να δείχνει πως η μεταβλητή αυξάνει την πιθανότητα της εμφάνισης του γεγονότος και την αρνητική τιμή να δείχνει μείωση της πιθανότητας εμφάνισης. Επιπλέον η υψηλή τιμή του συντελεστή, συνεπάγεται πως η ανεξάρτητη μεταβλητή επηρεάζει πολύ ισχυρά την πιθανότητα εμφάνισης του γεγονότος, με την χαμηλή τιμή να δείχνει μικρή ή ελάχιστη επίδραση στην εμφάνιση του γεγονότος. Οι συντελεστές των ανεξάρτητων μεταβλητών υπολογίζονται σύμφωνα με τη μέθοδο Μεγίστης Πιθανοφάνειας.

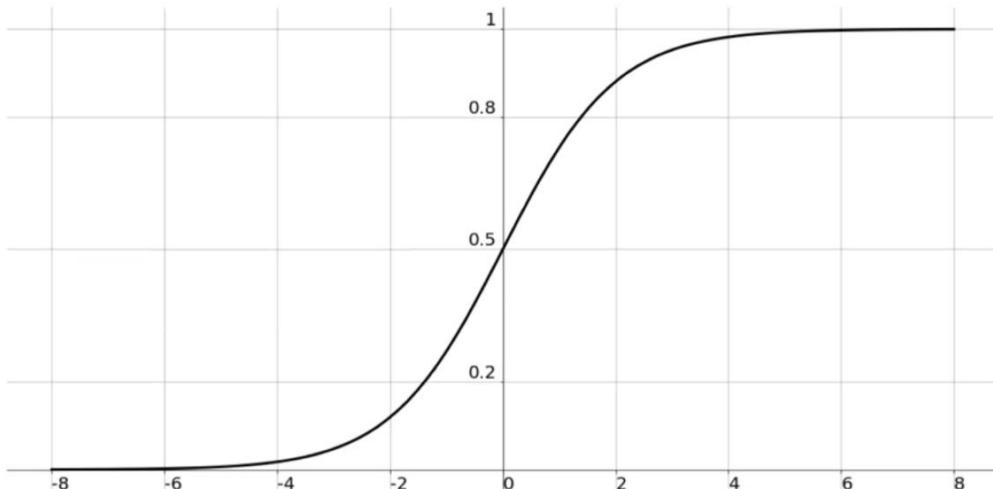
Το δεξί μέρος της εξίσωσης αποτελεί το σύνολο του γραμμικού συνδυασμού των ανεξάρτητων

μεταβλητών που συμμετέχουν στο μοντέλο παλινδρόμησης, ενώ το αριστερό μέρος περιλαμβάνει τις τιμές της εξαρτημένης μεταβλητής, με μορφή του λογαρίθμου της πιθανότητας (odds), για το αν θα συμβεί (p) ή όχι ($1-p$) κάποιο γεγονός.

Η LR χρησιμοποιείται σε αρκετούς τομείς, όπως στις στατιστικές και κοινωνικές επιστήμες, σε αρκετά ιατρικά πεδία καθώς και στον τομέα της Μηχανικής Μάθησης. Ειδικότερα στον

τομέα της Μηχανικής Μάθησης χρησιμοποιείται για δημιουργία ενός μοντέλου παλινδρόμησης με σκοπό την πρόβλεψη του αν μια δεδομένη καταχώριση δεδομένων (δείγμα) ανήκει στην κατηγορία που αριθμείται ως "1" ή όχι. Η LR μοντελοποιεί τα δεδομένα (συνεχή και διακριτά) με τη χρήση της σιγμοειδούς συνάρτησης μέσω της οποίας πραγματοποιούνται και οι προβλέψεις για το ποιο από τα δύο πιθανά αποτελέσματα θα συμβεί (0 ή 1) [479].

Ο τύπος της σιγμοειδούς συνάρτησης (**Εικόνα 54**) είναι: $f(x) = 1/(1+e^{-x})$



Εικόνα 54: Η σιγμοειδής καμπύλη [480]

Η LR βρίσκει εφαρμογές σε ευρύ φάσμα τομέων και πεδίων, ορισμένα παραδείγματα είναι και τα παρακάτω:

➤ Εκπαίδευση:

- Ως κριτήριο για το αν κάποιος-α φοιτητής-τρια θα εισαχθεί σε ένα πανεπιστημιακό πρόγραμμα ή όχι, με βάσει τις βαθμολογίες του σε συγκεκριμένα μαθήματα.
- Για το αν κάποιος-α μαθητής-τρια θα προλάβει να ολοκληρώσει την εργασία που του έχει ανατεθεί, με βάσει κυρίως τις προηγούμενες δραστηριότητες και άλλα στατιστικά στοιχεία.

➤ Ιατρική:

- Στην πρόβλεψη το αν ένα άτομο έχει ή όχι, κάποια ασθένεια, με βάσει τις τιμές που έχει σε κάποιους βιολογικούς δείκτες, με βάσει τις τιμές αναφοράς από πληθώρα δειγμάτων θετικών και μη, για την συγκεκριμένη ασθένεια.
- Στην πρόβλεψη για το αν κάποιος πάσχει από COVID-19 με σημείο αναφοράς, ακτινογραφίες ασθενών και υγειών ατόμων.

➤ Άλλες εφαρμογές:

- Στην πρόβλεψη για το αν μια συναλλαγή μέσω πιστωτικής κάρτας που είναι παράνομη ή όχι.

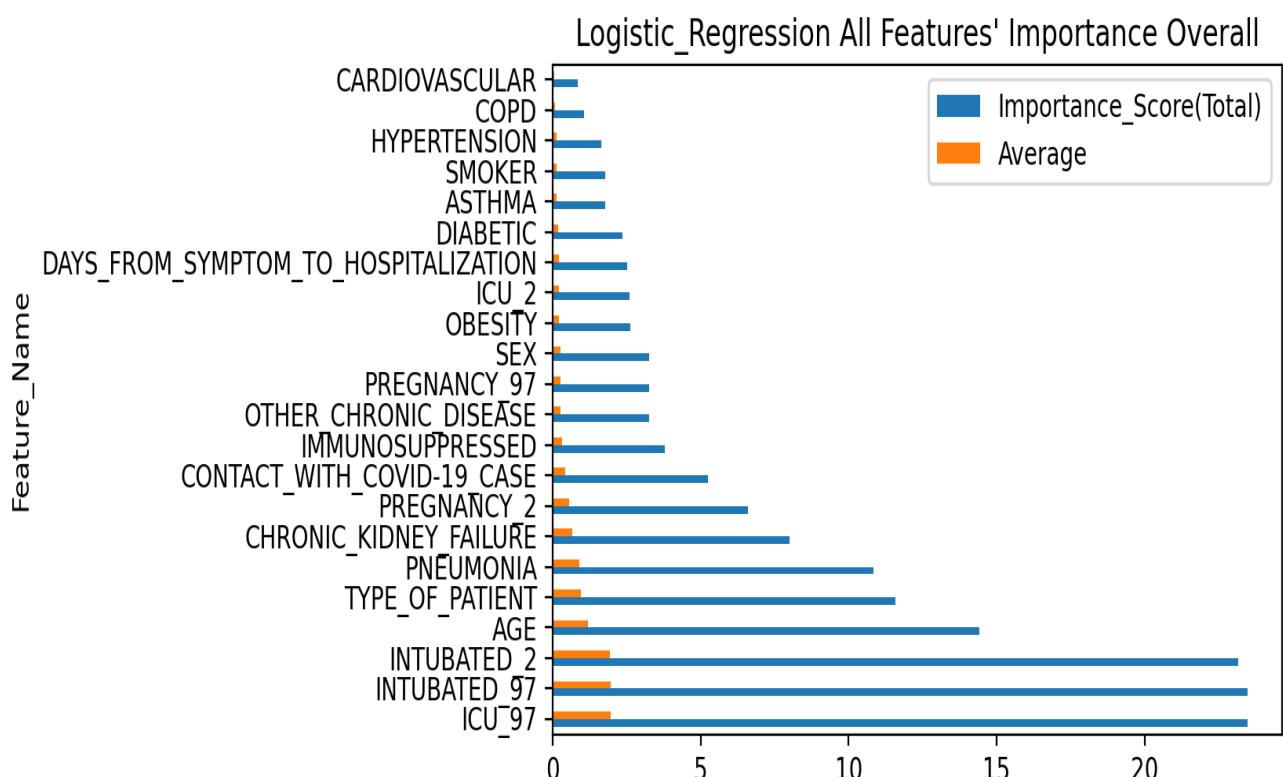
- Στην ανάλυση συναισθήματος, ώστε να διαπιστωθεί αν κάποιος συνομιλητής είναι χαρούμενος, λυπημένος, οργισμένος κ.α.
- Στην ανίχνευση και ταξινόμηση αντικειμένων, όπως π.χ. την κατηγοριοποίηση των εικόνων που ανήκουν σε διαφορετικά είδη φυτών.

Στην παρούσα εργασία χρησιμοποιήθηκε η **LogisticRegression** της βιβλιοθήκης **sklearn**. Σε ένα **LR** μοντέλο είναι δυνατόν να ανακτηθούν οι συντελεστές σημαντικότητας που βρέθηκαν για κάθε χαρακτηριστικό μέσω της `coeff_` ιδιότητας η οποία περιέχει τους συντελεστές που βρέθηκαν για κάθε χαρακτηριστικό και να το εξάγουμε σε μορφή πίνακα (`feat_importances`), με τη χρήση του παρακάτω κώδικα:

```
feat_importances = pd.Series(model.coef_[0], index=features.columns)
```

Οι συντελεστές είναι θετικοί και αρνητικοί. Οι θετικές βαθμολογίες υποδεικνύουν ένα χαρακτηριστικό που προβλέπει την κατηγορία 1, ενώ οι αρνητικές βαθμολογίες δείχνουν ένα χαρακτηριστικό που προβλέπει την κατηγορία 0.

Έτσι αφού υπολογίσαμε τους συντελεστές σημαντικότητας του κάθε χαρακτηριστικού, για κάθε ένα από τα 6 διαφορετικά κωδικοποιημένα σύνολα δεδομένων, των αρχείων csv, ώστε να έχουν την ίδια κωδικοποίηση, προσθέσαμε τις απόλυτες τιμές τους και καταλήξαμε στο παρακάτω Διάγραμμα (**Εικόνα 55**).



Εικόνα 55: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο της **Λογιστικής Παλινδρόμησης** (LR).

Πιο αναλυτικά, για την **Λογιστική Παλινδρόμηση**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 24**.

Πίνακας 24: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για την LR.

1	ICU_97	23,464210
2	INTUBATED_97	23,464210
3	INTUBATED_2	23,161639
4	AGE	14,425526
5	TYPE_OF_PATIENT	11,591325
6	PNEUMONIA	10,828989
7	CHRONIC_KIDNEY_FAILURE	8,007899
8	PREGNANCY_2	6,608863
9	CONTACT_WITH_COVID-19_CASE	5,238136
10	IMMUNOSUPPRESSED	3,799076
11	OTHER_CHRONIC_DISEASE	3,260162
12	PREGNANCY_97	3,255442
13	SEX	3,255442
14	OBESITY	2,638171
15	ICU_2	2,602937
16	DAYs_FROM_SYMPTOM_TO_HOSPITALIZATION	2,511025
17	DIABETIC	2,353492
18	ASTHMA	1,782358
19	SMOKER	1,775976
20	HYPERTENSION	1,642227
21	COPD	1,072396
22	CARDIOVASCULAR	0,851999

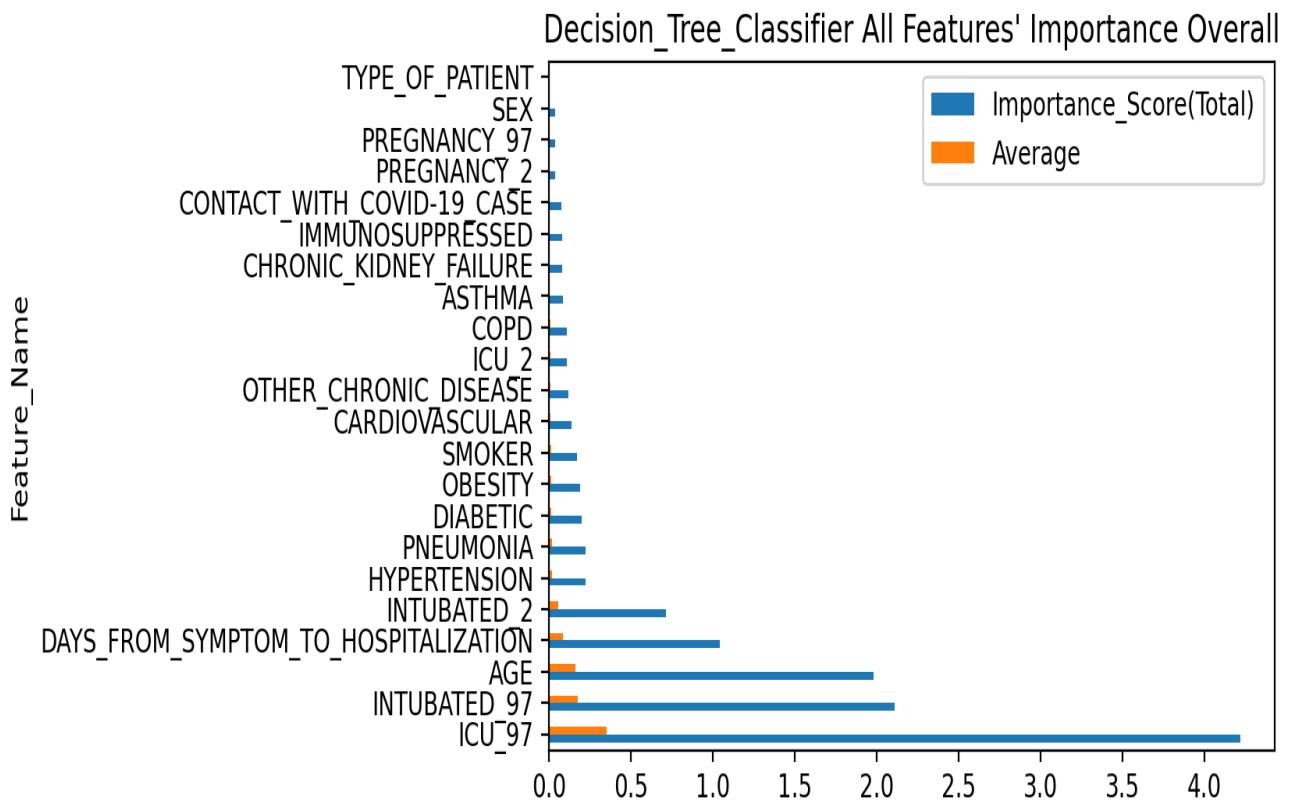
4.5.2 Δέντρα Απόφασης

Τα Δέντρα Αποφάσεων (Decision Trees, DT), όπως αναλύθηκε και σε προηγούμενη ενότητα, είναι μια μη παραμετρική εποπτευόμενη μέθοδος μάθησης που χρησιμοποιείται σε προβλήματα ταξινόμησης ή παλινδρόμησης. Κύριος στόχος τους είναι η δημιουργία ενός μοντέλου που να προβλέπει την τιμή μιας μεταβλητής στόχου, μαθαίνοντας απλούς κανόνες απόφασης που συνάγονται από τα χαρακτηριστικά δεδομένων. Ένα δέντρο μπορεί να θεωρηθεί ως μια τμηματικά σταθερή προσέγγιση.

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής **DecisionTreeClassifier** της βιβλιοθήκης **sklearn**. Στο DT μοντέλο του αφού αυτό προσαρμοστεί, είναι δυνατόν να ανακτηθούν, σε μορφή πίνακα (**feat_importances**), οι συντελεστές σημαντικότητας που βρέθηκαν για κάθε χαρακτηριστικό μέσω της ιδιότητας `feature_importances_`, με τη χρήση του παρακάτω κώδικα:

```
feat_importances = pd.Series(model.feature_importances_, index=features.columns)
```

Εδώ Οι συντελεστές είναι μόνο θετικοί αριθμοί. Στη συνέχεια υπολογίστηκαν οι συντελεστές σημαντικότητας του κάθε χαρακτηριστικού, για κάθε ένα από τα 6 διαφορετικά κωδικοποιημένα σύνολα δεδομένων, των αρχείων csv, ώστε να έχουν την ίδια κωδικοποίηση, και προστέθηκαν οι απόλυτες τιμές τους, καταλήγοντας στο παρακάτω Διάγραμμα (**Εικόνα 56**).



Εικόνα 56: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο των **Δέντρων Απόφασης** (DT).

Πιο αναλυτικά, για τον ταξινομητή των **Δέντρων Απόφασης**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 25**.

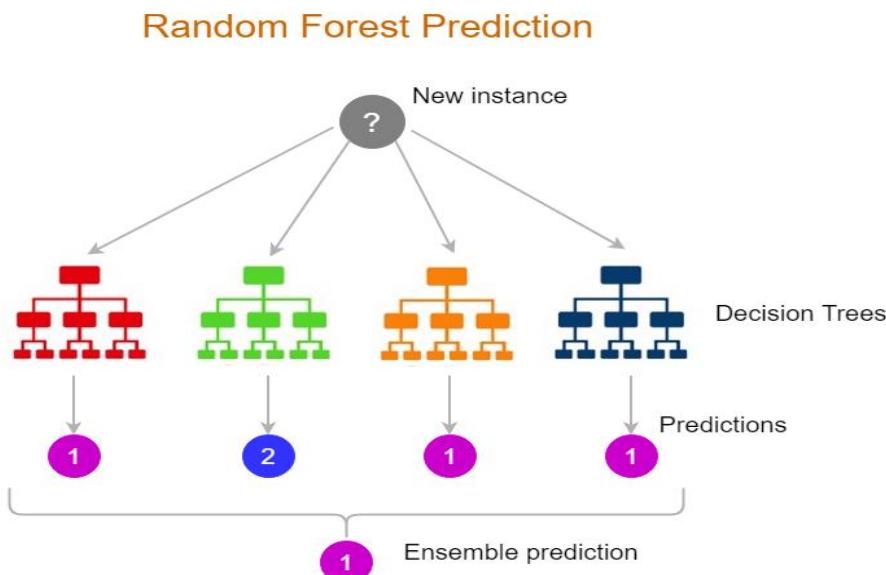
Πίνακας 25: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για τα DTs.

1	ICU_97	4,219194
2	INTUBATED_97	2,109597
3	AGE	1,981054
4	DAYs_FROM_SYMPTOM_TO_HOSPITALIZATION	1,043425
5	INTUBATED_2	0,714463
6	HYPERTENSION	0,227111
7	PNEUMONIA	0,225179
8	DIABETIC	0,198828
9	OBESITY	0,192870
10	SMOKER	0,170398
11	CARDIOVASCULAR	0,136990
12	OTHER_CHRONIC_DISEASE	0,122133
13	ICU_2	0,109808
14	COPD	0,108379
15	ASTHMA	0,084368
16	CHRONIC_KIDNEY_FAILURE	0,083769
17	IMMUNOSUPPRESSED	0,082815
18	CONTACT_WITH_COVID-19_CASE	0,076441
19	PREGNANCY_2	0,038269
20	PREGNANCY_97	0,038117
21	SEX	0,036790
22	TYPE_OF_PATIENT	0,000000

4.5.3 Τυχαίο Δάσος

Το **Τυχαίο Δάσος** (Random Forest, RF) αποτελεί μια μέθοδο μηχανικής εποπτευόμενης μάθησης που μπορεί να εφαρμοστεί τόσο στην ταξινόμηση όσο και σε προβλήματα παλινδρόμησης. Στηρίζεται στην έννοια της μάθησης συνόλου (Ensemble), σύμφωνα με την οποία χρησιμοποιείται ένας συνδυασμός πολλαπλών ταξινομητών με σκοπό την επίλυση για την επίλυση ενός σύνθετου προβλήματος και για τη βελτίωση της απόδοσης του μοντέλου [481].

Πιο συγκεκριμένα, το RF είναι ένας ταξινομητής ο οποίος εμπεριέχει έναν αριθμό (n) δέντρων αποφάσεων, το καθένα από τα οποία έχει εκπαιδευτεί σε διαφορετικό υποσύνολο του συνόλου εκπαίδευσης. Η τελική απόφαση (πρόβλεψη) να λαμβάνεται από την πλειοψηφία, για τις προβλέψεις κατηγορικών μεταβλητών ή από το μέσο όρο των τιμών για τις προβλέψεις των αριθμητικών μεταβλητών, έτσι ώστε να ενισχύσει την ακρίβεια του εκπαίδευσης μοντέλου. Όσο μεγαλύτερος ο αριθμός των δέντρων αποφάσεων από τα οποία απαρτίζεται ένα δάσος, τόσο μεγαλύτερη και η ακρίβεια του μοντέλου και τόσο μικρότερος ο κίνδυνος για υπερπροσαρμογή. Στην παρακάτω εικόνα (Εικόνα 57) απεικονίζεται το διάγραμμα λειτουργίας του RF.



Εικόνα 57: Το διάγραμμα λειτουργίας των προβλέψεων του **Τυχαίου Δάσους** (RF). [482]

Το πρωτόκολλο του RF περιέχει 2 φάσεις, με την πρώτη να είναι η δημιουργία ενός τυχαίου δάσους συνδυάζοντας ένα αριθμό N από DT και τη δεύτερη να περιλαμβάνει τις προβλέψεις του κάθε DT, που δημιουργήθηκε στην πρώτη φάση. Πιο αναλυτικά τα βήματα που ακολουθεί το RF είναι τα παρακάτω [481]:

- Βήμα 1** Δημιουργία σετ Εκπαιδεύσεως και σετ Δοκιμής (προβλέψεων)
- Βήμα 2** Επιλογή του αριθμού N των DT που θα απαρτίζουν το RF.
- Βήμα 3** Τυχαία επιλογή K δειγμάτων από το σετ εκπαίδευσης.
- Βήμα 4** Επανάληψη του **Βήματος_3**, N φορές, μια για κάθε DT
- Βήμα 5** Δημιουργία και εκπαίδευση του κάθε DT στο αντίστοιχο για αυτό σετ.
- Βήμα 6** Δημιουργία Προβλέψεων από το κάθε DT για τα δείγματα του σετ Δοκιμής.
- Βήμα 7** Δημιουργία Πρόβλεψης με βάσει την πλειοψηφία ή τον μέσο όρο των DT.

Έτσι αφού το **RF** συνδυάζει πολλά δέντρα για να πραγματοποιήσει τις προβλέψεις, είναι πιθανό κάποια δέντρα απόφασης να προβλέψουν τη σωστή έξοδο, ενώ άλλα όχι, αλλά όλα μαζί ως σύνολο δύνανται να εμφανίσουν υψηλή ακρίβεια πρόβλεψης.

Περιπτώσεις που συμβάλει σημαντικά η χρήση του **RF** [481]:

- Η Δυνατότητα εκτέλεσης εργασιών Ταξινόμησης και Παλινδρόμησης.
- Η Διαχείριση μεγάλων συνόλων δεδομένων με υψηλή διαστατικότητα.
- Η Διατήρηση της Ακρίβειας των Προβλέψεων, ακόμα και όταν λείπει μεγάλο ποσοστό δεδομένων, αποτρέπει σε μεγάλο βαθμό την υπερπροσαρμογή.

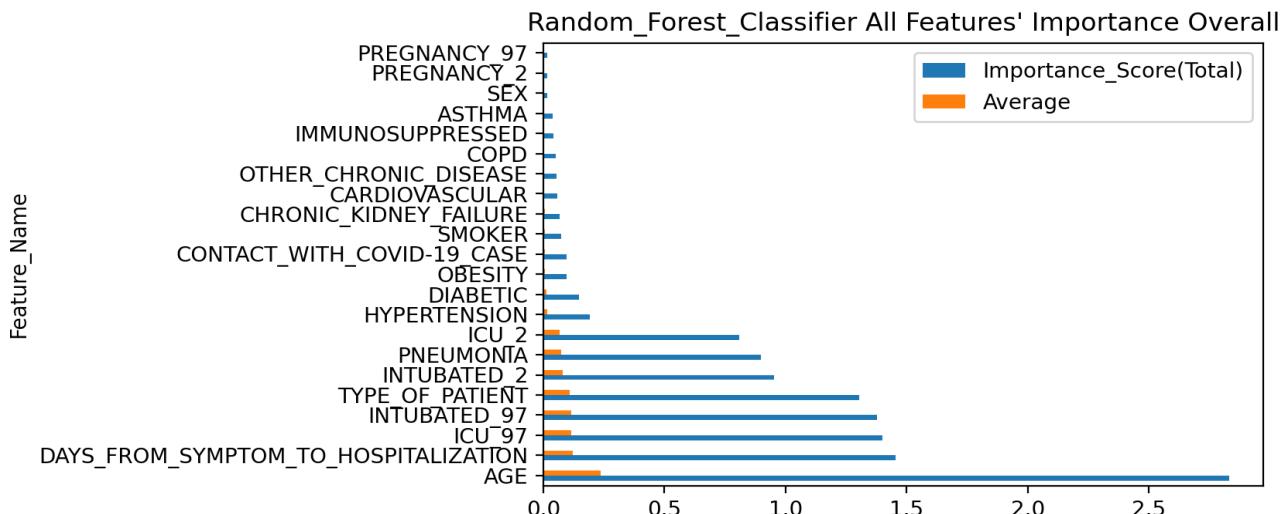
Το **RF** χρησιμοποιείται περισσότερο στους παρακάτω τομείς [481]:

- Στον Τραπεζικό τομέα χρησιμοποιείται, για τον προσδιορισμό του δανειακού κινδύνου
- Στον τομέα της Ιατρικής χρησιμοποιείται για τον εντοπισμό των πασχόντων, από μια νόσο, των επιπλοκών της νόσου κ.α.
- Στον τομέα του Μάρκετινγκ μπορεί να χρησιμοποιηθεί για τον εντοπισμό τάσεων μάρκετινγκ.

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής **RandomForestClassifier** της βιβλιοθήκης **sklearn**. Στο **RF** μοντέλο, όπως και με τα **DT**, αφού προσαρμοστεί, είναι δυνατόν να ανακτηθούν, σε μορφή πίνακα (**feat_importances**), οι συντελεστές σημαντικότητας που βρέθηκαν για κάθε χαρακτηριστικό μέσω της ιδιότητας **feature_importances_**, με τη χρήση του παρακάτω κώδικα:

```
feat_importances = pd.Series(model.feature_importances_, index=features.columns)
```

Και εδώ Οι συντελεστές είναι μόνο θετικοί αριθμοί. Αφού υπολογίστηκαν οι συντελεστές σημαντικότητας του κάθε χαρακτηριστικού, για κάθε ένα από τα 6 διαφορετικά κωδικοποιημένα σύνολα δεδομένων, των αρχείων csv, ώστε να έχουν την ίδια κωδικοποίηση, και προστέθηκαν οι απόλυτες τιμές τους, καταλήγοντας στο παρακάτω Διάγραμμα (**Εικόνα 58**).



Εικόνα 58: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο του **Τυχαίου Δάσους** (RF).

Πιο αναλυτικά, για το **Τυχαίο Δάσος**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 26**.

Πίνακας 26: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για το RF.

1	AGE	2,829940
2	DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION	1,453450
3	ICU_97	1,399297
4	INTUBATED_97	1,376842
5	TYPE_OF_PATIENT	1,303206
6	INTUBATED_2	0,953464
7	PNEUMONIA	0,899134
8	ICU_2	0,809578
9	HYPERTENSION	0,191666
10	DIABETIC	0,147699
11	OBESITY	0,096824
12	CONTACT_WITH_COVID-19_CASE	0,095596
13	SMOKER	0,074086
14	CHRONIC_KIDNEY_FAILURE	0,069094
15	CARDIOVASCULAR	0,059447
16	OTHER_CHRONIC_DISEASE	0,056126
17	COPD	0,050428
18	IMMUNOSUPPRESSED	0,042718
19	ASTHMA	0,038430
20	SEX	0,017956
21	PREGNANCY_2	0,017754
22	PREGNANCY_97	0,017264

4.5.4 Ακραία Ενίσχυση Κλίσης

Η Ακραία Ενίσχυση Κλίσης (eXtreme Gradient Boosting, **XGB**) είναι μια από τις πιο γνωστές παραλλαγές του αλγορίθμου Ενίσχυσης Κλίσης (Gradient Boosting Algorithm), αλγορίθμου ο οποίος αναπτύχθηκε με στόχο την αύξηση της ακρίβειας πρόβλεψης, του οποίου αρχικά η χρήση ήταν περιορισμένη καθώς απαιτούσε μεγάλο χρονικό διάστημα για την εκπαίδευση μοντέλων, ακόμα και αν αυτά ήταν μικρά σε μέγεθος. Στην **XGB** δημιουργούνται μεμονωμένα **DTs** με την χρήση πολλαπλών πυρήνων, παράλληλα με την οργάνωση των δεδομένων έτσι ώστε να μειωθούν στο ελάχιστο οι χρόνοι αναζήτησης. Αποτέλεσμα αυτού ήταν η μείωση του χρόνου εκπαίδευσης των μοντέλων, αυξάνοντας έτσι την απόδοση [483].

Η **XGB** κατάφερε να βελτιώσει την ταχύτητα και την απόδοση των αλγορίθμων που βασίζονται στα **DTs**. Η μέθοδος αρχικά αναπτύχθηκε από τον Tianqi Chen και περιγράφηκε από το ίδιο και τον Carlos Guestrin το 2016 [384], αποτελώντας σήμερα τον πιο διαδεδομένο αλγόριθμο της Μηχανικής Μάθησης (ML), κερδίζοντας μεγάλη δημοτικότητα σε διαγωνισμούς ML [384]. Ανήκει στην κατηγορία των μεθόδων Συνόλων Μάθησης (Ensemble Learning), με βάση τα **DTs** πάνω στα οποία χρησιμοποιείται ένα πλαίσιο Ενίσχυσης Κλίσης. Στην τεχνική Ενίσχυσης τα λάθη τα οποία έγιναν στα μοντέλα που προηγήθηκαν συμβάλουν στην διόρθωση των επόμενων μοντέλων, τροποποιώντας κάποια βάρη των εισόδων.

Η διαφοροποίηση της XGB, σε σχέση με τις υπόλοιπες τεχνικές Μηχανών Ενίσχυσης Κλίσης, είναι ότι αυτή επιτυγχάνει την βελτίωση της ακρίβειας των προβλέψεων, μέσω των παρακάτω τεχνικών βελτιστοποίησης των συστημάτων και των τεχνικών βελτιστοποίησης των αλγορίθμικών λειτουργειών [384]:

Οι τεχνικές βελτιστοποίησης των συστημάτων που χρησιμοποιεί η XGB είναι [384]:

- **Η Παραλληλοποίηση:** Εδώ η **XGB** εφαρμόζει την παράλληλη αντί για τη διαδοχική κατασκευή των **DTs**, το όποιο καθίσταται εφικτό από την ίδια τη φύση των βρόχων των **DTs**, καθώς αυτοί μπορούν να εναλλάσσονται κατά την σύσταση των βασικών DT μαθητών. Εδώ η βελτίωση του χρόνου εκτέλεσης επιτυγχάνεται, με τη εναλλαγή της σειράς των βρόγχων χρησιμοποιώντας την αρχικοποίηση από την καθολική σάρωση όλων των πιθανών περιπτώσεων και την ταξινόμηση τους με την χρήση παράλληλων νημάτων (**multi-threads**). Κατ' αυτόν τον τρόπο επιτυγχάνεται βελτίωση της απόδοσης της XGB, αντισταθμίζοντας οποιαδήποτε κακή χρήση της παραλληλοποίησης που μπορεί να έγινε κατά τον εκτέλεση της.
- **Το Κλάδεμα (Pruning) Δέντρων:** Όπου η XGB χρησιμοποιεί πρώτα την παράμετρο «μέγιστου βάθους» (max-depth), έναντι του κριτήριου διακοπής και για γι' αυτό και εκκινεί το κλάδεμα των δέντρων προς τα πίσω. Η τεχνική αυτή της προσέγγισης πρώτα με βάσει το βάθος βελτιώνει σημαντικά την υπολογιστική απόδοση.
- **Η Βελτιστοποίηση χρήσης του Hardware:** Η XGB σχεδιάστηκε ώστε να χρησιμοποιεί αποτελεσματικότερα το hardware των μηχανών που εκτελείται. Αυτό το πετυχαίνει με την με την σωστή αναγνώριση της χωρητικότητας τη RAM, καθώς και με την κατανομή εσωτερικών «buffer» σε κάθε νήμα (thread) όπου αποθηκεύονται οι στατιστικές διαβαθμίσεις. Άλλη μια τεχνική βελτίωσης που χρησιμοποιεί η XGB είναι και ο «εκτός πυρήνα υπολογισμός» (out-of-core) καθώς βελτιστοποιούν την χρήση του διαθέσιμου αποθηκευτικού χώρου του δίσκου, έτσι ώστε να είναι εύκολη η διαχείριση μεγάλων συνόλων δεδομένων τα οποία δεν μπορεί να υποστηρίξει η RAM.

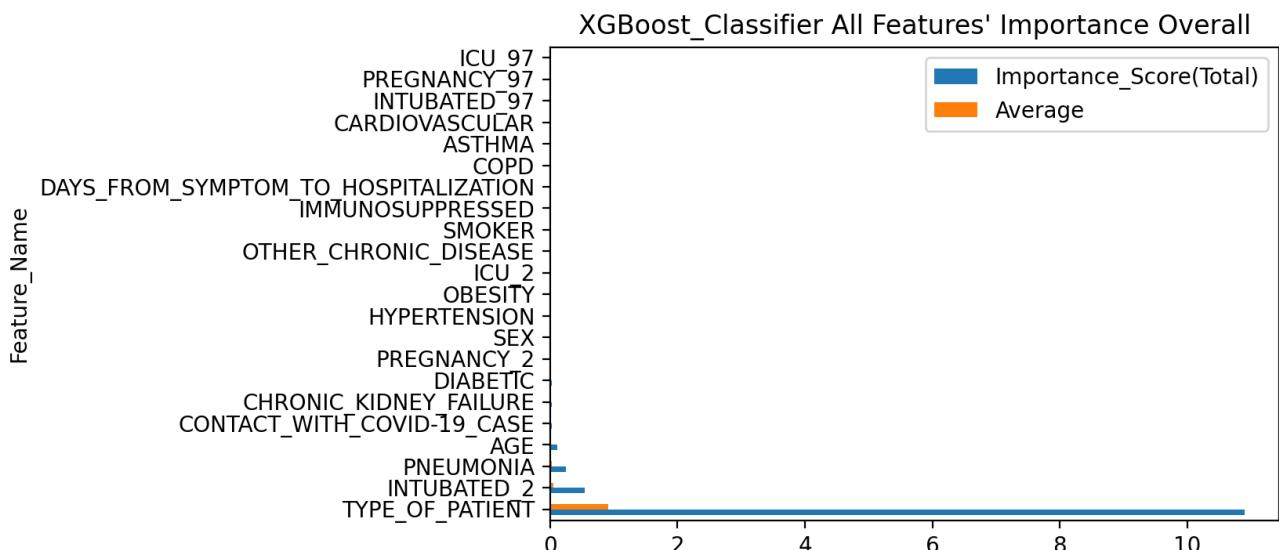
Οι αλγορίθμικές τεχνικές βελτιστοποίησης των συστημάτων που χρησιμοποιεί η XGB είναι [384]:

- Η Κανονικοποίηση, όπου μέσω αυτής επιβάλλονται κυρώσεις στα πολύπλοκα μοντέλα μέσω των μεθόδων συρρίκνωσης των συντελεστών μοντέλου Ridge και LASSO με σκοπό την αποφυγή της υπερπροσαρμογής (overfitting).
- Οι δυνατότητα χειρισμού δεδομένων με Ελλιπείς τιμές, καθώς η XGB μπορεί να δεχθεί δείγματα με ελλιπή χαρακτηριστικά και να «μάθει» αυτόματα την καλύτερη τιμή, σύμφωνα με την απώλεια κατάρτισης. Έτσι καταφέρνει τον βέλτιστο χειρισμό διαφορετικών τύπων ελλιπών μοτίβων που μπορεί να εντοπιστούν στα δεδομένα.
- Η Διασταυρούμενη Επικύρωση, καθώς η XGB διαθέτει ενσωματωμένη μέθοδο Διασταυρούμενης Επικύρωσης για κάθε επανάληψη, χωρίς έτσι να χρειάζεται η ανάγκη για τον προγραμματισμό και παραμετροποίησης αυτής της από τον χρήστη.

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής **XGBClassifier** της βιβλιοθήκης **xgboost**. Στο **XGB** μοντέλο του αφού αυτό προσαρμοστεί, είναι δυνατόν να ανακτηθούν οι συντελεστές σημαντικότητας που βρέθηκαν για κάθε χαρακτηριστικό μέσω της ιδιότητας `feature_importances_`, μέσω της οποίας είναι δυνατόν να ανακτηθούν οι σχετικοί βαθμοί σημαντικότητας για κάθε χαρακτηριστικό, με τη χρήση του παρακάτω κώδικα:

```
feat_importances = pd.Series(model.feature_importances_, index=features.columns)
```

Αφού όπως και παραπάνω υπολογίστηκαν οι συντελεστές σημαντικότητας του κάθε χαρακτηριστικού, για κάθε ένα από τα 6 διαφορετικά κωδικοποιημένα σύνολα δεδομένων, των αρχείων csv, ώστε να έχουν την ίδια κωδικοποίηση, και προστέθηκαν οι απόλυτες τιμές τους, καταλήγοντας στο παρακάτω Διάγραμμα (**Εικόνα 59**).



Εικόνα 59: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο της **Ακραίας Ενίσχυσης Κλίσης** (XGB).

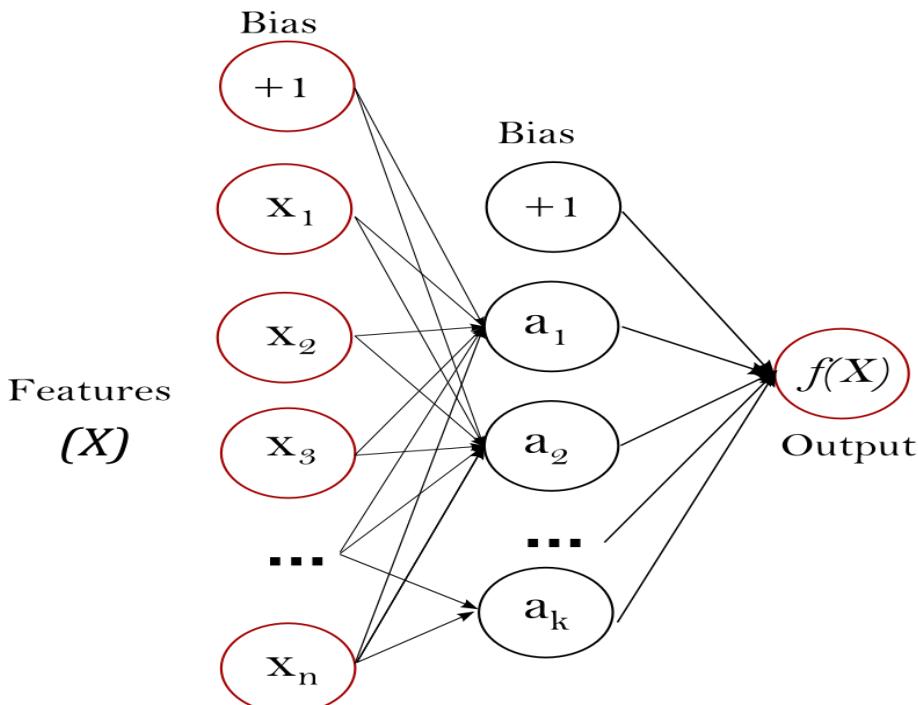
Πιο αναλυτικά, για τον ταξινομητή της **Ακραίας Ενίσχυσης Κλίσης**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 27**.

Πίνακας 27: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για την XGBoost.

1	TYPE_OF_PATIENT	10,891963
2	INTUBATED_2	0,545206
3	PNEUMONIA	0,253355
4	AGE	0,110003
5	CONTACT_WITH_COVID-19_CASE	0,031930
6	CHRONIC_KIDNEY_FAILURE	0,028626
7	DIABETIC	0,023003
8	PREGNANCY_2	0,020081
9	SEX	0,019567
10	HYPERTENSION	0,016958
11	OBESITY	0,013037
12	ICU_2	0,009135
13	OTHER_CHRONIC_DISEASE	0,007349
14	SMOKER	0,005765
15	IMMUNOSUPPRESSED	0,005399
16	DAYs_FROM_SYMPTOM_TO_HOSPITALIZATION	0,005176
17	COPD	0,005067
18	ASTHMA	0,004292
19	CARDIOVASCULAR	0,004089
20	INTUBATED_97	0,000000
21	PREGNANCY_97	0,000000
22	ICU_97	0,000000

4.5.5 Τεχνητά Νευρωνικά Δίκτυα Πολυστρωματικών Αντιλήπτρων

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) όπως αναλύσαμε και σε προηγούμενη ενότητα, μιμούνται την λειτουργία των νευρώνων μέσα από αλγορίθμικές τεχνικές εστιάζοντας στην επίλυση σύνθετων υπολογιστικών προβλημάτων [373, 374]. Σε ένα ΤΝΔ ο κάθε τεχνητός νευρώνας ονομάζεται «Αντίληπτρο» ή «νευρώνας Perceptron» [375], για αυτό και πολλές φορές καλούνται και **Multi-Layer Perceptrons** (MLPs). Τα **MLPs** είναι αλγόριθμοι εποπτεύμενης μάθησης οι οποίοι εκπαιδεύονται «μαθαίνοντας» μια συνάρτηση $f(): \mathbb{R}^m \rightarrow \mathbb{R}^o$ σε ένα σύνολο δεδομένων, με τον όρο m να δηλώνει τον αριθμό των διαστάσεων για την είσοδο και τον όρο O να δηλώνει τον αριθμό των διαστάσεων για την έξοδο. Οπότε δεδομένου ενός συνόλου χαρακτηριστικών $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ και ενός στόχου y , το εκάστοτε MLP μπορεί να «μάθει» μια μη γραμμική προσέγγιση της συνάρτησης για ταξινόμηση ή για παλινδρόμηση. Βασική διαφορά τους από την LR είναι πως μεταξύ της εισόδου και της εξόδου μπορεί να υπάρχουν ένα ή περισσότερα μη γραμμικά επίπεδα, τα λεγόμενα κρυφά επίπεδα [484]. Στην παρακάτω εικόνα (**Εικόνα 60**) απεικονίζεται ένα απλό ΤΝΔ.



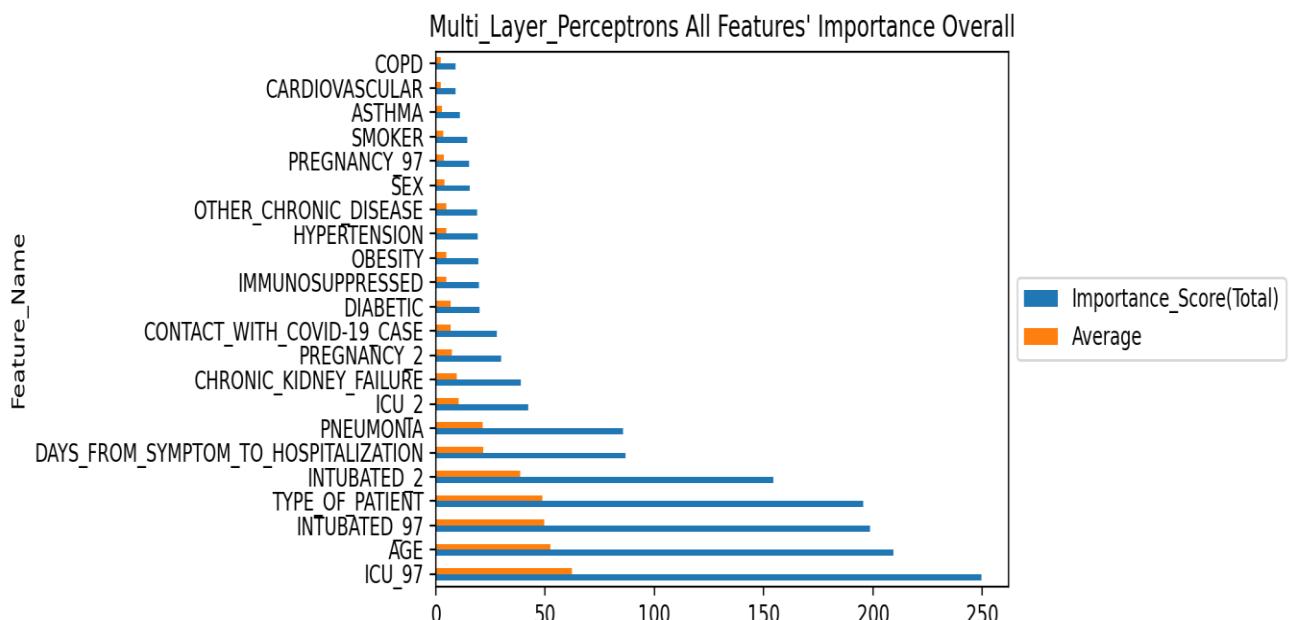
Εικόνα 60: Σχηματική απεικόνιση MLP το οποίο περιέχει 1 κρυφό στρώμα. [485]

Το πρώτο επίπεδο, γνωστό ως στρώμα εισόδου, αποτελείται από ένα σύνολο νευρώνων (x_1, x_2, \dots, x_m) που ο καθένας αντιπροσωπεύει ένα από τα χαρακτηριστικά της εισόδου δεχόμενος και την αντίστοιχη τιμή. Έπειτα στα μεσαία (κρυφά), στρώματα έχουμε τον μετασχηματισμό των τιμών του προηγουμένου επιπέδου με τη χρήση ενός σταθμισμένου παράγοντα (w), διαφορετικό για κάθε χαρακτηριστικό εισόδου (w_1, w_2, \dots, w_m), και έτσι προσθέτοντας όλες τις μετασχηματισμένες τιμές ($x_1w_1, x_2w_2, \dots, x_mw_m$) προκύπτει το γραμμικό άθροισμα ($x_1w_1 + x_2w_2 + \dots + x_mw_m$) το οποίο επεξεργάζεται από μια μη γραμμική συνάρτηση ενεργοποίησης (Activation Function) που μοιάζει με τη σιγμοειδή συνάρτηση. Τέλος το επίπεδο εξόδου λαμβάνει τις τιμές από το τελευταίο κρυφό στρώμα και τις μετατρέπει σε τιμές εξόδου, ανάλογες του εκάστοτε προβλήματος ταξινόμησης ή παλινδρόμησης.

Τα MLPs έχουν δυνατότητα μάθησης μη γραμμικών μοντέλων σε πραγματικό χρόνο, αλλά λόγω των κρυφών επίπεδων, παρουσιάζουν μια μη κυρτή συνάρτηση απώλειας, λόγω των

πολλαπλών τοπικών ελαχίστων. Επομένως, διαφορετικές αρχικοποιήσεις τυχαίου βάρους μπορούν να οδηγήσουν σε διαφορετική ακρίβεια επικύρωσης, καθώς τα MLPs είναι ευαίσθητα στην κλιμάκωση των βαρών των χαρακτηριστικών [484].

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής **MLPClassifier** της βιβλιοθήκης **sklearn**, ο οποίος εφαρμόζει έναν αλγόριθμο πολλαπλών στρωμάτων Perceptrons, που εκπαιδεύεται χρησιμοποιώντας Οπίσθια Τροφοδότηση (Backpropagation). Λόγω του ότι ο **MLPClassifier** δεν υποστηρίζει κάποια μέθοδο (function), για να εξαχθούν τα σκορ σημαντικότητας των χαρακτηριστικών, εδώ το σκορ για κάθε χαρακτηριστικό υπολογίστηκε ως το κανονικοποιημένο άθροισμα των 4 σκορ, που αυτό έχει, για τους 4 προηγουμένους αλγορίθμους (LogisticRegression, DecisionTreeClassifier, RandomForestClassifier, XGBClassifier). Τα υπολογισμένα σκορ απεικονίζονται στο παρακάτω Διάγραμμα (**Εικόνα 61**).



Εικόνα 61: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο των **Τεχνητών Νευρωνικών Δικτύων Πολυστρωματικών Αντιλήπτρων** (MLPs).

Πιο αναλυτικά, για τον ταξινομητή των **Τεχνητών Νευρωνικών Δικτύων Πολυστρωματικών Αντιλήπτρων**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 28**.

Πίνακας 28: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για τα MLPs.

1	ICU_97	249,446172
2	AGE	209,442170
3	INTUBATED_97	198,652692
4	TYPE_OF_PATIENT	195,450681
5	INTUBATED_2	154,341739
6	DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION	86,839213
7	PNEUMONIA	85,586372
8	ICU_2	42,387277
9	CHRONIC_KIDNEY_FAILURE	38,817922
10	PREGNANCY_2	29,884467
11	CONTACT_WITH_COVID-19_CASE	27,806858
12	DIABETIC	20,172948
13	IMMUNOSUPPRESSED	19,712826
14	OBESITY	19,355744
15	HYPERTENSION	19,310151
16	OTHER_CHRONIC_DISEASE	18,839655
17	SEX	15,560189
18	PREGNANCY_97	15,387542
19	SMOKER	14,278375
20	ASTHMA	10,993079
21	CARDIOVASCULAR	9,016073
22	COPD	8,967528

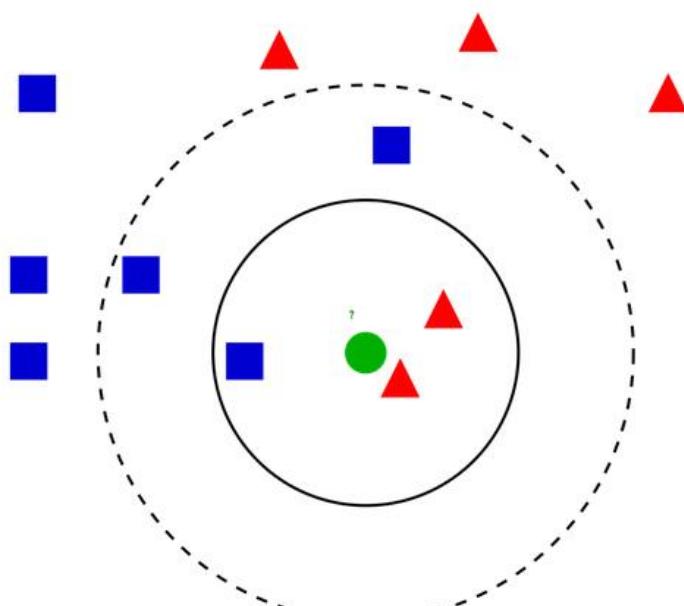
4.5.6 Κ Εγγύτεροι Γείτονες

Ο ταξινομητής των **Κ Εγγύτερων Γειτόνων** (K Nearest Neighbors, **KNN**) που εφαρμόζει την ψηφοφορία κ-πλησιέστερων γειτόνων, όπως αναλύθηκε και σε προηγούμενη ενότητα, είναι μια μη παραμετρική μέθοδος επιβλεπόμενης μάθησης που χρησιμοποιείται σε προβλήματα ταξινόμησης.

Ο **KNN** στηρίζεται στην απόσταση των τιμών των γνωστών δειγμάτων ώστε να μπορέσει να ταξινομήσει τα καινούρια. Η μέθοδος KNN, ανακαλύφθηκε από τους Fix και Hodges το 1951 και εν συνεχεία αναπτύχθηκε από τον Cover [486]. Βασική λογική του τρόπου λειτουργίας του ταξινομητή KNN είναι πως «τα όμοια δείγματα ανήκουν στην ίδια κλάση» [487].

Ο όρος μη-παραμετρική εκτίμηση, δηλώνει πως τα δεδομένα της εισόδου, θα έχουν δεδομένα εξόδου όμοιου τύπου. Πιο συγκεκριμένα η μέθοδος αυτή είναι κατασκευασμένη με σκοπό την εύρεση δειγμάτων, τα οποία θα χρησιμοποιηθούν ως παραδείγματα, από τα δεδομένα της εκπαίδευσης και με τη χρήση των κατάλληλων μέτρων απόστασης να ταξινομήσει το νέο, άγνωστο δείγμα [488]. Ο KNN μπορεί να χρησιμοποιηθεί για είτε για προβλήματα Κατηγοριοποίησης, για διακριτές μεταβλητές στόχους, είτε για προβλήματα Παλινδρόμησης, για συνεχείς μεταβλητές στόχους.

Η κλάση στην οποία θα ανήκει το κάθε νέο δείγμα εξαρτάται από την απόσταση του από τα κ προηγούμενα δείγματα του συνόλου των δεδομένων της εκπαίδευσης, πιο συγκεκριμένα, αν θεωρήσουμε ένα σύνολο δειγμάτων $X_i: x_1, x_2, \dots, x_m \in \mathbb{R}^m$, ένα σύνολο κλάσεων $C_i: c_1, c_2, \dots, c_m$, και ένα νέο δείγμα $y_k \in \mathbb{R}^m$ που είναι να ταξινομηθεί, τότε μετά από την απεικόνιση των δεδομένων στον χώρο \mathbb{R}^m , το y_k θα ταξινομηθεί με βάσει την κλάση της περισσότερων k κοντινότερων δειγμάτων του που ανήκουν στο σύνολο X_i . Στην **Εικόνα 62** απεικονίζεται ένα παράδειγμα ταξινόμησης KNN.

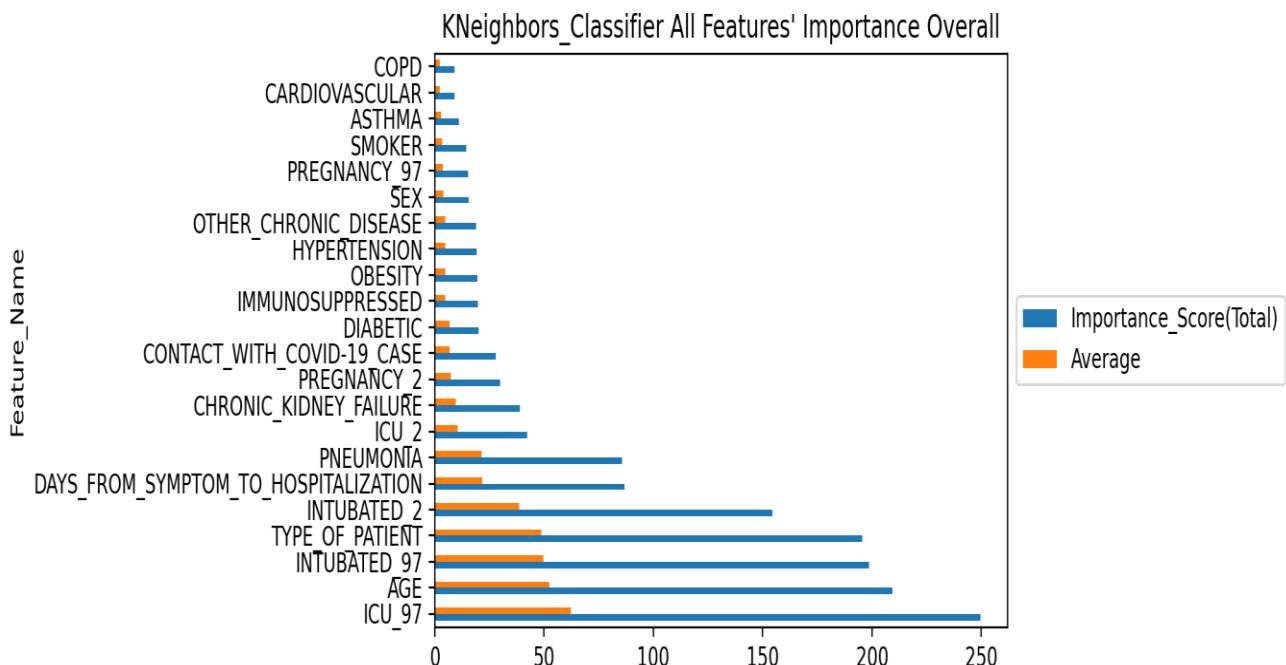


Εικόνα 62: Παράδειγμα Κ Εγγύτερων Γειτόνων - Το νέο δείγμα (πράσινος κύκλος), θα ταξινομηθεί στην κλάση των κόκκινων τριγώνων, σύμφωνα με τον αλγόριθμο των 3ων-Κοντινότερων Γειτόνων, ενώ αντίθετα αν πάρουμε τους 5 Κοντινότερους γείτονες τότε ταξινομείται στην κλάση των μπλε τετραγώνων. [489]

Ο **KNN** αποτελεί μία απλή και εύκολη, ως προς την κατανόηση, μέθοδο η οποία μπορεί να την χρησιμοποιήσει ως μια αρχική προσπάθεια για την εξαγωγή κάποιων αρχικών συμπερασμάτων. Ακόμα το γεγονός πως είναι μη παραμετρική-μέθοδος ταξινόμησης, δηλαδή δεν είναι απαραίτητο να πληρούνται κάποιες προϋποθέσεις για την εφαρμογή της.

Επιπλέον ο KNN δύναται να χρησιμοποιηθεί και σε προβλήματα πολλαπλών κλάσεων. Στον αντίποδα η εύρεση της παραμέτρου K και της κατάλληλης μετρικής μπορεί να αποτελέσει μια χρονοβόρα και δύσκολη διαδικασία, με τη δυσκολία και την αποδοτικότητα του KNN, να αυξάνεται ανάλογα με το μέγεθος του συνόλου δεδομένων [490].

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής **KNeighborsClassifier** της βιβλιοθήκης **sklearn**, ο οποίος δεν υποστηρίζει κάποια μέθοδο (function), για να εξαχθούν τα σκορ σημαντικότητας των χαρακτηριστικών, εδώ το σκορ για κάθε χαρακτηριστικό υπολογίστηκε, όπως και για τον **MLPClassifier**, ως το κανονικοποιημένο άθροισμα των 4 σκορ, που αυτό έχει, για τους 4 προηγουμένους αλγορίθμους (LogisticRegression, DecisionTreeClassifier, RandomForestClassifier, XGBClassifier). Τα υπολογισμένα σκορ απεικονίζονται στο παρακάτω Διάγραμμα (**Εικόνα 63**).



Εικόνα 63: Διάγραμμα κατάταξης των βαθμών σημαντικότητας όλων των 22 χαρακτηριστικών του σύνολο των δεδομένων για τον αλγόριθμο των **K Εγγύτερων Γειτόνων** (KNN).

Πιο αναλυτικά, για τον ταξινομητή **Κ Εγγύτερων Γειτόνων**, οι βαθμολογίες των χαρακτηριστικών (στηλών), σε φθίνουσα σειρά, απεικονίζονται στον **Πίνακα 29**.

Πίνακας 29: Οι Βαθμολογίες της σημαντικότητας των χαρακτηριστικών για τους KNN.

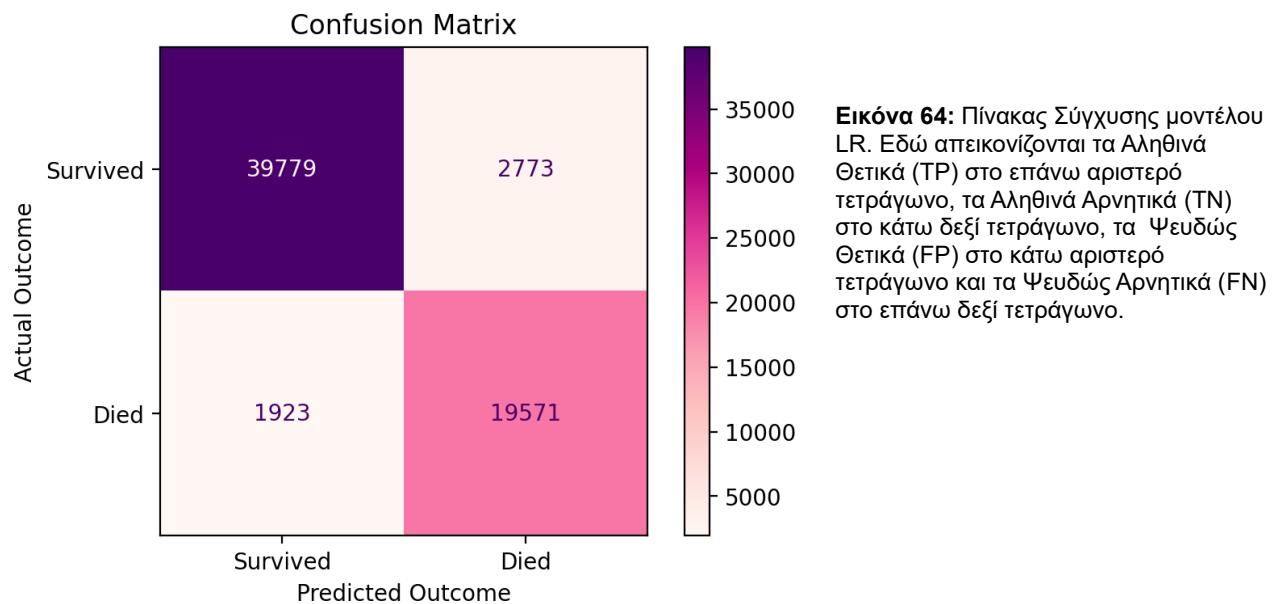
1	ICU_97	249,446172
2	AGE	209,442170
3	INTUBATED_97	198,652692
4	TYPE_OF_PATIENT	195,450681
5	INTUBATED_2	154,341739
6	DAYS_FROM_SYMPTOM_TO_HOSPITALIZATION	86,839213
7	PNEUMONIA	85,586372
8	ICU_2	42,387277
9	CHRONIC_KIDNEY_FAILURE	38,817922
10	PREGNANCY_2	29,884467
11	CONTACT_WITH_COVID-19_CASE	27,806858
12	DIABETIC	20,172948
13	IMMUNOSUPPRESSED	19,712826
14	OBESITY	19,355744
15	HYPERTENSION	19,310151
16	OTHER_CHRONIC_DISEASE	18,839655
17	SEX	15,560189
18	PREGNANCY_97	15,387542
19	SMOKER	14,278375
20	ASTHMA	10,993079
21	CARDIOVASCULAR	9,016073
22	COPD	8,967528

4.6 Μετρικές Αξιολόγησης Αλγορίθμων

Μετά την εφαρμογή των μοντέλων, το επόμενο βήμα είναι ο έλεγχος της αποτελεσματικότητας του κάθε διαφορετικού αλγορίθμου. Στην παρούσα εργασία, για την μέτρηση της απόδοσης της ταξινόμησης θα χρησιμοποιήσουμε τις μετρικές, της Ακρίβειας (Accuracy), της Ευαισθησίας-Ανάκλησης (Recall), το F1 σκορ, της Διάρκειας (Duration) εκτέλεσης, της Περιοχής Κάτω από την Καμπύλη ROC (Area Under ROC Curve, AUC_ROC) [491].

Πίνακας Σύγχυσης

Ο Πίνακας Σύγχυσης (Confusion Matrix) αποτελεί μια συνοπτική απεικόνιση των αποτελεσμάτων της πρόβλεψης για 1 εκτέλεση ενός μοντέλου. Ο πίνακας σύγχυσης (**Εικόνα 64**) δείχνει τους τρόπους με τους οποίους το εκάστοτε μοντέλο σας προβλέπει σωστά ή λάθος, εδώ ο αριθμός των σωστών και λάθος προβλέψεων συνοψίζεται με τιμές μέτρησης και αναλύεται ανά κατηγορία, παρέχοντας πληροφορίες τόσο για τα σφάλματα του μοντέλου, όσο για τους τύπους των σφαλμάτων αυτών [492].



Αληθινά Θετικά (True Positives, TP): Είναι οι σωστά προβλεπόμενες θετικές τιμές (1:Yes), για τα δείγματα του συνόλου εκπαίδευσης, πιο συγκεκριμένα, σε κάθε ένα από τα δείγματα αυτά, η πραγματική θετική τιμή της κλάσης, ταυτίζεται με την προβλεπόμενη τιμή της κλάσης [491], π.χ. Σε μια Αληθινά Θετική πρόβλεψη, αν η πρόβλεψη του μοντέλου για ένα δείγμα του συνόλου εκπαίδευσης είναι πως ο ασθενής επιβίωσε και η πραγματική τιμή του δείγματος είναι η ίδια, δηλαδή πως ο ασθενής αυτός ζόντως επιβίωσε.

Αληθινά Αρνητικά (True Negatives, TN): Είναι οι σωστά προβλεπόμενες αρνητικές τιμές (0:No), για τα δείγματα του συνόλου εκπαίδευσης, πιο συγκεκριμένα, σε κάθε ένα από τα δείγματα αυτά, η πραγματική αρνητική τιμή της κλάσης, ταυτίζεται με την προβλεπόμενη τιμή της κλάσης [491], π.χ. Σε μια Αληθινά Αρνητική πρόβλεψη, αν η πρόβλεψη του μοντέλου για ένα δείγμα του συνόλου εκπαίδευσης είναι πως ο ασθενής απεβίωσε και η πραγματική τιμή του δείγματος είναι η ίδια, δηλαδή πως ο ασθενής αυτός ζόντως απεβίωσε.

Ψευδώς Θετικά (False Positives, FP): Είναι οι λάθος προβλεπόμενες θετικές τιμές (1:Yes), για τα δείγματα του συνόλου εκπαίδευσης, πιο συγκεκριμένα, σε κάθε ένα από τα δείγματα αυτά, η προβλεπόμενη θετική τιμή της κλάσης, δεν ταυτίζεται με την πραγματική τιμή της

κλάσης [491], π.χ. Σε μια Ψευδώς Θετική πρόβλεψη, αν η πρόβλεψη του μοντέλου για ένα δείγμα του συνόλου εκπαίδευσης είναι πως ο ασθενής επιβίωσε και η πραγματική τιμή του δείγματος είναι η αντίθετη, δηλαδή πως ο ασθενής αυτός απεβίωσε.

Ψευδώς Αρνητικά (False Negatives, FN): Είναι οι λάθος προβλεπόμενες αρνητικές τιμές (0:No), για τα δείγματα του συνόλου εκπαίδευσης, πιο συγκεκριμένα, σε κάθε ένα από τα δείγματα αυτά, η προβλεπόμενη αρνητική τιμή της κλάσης, δεν ταυτίζεται με την πραγματική τιμή της κλάσης [491], π.χ. Σε μια Ψευδώς Αρνητική πρόβλεψη, αν η πρόβλεψη του μοντέλου για ένα δείγμα του συνόλου εκπαίδευσης είναι πως ο ασθενής απεβίωσε και η πραγματική τιμή του δείγματος είναι η αντίθετη, δηλαδή πως ο ασθενής αυτός απεβίωσε.

Έπειτα με βάσει αυτές τις τέσσερις παραμέτρους, μπορούμε να υπολογίσουμε την Ακρίβεια (Precision), την, την Ευαισθησία-Ανάκληση (Recall) και τη βαθμολογία F1 (F1 Score), και την Περιοχής Κάτω από την Καμπύλη ROC (Area Under ROC Curve, AUC_ROC).

Ακρίβεια (Precision): Είναι ο λόγος των Αληθινά Θετικά (TP) προς τις συνολικές προβλεπόμενες Θετικές παρατηρήσεις (TP+FP), μας δίνει το ποσοστό των Θετικών προβλέψεων του μοντέλου, οι οποίες έχουν πραγματικά Θετική τιμή. Δίνεται από τον τύπο [491]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Για τον συγκεκριμένο Πίνακα Σύγχυσης η Ακρίβεια είναι 0,954. Για τον υπολογισμό της χρησιμοποιήθηκε η μέθοδος `precision_score` της `sklearn`.

Ευαισθησία-Ανάκληση(Recall): Είναι ο λόγος των Αληθινά Θετικά (TP) προς τις όλες της πραγματικά Θετικές τιμές (TP+FN), μας δίνει το ποσοστό των πραγματικά Θετικών τιμών που προβλέφθηκαν σωστά από το μοντέλο. Δίνεται από τον τύπο [491]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Για τον συγκεκριμένο Πίνακα Σύγχυσης η Ακρίβεια είναι 0,935. Για τον υπολογισμό της χρησιμοποιήθηκε η μέθοδος `recall_score` της `sklearn`.

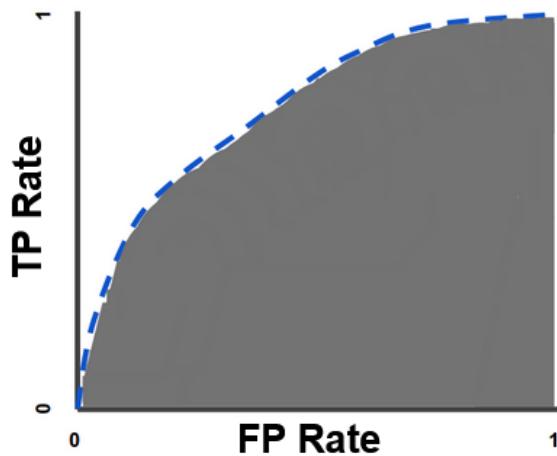
Βαθμολογία F1 (F1 Score): είναι ο σταθμισμένος μέσος όρος **Precision** και **Recall**, λαμβάνει υπόψη τόσο τα FP όσο και τα FN και συνήθως είναι πιο χρήσιμο από την ακρίβεια, ειδικότερα αν υπάρχει άνιση κατανομή κλάσης στόχου. Η ακρίβεια λειτουργεί καλύτερα εάν τα FP και τα FN έχουν παρόμοιο κόστος. Αν το κόστος των FP και των FN είναι αρκετά διαφορετικό, τότε είναι καλύτερα να εξεταστούν τόσο η Precision όσο και η Recall. Δίνεται από τον τύπο [491]:

$$\text{F1 Score} = 2 \times (\text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision})$$

Για τον συγκεκριμένο Πίνακα Σύγχυσης η Βαθμολογία F1 είναι 0,944. Για τον υπολογισμό της χρησιμοποιήθηκε η μέθοδος `f1_score` της `sklearn`.

Διάρκεια (Runtime): Στην παρούσα εργασία χρησιμοποιήθηκε ως επιπλέον μετρικό η Διάρκεια της κάθε εκτέλεσης της κάθε επανάληψης(iteration) του Αλγορίθμου.

Περιοχή κάτω από την καμπύλη ROC (Area Under Curve ROC, AUC_ROC): υπολογίζεται ως ολόκληρη η δισδιάστατη περιοχή κάτω από την καμπύλη ROC (**Εικόνα 65**), από το 0,0 έως το 1,1 [493].



Εικόνα 65: AUC (Περιοχή κάτω από την καμπύλη ROC). [494]

Η καμπύλη ROC (Receiver Operating characteristic Curve) είναι ένα γράφημα το οποίο απεικονίζει την απόδοση ενός μοντέλου ταξινόμησης σε όλα τα κατώφλια της ταξινόμησης. Πάνω σε αυτή την καμπύλη παρουσιάζονται οι παράμετροι του **Αληθινά Θετικού Ποσοστού** (True Positive Rate, TPR) και του **Ψευδώς Θετικού Ποσοστού** (False Positive Rate, FPR) [493].

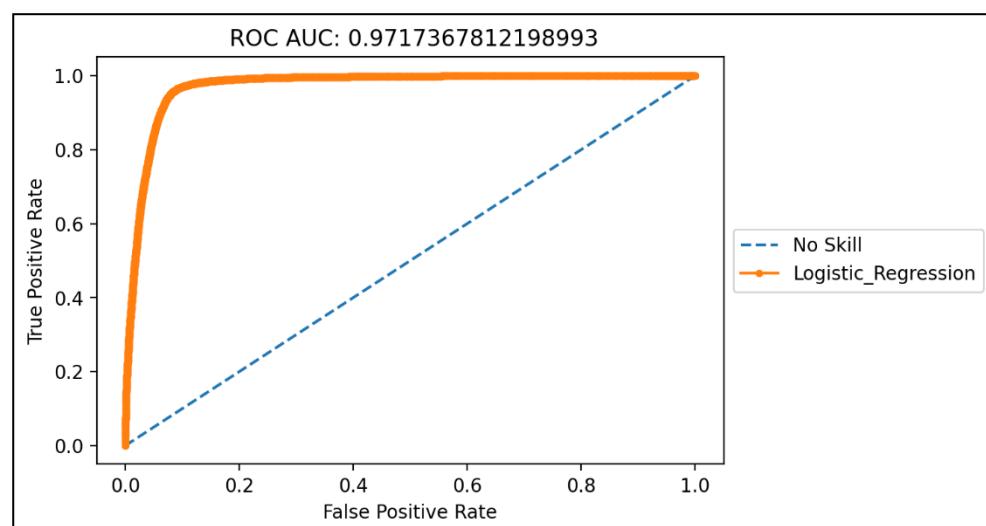
To TPR που είναι συνώνυμο της Recall και υπολογίζεται από τον τύπο:

$$TPR = TP / (TP + FN)$$

To FPR υπολογίζεται από τον τύπο:

$$FPR = FP / (FP + TN)$$

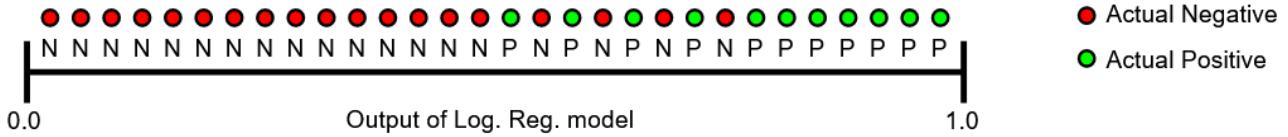
Η καμπύλη ROC παρουσιάζει το TPR έναντι του FPR στα διάφορα κατώφλια της ταξινόμησης, όσο μειώνεται το κατώφλι ταξινόμησης το μοντέλο ταξινομεί περισσότερα στοιχεία ως θετικά, με αποτέλεσμα να αυξάνονται τόσο τα FP όσο και τα TP [493]. Το παρακάτω σχήμα (Εικόνα 66) δείχνει την καμπύλη ROC που αντιστοιχεί στο μοντέλο του παραπάνω πίνακα σύγχυσης (Εικόνα 64).



Εικόνα 66: Καμπύλη ROC όπου απεικονίζεται το ποσοστό TP έναντι του ποσοστού FP στα διάφορα κατώφλια ταξινόμησης.

Για να υπολογίσουμε τα σημεία (κατώφλια της ταξινόμησης) σε μια καμπύλη ROC (AUC_ROC), χρησιμοποιήθηκε η μέθοδος `roc_auc_score` της `sklearn`. Η AUC_ROC παρέχει ένα συνολικό μέτρο απόδοσης για όλα τα πιθανά κατώφλια ταξινόμησης και μπορεί

να ερμηνευτεί ως η πιθανότητα το μοντέλο να ταξινομήσει ένα τυχαίο δείγμα που έχει θετική τιμή, για τη μεταβλητή στόχο, υψηλότερα από ένα τυχαίο δείγμα με αρνητική τιμή, για τη μεταβλητή στόχο, παραδείγματος χάριν για τα δείγματα της **Εικόνας 67** (**πράσινα**= δείγματα που έχουν θετική τιμή, **κόκκινα**= δείγματα που έχουν αρνητική τιμή), τα οποία είναι διατεταγμένα από αριστερά προς τα δεξιά σε αύξουσα σειρά με βάσει τις προβλέψεις ενός μοντέλου LR η AUC εκφράζει την πιθανότητα ένα τυχαίο **θετικό** δείγμα να τοποθετηθεί στα δεξιά ενός τυχαίου **αρνητικού** δείγματος.



Εικόνα 67: Οι προβλέψεις ταξινομήθηκαν με αύξουσα σειρά, όσον αφορά την βαθμολογία του μοντέλου της LR. [495]

Η τιμή της AUC_ROC κυμαίνεται από 0 έως 1, πιο συγκεκριμένα σε ένα μοντέλο του οποίου οι προβλέψεις είναι 100% λανθασμένες η AUC_ROC θα είναι ίση με 0,00, ενώ αντίθετα σε μοντέλο του οποίου οι προβλέψεις είναι 100% σωστές η AUC_ROC θα είναι ίση με 1,00 [493].

Αποτελέσματα

Με βάση τα αποτελέσματα από τις συνολικά **3.240** εκτελέσεις (iterations), **540** για τον καθένα από τους **6** αλγόριθμους, κατασκευάστηκαν οι παρακάτω **18** πίνακες, **3** για κάθε αλγόριθμο.

5.1 Επεξήγηση Πινάκων

Ο κάθε διαφορετικός αλγόριθμος απεικονίζεται με διαφορετικό χρώμα σύμφωνα με την παρακάτω αντιστοιχία:

- **Λογιστική Παλινδρόμηση:** **Logistic Regression**
- **Δέντρα Αποφάσεων:** **Decision Trees**
- **Τυχαίο Δάσος:** **Random Forest**
- **Ακραία Ενίσχυση Κλίσης:** **eXtreme Gradiend Boosting**
- **ΤΝΔ Πολυστρωματικών Αντιλήπτρων:** **Multi-Layer Perceptrons**
- **Κ Εγγύτεροι Γείτονες:** **K-Nearest Neighbors**

Για κάθε αλγόριθμο

Στους 3 πίνακες που ανήκουν στο ίδιο αλγόριθμο στις **10 υψηλότερες** τιμές για τα μετρικά **Precision, Recall, F1-score** και **AU_ROC** και στις **10 χαμηλότερες** για την **Runtime**, έχει σημειωθεί σε παρένθεση με **πράσινο** η θέση της κάθε τιμής, με τις **3 πρώτες** να έχουν διαφορετικό χρώμα, ανάλογα την με την θέση τους (1^η: **green**, 2^η: **light green**, 3^η: **yellow**).

Ομοίως στις **10 χαμηλότερες** τιμές για τα μετρικά **Precision, Recall, F1-score** και **AU_ROC** και στις **10 υψηλότερες** για την **Runtime**, έχει σημειωθεί σε παρένθεση με **κόκκινο** η θέση της κάθε τιμής, με τις **3 πρώτες** να έχουν διαφορετικό χρώμα, ανάλογα την με την θέση τους (54^η: **red**, 53^η: **orange**, 52^η: **gold**).

Για όλους τους αλγόριθμους

Σε όλους τους πίνακες των αλγορίθμων στις **20 υψηλότερες** τιμές για τα μετρικά Precision, Recall, F1-score και AU_ROC και στις **20 χαμηλότερες** για την Runtime, έχει σημειωθεί σε παρένθεση με **πράσινο και αστερίσκο(*)** η θέση της κάθε τιμής, με τις **3 πρώτες** να έχουν διαφορετικό χρώμα, ανάλογα την με την θέση τους (1^η: **purple**, 2^η: **magenta**, 3^η: **pink**)

Ομοίως στις **20 χαμηλότερες** τιμές για τα μετρικά Precision, Recall, F1-score και AU_ROC και στις **20 υψηλότερες** για την Runtime, έχει σημειωθεί σε παρένθεση με **κόκκινο και αστερίσκο(*)** η θέση της κάθε τιμής, με τις **3 πρώτες** να έχουν διαφορετικό χρώμα, ανάλογα την με την θέση τους (324^η: **dark blue**, 323^η: **blue**, 322^η: **teal**)

5.2 Πίνακες Αποτελεσμάτων

Για την **Λογιστική παλινδρόμηση (LR)** έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 30)

Logistic Regression 22 Features																					
Scaler Type	Parameter Type	Precision			Recall				F1-score				ROC AUC				Runtime(secs)				
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92474	0,00097	0,9231	0,92645	0,9085(46)	0,00095	0,90696	0,90983	0,89006	0,0014	0,88748	0,89241	0,97057(2)	0,00064	0,96948	0,97151	2,98757(48)	0,13926	2,78059	3,22352
	Optimal_01	0,92531(6)	0,00127	0,9226	0,92704	0,91291	0,00314	0,90807	0,91695	0,89134(5)	0,00183	0,88757	0,89381	0,96896	0,00054	0,9679	0,96978	1,93554	0,09955	1,78934	2,08105
	Optimal_02	0,92479	0,00109	0,92313	0,92637	0,90889	0,00416	0,90115	0,91429	0,89028	0,00171	0,88809	0,89299	0,9702	0,0005	0,96922	0,9709	2,19047	0,09597	2,04629	2,35885
Standard Scaler	Default	0,92564(2)	0,00123	0,92324	0,92739	0,90994	0,00257	0,90584	0,91436	0,89136(3)	0,00181	0,88792	0,89386	0,97081(1)	0,00053	0,97014	0,97155	2,99462(49)	0,27446	2,35883	3,42108
	Optimal_01	0,92482	0,00116	0,92319	0,92695	0,9117	0,00123	0,90948	0,91368	0,89067	0,00146	0,88857	0,89322	0,96856	0,00051	0,96756	0,96922	1,73927	0,11228	1,61439	1,9323
	Optimal_02	0,9248	0,00106	0,92351	0,92668	0,90864	0,00194	0,90627	0,91246	0,89037	0,00154	0,88874	0,89355	0,97024	0,00081	0,96889	0,9718	2,06294	0,1585	1,91427	2,46622
MinMax Scaler 0-1	Default	0,92542(5)	0,00111	0,92334	0,92752	0,90945	0,00284	0,90445	0,91704	0,89121(7)	0,00165	0,88831	0,8946	0,97054(3)	0,00067	0,9693	0,9722	3,15965(50)	0,16066	2,88995	3,67392
	Optimal_01	0,92472	0,00104	0,92242	0,92632	0,91673(9)	0,00664	0,90586	0,92857	0,89094(10)	0,00175	0,88672	0,89393	0,96731	0,0014	0,96526	0,96963	2,14899	0,69917	1,49811	3,63623
	Optimal_02	0,92526(7)	0,00159	0,92225	0,92949	0,91153	0,00677	0,8979	0,92546	0,89116(8)	0,00264	0,88563	0,89806	0,9696	0,0021	0,96526	0,97242	2,49385	0,37812	1,95268	3,48616
MinMax Scaler 0-10	Default	0,92545(4)	0,00116	0,92334	0,92752	0,90999	0,003	0,90445	0,91704	0,89128(6)	0,00175	0,88831	0,8946	0,97047(7)	0,00061	0,9693	0,97168	3,18224(52)	0,17457	2,88995	3,67392
	Optimal_01	0,92472	0,00111	0,92242	0,92632	0,91355	0,00402	0,90586	0,9212	0,89058	0,00172	0,88672	0,89305	0,96778	0,00129	0,96529	0,96963	2,32189	0,72922	1,60729	3,63623
	Optimal_02	0,92504	0,00139	0,92225	0,92862	0,91141	0,00675	0,8979	0,92546	0,89084	0,00227	0,88563	0,89567	0,96917	0,00224	0,96526	0,97202	2,35542	0,26533	1,95268	2,86337
MinMax Scaler 0-100	Default	0,92562(3)	0,0012	0,92334	0,92752	0,91068	0,00322	0,90445	0,91704	0,89157(2)	0,00181	0,88831	0,8946	0,97041	0,00066	0,9693	0,97168	3,23525(53)	0,16606	3,01069	3,67392
	Optimal_01	0,92463	0,00094	0,92316	0,92602	0,9148	0,00417	0,90586	0,9212	0,8906	0,00151	0,88796	0,89305	0,96733	0,00128	0,96529	0,96921	2,62905	0,71447	1,81059	3,63623
	Optimal_02	0,92478	0,00128	0,92225	0,92721	0,91261	0,00759	0,8979	0,92546	0,89063	0,00227	0,88563	0,89456	0,96826	0,0022	0,96526	0,97102	2,33779	0,28349	1,95268	2,86337
MinMax Scaler 0-1000	Default	0,92626(1)	0,00085	0,92472	0,92752	0,91287	0,00272	0,90986	0,91704	0,89262(1)	0,00138	0,88996	0,8946	0,97047(6)	0,00056	0,96961	0,97153	3,17691(51)	0,20378	3,01069	3,67392
	Optimal_01	0,92476	0,00095	0,92345	0,92602	0,91777(5)	0,00247	0,91393	0,9212	0,89114(9)	0,00142	0,88883	0,89305	0,96619(50)	0,00063	0,96529	0,96705	3,31659(54)	0,14032	3,14007	3,63623
	Optimal_02	0,92489	0,00098	0,92356	0,92655	0,91836(2)	0,00273	0,91612	0,92546	0,89135(4)	0,00132	0,88936	0,89385	0,96621(49)	0,00071	0,96526	0,96735	2,53675	0,10191	2,43697	2,75176

❖ Για τα 15 χαρακτηριστικά (Πίνακας 31)

Logistic Regression 15 Features																					
Scaler Type	Parameter Type	Precision			Recall				F1-score				ROC AUC				Runtime(secs)				
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92496	0,00131	0,92311	0,92776	0,90951	0,002	0,90795	0,91467	0,8905	0,00179	0,88807	0,89382	0,97044(9)	0,00099	0,96944	0,97296	2,69977(46)	0,11924	2,53394	2,93307
	Optimal_01	0,92474	0,00135	0,92189	0,92685	0,91107	0,00326	0,90387	0,91495	0,89066	0,002	0,88628	0,89316	0,96836	0,00061	0,96704	0,96922	1,70605(6)	0,06865	1,59345	1,81718
	Optimal_02	0,92458	0,00108	0,92296	0,92623	0,90814(50)	0,00397	0,90214	0,91652	0,88984	0,00171	0,88717	0,89296	0,97002	0,00046	0,96954	0,97071	1,90726	0,11913	1,74552	2,12278
Standard Scaler	Default	0,92511(4)	0,00085	0,92389	0,92646	0,90919	0,00193	0,90643	0,91285	0,89083	0,00125	0,88881	0,89281	0,97038	0,00075	0,96913	0,97141	2,54578	0,19255	2,13487	2,8392
	Optimal_01	0,92478	0,00135	0,92293	0,92752	0,91342	0,00221	0,90955	0,91573	0,8908	0,00183	0,88787	0,89417	0,96843	0,00071	0,96759	0,96986	1,46687(2)	0,10419	1,35716	1,68715
	Optimal_02	0,9251	0,00104	0,92331	0,92728	0,90938	0,0029	0,90423	0,91407	0,89078	0,00166	0,88777	0,89435	0,97044(8)	0,00041	0,96973	0,97126	2,18624	0,21749	1,81093	2,42306
MinMax Scaler 0-1	Default	0,92499	0,00082	0,92291	0,92638	0,90845(48)	0,00238	0,90331	0,91233	0,89044	0,00125	0,88737	0,8927	0,97043(10)	0,00055	0,9692	0,97198	2,48206	0,29564	1,85031	3,0329
	Optimal_01	0,92434	0,00133	0,92219	0,92709	0,91794(4)	0,00741	0,90638	0,93316	0,89066	0,00172	0,88655	0,89417	0,96702	0,00165	0,9647	0,9699	1,73141(10)	0,46281	1,23412	2,61708
	Optimal_02	0,92421	0,00171	0,91955	0,92706	0,90917	0,00759	0,88398	0,92174	0,88957(48)	0,00273	0,88122	0,89403	0,96889	0,00252	0,96315	0,97151	2,02179	0,35249	1,62687	3,33815
MinMax Scaler 0-10	Default	0,92511(10)	0,00086	0,92291	0,92638	0,90807	0,00239	0,90331	0,91233	0,89062	0,00126	0,88737	0,8927	0,97048(5)	0,00057	0,9692	0,97198	2,59504	0,20824	2,18673	3,0329
	Optimal_01	0,92461	0,00128	0,92219	0,92709	0,91418	0,00375	0,90638	0,92071	0,89061	0,00177	0,88655	0,89417	0,96746	0,00164	0,9647	0,9699	1,85957	0,46753	1,3258	2,61708
	Optimal_02	0,92418	0,00181	0,91955	0,92706	0,91021	0,00777	0,88398	0,92174	0,88961(47)	0,00281	0,88122	0,89403	0,96843	0,00275	0,96315	0,97151	1,89051	0,18395	1,62687	2,42137
MinMax Scaler 0-100	Default	0,92511(8)	0,001	0,92291	0,92638	0,90849(47)	0,00242	0,90331	0,91233	0,89063	0,00145	0,88737	0,8927	0,97051(4)	0,00067	0,9692	0,97198	2,66901(45)	0,18622	2,23193	3,0329
	Optimal_01	0,92437	0,00119	0,92284	0,92677	0,91522(10)	0,00374	0,90638	0,92071	0,89039	0,00157	0,88849	0,89343	0,96694(45)	0,00174	0,9647	0,9699	2,0862	0,41244	1,53931	2,61708
	Optimal_02	0,92358(51)	0,00186	0,91955	0,92682	0,91031	0,00941	0,88398	0,92174	0,88879(54)	0,003	0,88122	0,89299	0,96746	0,0029	0,96315	0,97151	1,83556	0,13363	1,62687	2,15575
MinMax Scaler 0-1000	Default	0,9246																			

❖ Για τα 10 χαρακτηριστικά (Πίνακας 32)

Logistic Regression		10 Features																			
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92448	0,00155	0,92203	0,92663	0,90815(51)	0,00266	0,90522	0,91232	0,8898	0,00222	0,88617	0,89278	0,96991	0,0007	0,9688	0,971	2,60334	0,13226	2,37445	2,83846
	Optimal_01	0,92472	0,00124	0,92297	0,92671	0,91347	0,00374	0,90894	0,921	0,89077	0,00199	0,88817	0,89382	0,96815	0,00057	0,96736	0,96891	1,60786(4)	0,11982	1,4528	1,78086
	Optimal_02	0,92459	0,00106	0,92359	0,9265	0,90796(53)	0,00343	0,90155	0,91424	0,88985	0,00177	0,88816	0,89304	0,97001	0,00076	0,96906	0,97189	1,71157(7)	0,07259	1,609	1,84331
Standard Scaler	Default	0,92478	0,00112	0,92337	0,92641	0,90912	0,00171	0,90629	0,9113	0,89021	0,00153	0,88808	0,89237	0,97008	0,0007	0,96907	0,97099	1,99003	0,13167	1,76578	2,10992
	Optimal_01	0,92448	0,00077	0,92305	0,92596	0,91213	0,00231	0,90903	0,91636	0,89022	0,0012	0,88818	0,8925	0,96801	0,00063	0,96673	0,96895	1,3429(1)	0,07063	1,26534	1,48629
	Optimal_02	0,92492	0,0009	0,92379	0,9269	0,91022	0,00134	0,90739	0,91157	0,89063	0,00135	0,8888	0,89348	0,96983	0,00048	0,96912	0,97046	1,77832	0,11577	1,65586	2,04643
MinMax Scaler 0-1	Default	0,9247	0,00091	0,92297	0,92641	0,90853	0,00249	0,90379	0,91329	0,89018	0,00135	0,88764	0,89253	0,9701	0,00049	0,96891	0,97111	2,26139	0,40603	1,62458	3,2492
	Optimal_01	0,92365(50)	0,00099	0,92135	0,92589	0,91761(6)	0,00857	0,90705	0,93233	0,88978	0,00143	0,88582	0,8923	0,96659(47)	0,00128	0,96394	0,96841	1,60107(3)	0,34063	1,24418	2,25294
	Optimal_02	0,92413	0,00105	0,92209	0,92637	0,91108	0,00669	0,8981	0,92591	0,8897(45)	0,00164	0,88668	0,89443	0,96868	0,00246	0,96371	0,97143	1,85181	0,33679	1,49303	3,19909
MinMax Scaler 0-10	Default	0,92455	0,00085	0,92315	0,92594	0,90828(49)	0,00268	0,90379	0,91329	0,88997	0,00133	0,88764	0,89253	0,9701	0,00048	0,9691	0,97111	2,38942	0,38795	1,62458	3,2492
	Optimal_01	0,92375(48)	0,00102	0,92135	0,92589	0,91318	0,00411	0,90705	0,92139	0,88944(51)	0,00132	0,88582	0,89203	0,96685(46)	0,00136	0,96394	0,96841	1,69526(5)	0,34325	1,24972	2,25294
	Optimal_02	0,92412(45)	0,00095	0,92209	0,92633	0,91171	0,00616	0,90351	0,92542	0,88973	0,00125	0,88754	0,89232	0,96823	0,00269	0,96371	0,97143	1,72706(9)	0,13458	1,49303	2,0308
MinMax Scaler 0-1000	Default	0,92457	0,00088	0,92315	0,92594	0,90851(45)	0,00275	0,90436	0,91329	0,89002	0,00131	0,88809	0,89253	0,97013	0,0005	0,9691	0,97111	2,62144	0,20906	2,37444	3,2492
	Optimal_01	0,92373(49)	0,00081	0,92176	0,92482	0,91428	0,0044	0,90705	0,92139	0,88954(50)	0,00093	0,8874	0,89136	0,9665(48)	0,00147	0,96394	0,9683	1,87488	0,27486	1,4087	2,25294
	Optimal_02	0,92383(46)	0,00097	0,92209	0,92633	0,91362	0,0063	0,90351	0,92542	0,88961(46)	0,00129	0,88754	0,89232	0,96722	0,00277	0,96371	0,97085	1,71665(8)	0,13801	1,53092	2,0308
MinMax Scaler 0-1000	Default	0,92451	0,00079	0,92355	0,92568	0,90802(52)	0,00292	0,90436	0,91309	0,88995	0,00126	0,88815	0,89199	0,97012	0,0003	0,96967	0,97059	2,64538	0,26321	2,37444	3,2492
	Optimal_01	0,92322(54)	0,00083	0,92176	0,9247	0,91811(3)	0,00209	0,91511	0,92139	0,88931(52)	0,00115	0,8874	0,89136	0,96516(52)	0,0007	0,96394	0,96649	2,12615	0,08847	1,97884	2,25294
	Optimal_02	0,92337(52)	0,00058	0,92209	0,92402	0,91899(1)	0,0025	0,91651	0,92542	0,88957(49)	0,0009	0,88754	0,89089	0,96459(54)	0,00058	0,96371	0,9657	1,79217	0,13926	1,67149	2,0308

Περιγραφή Αποτελεσμάτων Λογιστικής παλινδρόμησης

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,92626** σημείωσε το μοντέλο LR που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 3,17691 seconds**. Τον χαμηλότερο μέσο όρο precision= **0,92322**, μικρότερο κατά **0,00304 (-0,33%)**, σε σχέση με το μοντέλο LR με τη μέγιστη precision, σημείωσε το μοντέλο LR που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 2,12615 seconds**, που είναι χαμηλότερος κατά **1,05076 seconds (-33,07%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,91899** σημείωσε το μοντέλο LR που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 1,79217 seconds**. Τον χαμηλότερο μέσο όρο recall= **0,90785**, μικρότερο κατά **0,01114 (-1,21%)**, σε σχέση με το μοντέλο LR με τη μέγιστη recall, σημείωσε το μοντέλο LR που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 2,72038 seconds**, που είναι υψηλότερος κατά **0,92821 seconds (+51,79%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Baθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο (mean) F1 score= **0,89262** σημείωσε το μοντέλο LR που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler, με εύρος **0-1000**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **3,17691** seconds. Τον χαμηλότερο μέσο όρο F1 score= **0,88879**, ελάχιστα μικρότερο κατά **0,00383 (-0,43%)**, σε σχέση με το μοντέλο LR με το μέγιστο F1 Score, σημείωσε το μοντέλο LR που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,83556** seconds, που είναι χαμηλότερος κατά **1,34135** seconds (**-42,22%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

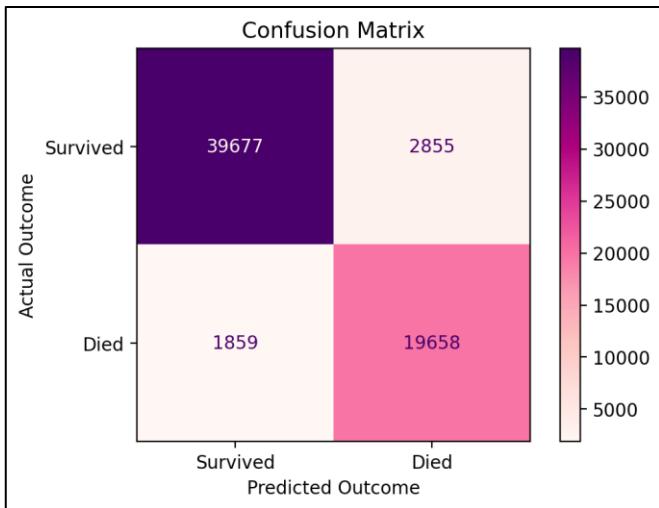
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,97081** σημείωσε το μοντέλο LR που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **2,99462** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,96459**, ελάχιστα μικρότερο κατά **0,00622 (-0,64%)**, σε σχέση με το μοντέλο LR με τη μέγιστη τιμή AUC_ROC, σημείωσε το μοντέλο LR που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,79217** seconds, που είναι χαμηλότερος κατά **1,20245** seconds (**-40,15%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

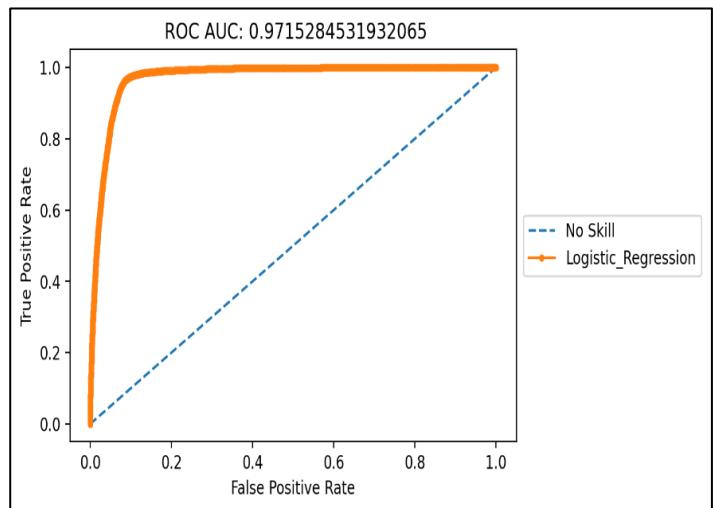
Τον χαμηλότερο μέσο όρο (mean) runtime= **1,3429** seconds σημείωσε το μοντέλο LR που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Standard** scaler, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο precision= **0,92448**. Τον υψηλότερο μέσο όρο runtime= **3,31659** seconds, υψηλότερο κατά **1,97369** seconds (**+146,97%**), σε σχέση με το μοντέλο LR με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο LR που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,92476**, ελάχιστα υψηλότερο κατά **0,00028 (+0,03%)**, σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα **LR** μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **default** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92626**, Recall: **0,91287**, F1-score: **0,89262**, AUC_ROC: **0,97047**, Runtime: **3,17691** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 68**) και η καμπύλη ROC (**Εικόνα 69**) του LR μοντέλου με την μέγιστη AUC_ROC.

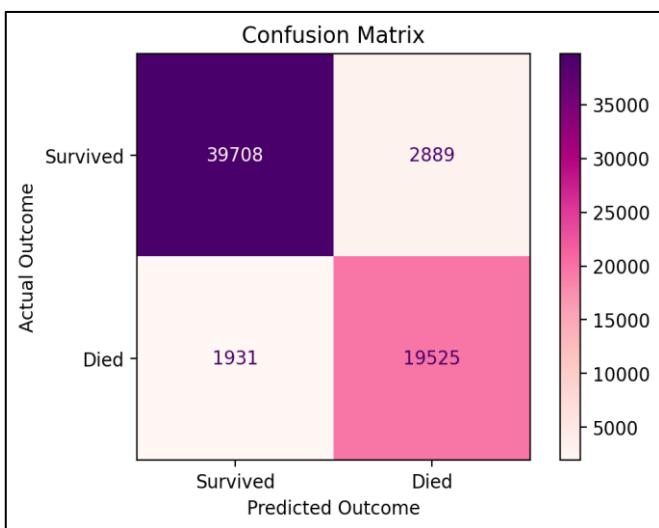


Εικόνα 68: Ο Πίνακας Σύγχυσης του **LR** μοντέλου **22_Min-Max_0-1000_default** με την μέγιστη **AUC_ROC**

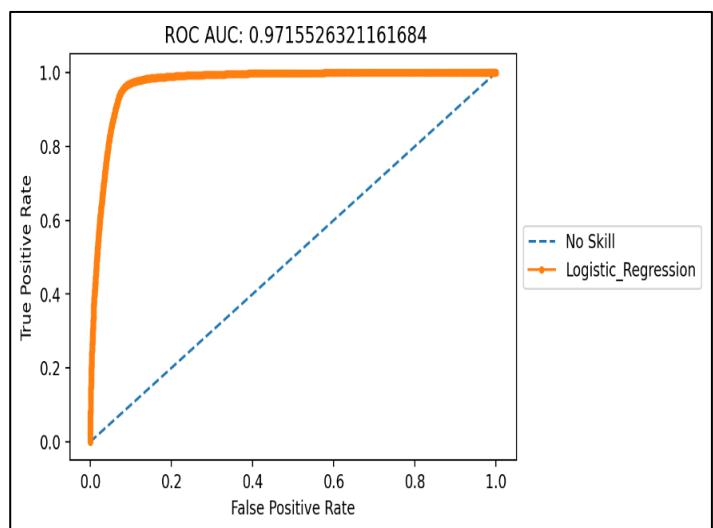


Εικόνα 69: Η καμπύλη ROC του **LR** μοντέλου **22_Min-Max_0-1000_default** με την μέγιστη **AUC_ROC**

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Standard** scaler και με τις **default** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92564**, Recall: **0,90994**, F1-score: **0,89136**, AUC_ROC: **0,97081**, Runtime: **2,99462** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 70**) και η καμπύλη ROC (**Εικόνα 71**) του LR μοντέλου με την μέγιστη AUC_ROC.

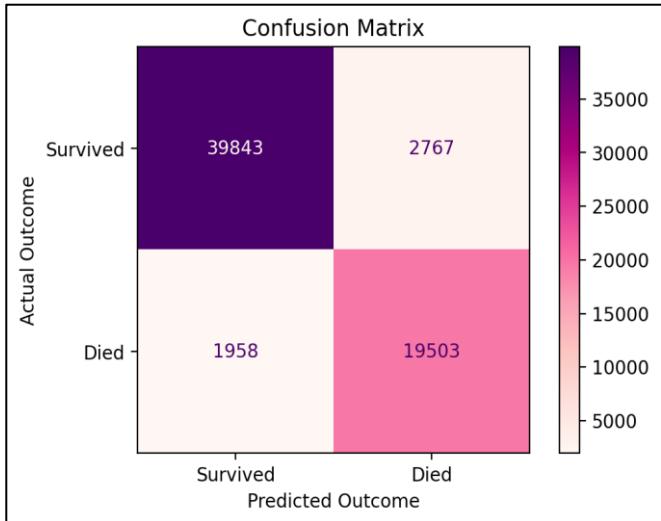


Εικόνα 70: Ο Πίνακας Σύγχυσης του **LR** μοντέλου **22_Standard_default** με την μέγιστη **AUC_ROC**

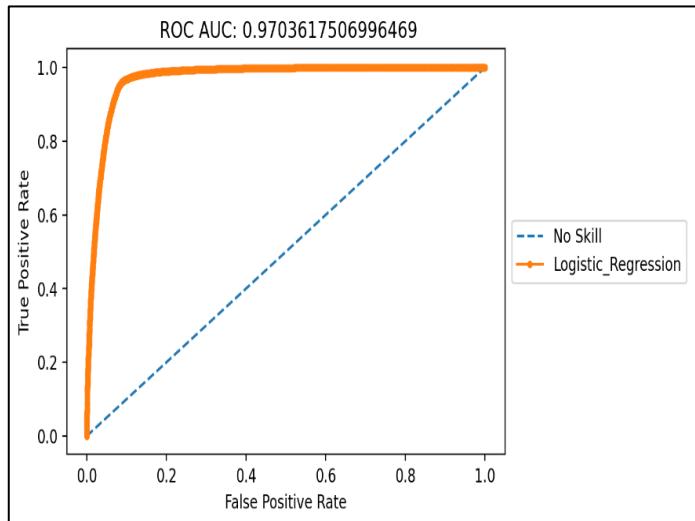


Εικόνα 71: Η καμπύλη ROC του **LR** μοντέλου **22_Standard_default** με την μέγιστη **AUC_ROC**

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-100** και με τις **default** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92562**, Recall: **0,91068**, F1-score: **0,89157**, AUC_ROC: **0,97041**, Runtime: **3,23525** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 72**) και η καμπύλη ROC (**Εικόνα 73**) του LR μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 72: Ο Πίνακας Σύγχυσης του **LR** μοντέλου **22_Min-Max_0-100_default** με την μέγιστη **AUC_ROC**



Εικόνα 73: Η καμπύλη ROC του **LR** μοντέλου **22_Min-Max_0-100_default** με την μέγιστη **AUC_ROC**

Με βάση τα παραπάνω για τον αλγόριθμο της LR, οι βέλτιστοι τρόποι scaling είναι ο **Min-Max** scaler, για τα εύρη **0-100** και **0-1000** και ο **Standard** scaler, με τις **default** υπερπαραμέτρους.

Για τα Δέντρα Απόφασης (DTs) έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 33)

Decision Trees		22 Features																				
Scaler Type	Parameter Type	Precision			Recall			F1-score				ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	
No Scaler	Default	0,90466(50)(320*)	0,00126	0,90217	0,9066	0,84284(50)(320*)	0,00281	0,83622	0,84604	0,85583(50)(320*)	0,00176	0,85218	0,8586	0,90454(50)(320*)	0,00237	0,90021	0,90901	1,72228(54)	0,42124	1,53088	2,9114	
	Optimal_01	0,92927(4)	0,00121	0,92724	0,93069	0,93538(4)	0,00297	0,93089	0,94011	0,89878(5)	0,00163	0,89577	0,90079	0,95577(3)	0,00096	0,9548	0,95718	1,45988(46)	0,06985	1,40595	1,64027	
	Optimal_02	0,92075	0,0012	0,91894	0,923	0,90394	0,00543	0,89564	0,91404	0,88446	0,00211	0,88178	0,88875	0,93035	0,00193	0,92756	0,93249	1,465(47)	0,11616	1,35525	1,76088	
Standard Scaler	Default	0,90092(54)(324*)	0,00117	0,89816	0,90204	0,8342(54)(324*)	0,00206	0,83095	0,83734	0,84959(54)(324*)	0,00185	0,84544	0,85159	0,90031(54)(324*)	0,00143	0,89689	0,902	1,61526(49)	0,07444	1,51536	1,74959	
	Optimal_01	0,92683	0,00139	0,92438	0,92819	0,92906	0,00336	0,92389	0,93503	0,89509	0,00202	0,89125	0,89715	0,95091	0,00151	0,94851	0,95334	1,49245(48)	0,14269	1,40592	1,88159	
	Optimal_02	0,91258(313*)	0,0013	0,90989	0,91426	0,87875(306*)	0,00353	0,87314	0,88449	0,87087(311*)	0,00197	0,86667	0,87322	0,91735(47)(317*)	0,00238	0,91229	0,92043	1,36029	0,0369	1,31218	1,43717	
MinMax Scaler 0-1	Default	0,90292(53)(323*)	0,00203	0,89924	0,90726	0,83897(53)(323*)	0,00504	0,82925	0,84926	0,85306(53)(323*)	0,00329	0,8472	0,85982	0,90268(53)(323*)	0,00217	0,89787	0,90687	1,62726(50)	0,08741	1,50063	1,93679	
	Optimal_01	0,92775	0,00185	0,92459	0,93158	0,93063	0,00425	0,92505	0,93929	0,89639	0,00274	0,89213	0,90232	0,9518	0,00332	0,94648	0,95847	1,44535	0,0484	1,37465	1,57625	
	Optimal_02	0,91549	0,00474	0,90964	0,92469	0,88741	0,01407	0,86929	0,91437	0,87573	0,00776	0,8661	0,8904	0,92234(311*)	0,00871	0,91094	0,93936	1,4058	0,05432	1,31042	1,5382	
MinMax Scaler 0-10	Default	0,90333(52)(322*)	0,00199	0,89924	0,90726	0,83977(52)(322*)	0,00514	0,83251	0,84926	0,85375(52)(322*)	0,00319	0,84827	0,85982	0,90312(52)(322*)	0,00208	0,89787	0,90687	1,63808(52)	0,09393	1,53463	1,93679	
	Optimal_01	0,92819(9)	0,0019	0,92459	0,93158	0,93168	0,00432	0,92505	0,93929	0,89706(10)	0,00281	0,89214	0,90232	0,95273	0,00326	0,94648	0,95847	1,44494	0,04815	1,37465	1,57625	
	Optimal_02	0,91659	0,00493	0,90964	0,92469	0,89059	0,01481	0,86929	0,91437	0,87752	0,00808	0,8661	0,8904	0,92441(309*)	0,00908	0,91094	0,93936	1,41971	0,04935	1,34201	1,5382	
MinMax Scaler 0-100	Default	0,90436(51)(321*)	0,00153	0,90144	0,90726	0,84238(51)(321*)	0,00423	0,8353	0,84926	0,85542(51)(321*)	0,00247	0,85134	0,85982	0,90423(51)(321*)	0,00129	0,90171	0,90687	1,65894(53)	0,10432	1,53463	1,93679	
	Optimal_01	0,92899(6)	0,00153	0,92721	0,93158	0,93296(8)	0,00438	0,92687	0,93929	0,89823(6)	0,00233	0,89549	0,90232	0,95424(7)	0,00271	0,95089	0,95847	1,44478	0,04781	1,37465	1,57581	
	Optimal_02	0,919	0,00423	0,91347	0,92469	0,89744	0,01328	0,87811	0,91437	0,88148	0,00692	0,87198	0,8904	0,92891	0,00765	0,9168	0,93936	1,44073	0,04281	1,37468	1,5382	
MinMax Scaler 0-1000	Default	0,90546(49)(319*)	0,0011	0,90424	0,90726	0,84593(49)(319*)	0,00207	0,84212	0,84926	0,85737(49)(319*)	0,00161	0,85573	0,85982	0,90502(49)(319*)	0,00103	0,90354	0,90687	1,63556(51)	0,07802	1,53463	1,80631	
	Optimal_01	0,93035(1)	0,00081	0,9292	0,93158	0,93688(2)	0,00167	0,93374	0,93929	0,90032(1)	0,00116	0,89885	0,90232	0,95666(1)	0,00129	0,95477	0,95847	1,45128(45)	0,061	1,37465	1,57581	
	Optimal_02	0,92299	0,00115	0,92153	0,92469	0,90996	0,00274	0,90738	0,91437	0,88803	0,00164	0,88605	0,8904	0,93599	0,00184	0,93326	0,93936	1,45123	0,04262	1,40595	1,5382	

❖ Για τα 15 χαρακτηριστικά (Πίνακας 34)

Decision Trees		15 Features																				
Scaler Type	Parameter Type	Precision			Recall			F1-score				ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	
No Scaler	Default	0,91191(314*)	0,00108	0,91075	0,91388	0,86633(314*)	0,00361	0,85919	0,87166	0,86837(314*)	0,00191	0,86603	0,87179	0,91992(312*)	0,00176	0,91723	0,92229	1,33817	0,05196	1,26536	1,43194	
	Optimal_01	0,9292(5)	0,00114	0,92777	0,93072	0,93662(3)	0,00388	0,93042	0,94117	0,89886(4)	0,00185	0,89623	0,9011	0,95542(4)	0,00152	0,95382	0,95896	1,27464	0,04585	1,22756	1,34547	
	Optimal_02	0,92137	0,00137	0,91948	0,92323	0,91008	0,00365	0,90637	0,9196	0,88604	0,002	0,88298	0,88942	0,93542	0,00118	0,93397	0,93749	1,20965(20*)	0,03486	1,17161	1,26532	
Standard Scaler	Default	0,90921(48)(318*)	0,00107	0,90738	0,91101	0,86004(48)(318*)	0,00326	0,85537	0,86617	0,8641(48)(318*)	0,00157	0,86141	0,86706	0,91647(48)(318*)	0,00121	0,9144	0,91828	1,32557	0,04407	1,28988	1,43056	
	Optimal_01	0,92739	0,00093	0,92529	0,92932	0,93071	0,00292	0,9265	0,93481	0,89602	0,0013	0,89352	0,89857	0,95277	0,00143	0,95056	0,95588	1,27687	0,04852	1,22693	1,37471	
	Optimal_02	0,91442(307*)	0,00102	0,91255	0,91561	0,88061	0,00329	0,87452	0,88614	0,87354(306*)	0,0017	0,87012	0,87583	0,92241(310*)	0,00164	0,9193	0,9246	1,15652(11*)	0,01837	1,13832	1,1871	
MinMax Scaler 0-1	Default	0,91024(47)(317*)	0,00182	0,90625	0,91434	0,86266(47)(317*)	0,00504	0,85602	0,87276	0,86578(47)(317*)	0,00294	0,85954	0,87237	0,91768(46)(316*)	0,00219	0,91242	0,92202	1,33292	0,07018	1,26533	1,54131	
	Optimal_01	0,92783	0,00178	0,92517	0,93246	0,93202(10)	0,00454	0,92163	0,94256	0,89665	0,00265	0,89259	0,90333	0,95238	0,0027	0,94841	0,95765	1,25739	0,0423	1,20098	1,35154	
	Optimal_02	0,91669	0,00408	0,91049	0,92462	0,88921	0,01445	0,8719	0,91676	0,87751	0,0069	0,86736	0,89078	0,92677(306*)	0,0069	0,9186	0,93935	1,21973	0,07157	1,13281	1,4684	
MinMax Scaler 0-10	Default	0,91071(46)(316*)	0,00173	0,90838	0,91434	0,86354(46)(316*)	0,00526	0,85614	0,87276	0,86649(46)(316*)	0,00286	0,8626	0,87237	0,91826(45)(315*)	0,00192	0,91439	0,92202	1,33636	0,07734	1,26533	1,54131	
	Optimal_01	0,92813(10)	0,00192	0,92517	0,93246	0,93271(9)	0,00463	0,92163	0,94256	0,89709(9)	0,00286	0,89259	0,90333	0,95292(10)	0,00286	0,94841	0,95765	1,26166	0,04554	1,20098	1,35154	
	Optimal_02	0,91749	0,00438	0,91049	0,92462	0,89274	0,01501	0,8719	0,91676	0,87897	0,00733	0,86736	0,89078	0,92831	0,00729	0,9186	0,93935	1,22565	0,06855	1,15431	1,4684	
MinMax Scaler 0-100	Default	0,91138(45)(315*)	0,00171	0,90908	0,91434	0,86558(45)(315*)	0,0053	0,85614	0,87276	0,86758(45)(315*)	0,00288	0,86345	0,87237	0,91916(314*)	0,00156	0,91657	0,92202	1,3383	0,08059	1,26533	1,54131	
	Optimal_01	0,92861(8)	0,00198	0,92517	0,93246	0,93343(7)	0,00479	0,92393	0,94256	0,89779(8)	0,00295	0,89259	0,90333	0								

❖ Για τα 10 χαρακτηριστικά (Πίνακας 35)

Decision Trees		10 Features																			
Scaler Type	Parameter Type	Precision				Recall				F1-score				ROC AUC				Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max												
No Scaler	Default	0,91531(305*)	0,00098	0,91327	0,91679	0,87689(307*)	0,00231	0,87241	0,88023	0,87437(305*)	0,0015	0,87138	0,87677	0,92888	0,00092	0,92785	0,93079	1,19258(19*)	0,02907	1,15599	1,24971
	Optimal_01	0,92875(7)	0,00101	0,92712	0,93053	0,93452(5)	0,00239	0,93079	0,93846	0,89806(7)	0,0014	0,8961	0,90074	0,95526(5)	0,00118	0,95358	0,95709	1,15722(12*)	0,04364	1,12101	1,26352
	Optimal_02	0,92156	0,00215	0,917	0,92441	0,90523	0,00579	0,89451	0,91486	0,88574	0,00317	0,87926	0,89029	0,93787	0,00164	0,93566	0,94055	1,13486(6)(6*)	0,05217	1,0913	1,24525
Standard Scaler	Default	0,91272(311*)	0,00096	0,91133	0,91459	0,87226(312*)	0,00196	0,86995	0,87603	0,87028(312*)	0,00146	0,86843	0,87325	0,92576(308*)	0,00107	0,92332	0,92673	1,18447(14*)	0,02636	1,15392	1,23245
	Optimal_01	0,92696	0,00077	0,92595	0,92812	0,92932	0,00281	0,92478	0,93383	0,89527	0,00118	0,89319	0,89693	0,95219	0,00098	0,95028	0,95366	1,15092(8)(8*)	0,02615	1,12422	1,20777
	Optimal_02	0,91347(310*)	0,00131	0,91201	0,91551	0,87623(309*)	0,00292	0,87115	0,87992	0,87168(310*)	0,00207	0,86871	0,87466	0,92644(307*)	0,00227	0,92202	0,92887	1,09347(2)(2*)	0,04634	1,04666	1,1716
MinMax Scaler 0-1	Default	0,91441(309*)	0,0017	0,90971	0,91781	0,87472(311*)	0,00419	0,86728	0,88573	0,87239(309*)	0,00274	0,86556	0,87818	0,92714(305*)	0,00164	0,92305	0,9296	1,18687(15*)	0,04132	1,1385	1,29657
	Optimal_01	0,92719	0,00191	0,92368	0,93051	0,92828	0,00423	0,92005	0,937	0,89543	0,00281	0,89048	0,90022	0,95208	0,00196	0,94843	0,95609	1,16534(13*)	0,05828	1,10455	1,35723
	Optimal_02	0,91717	0,0042	0,91276	0,92606	0,88751	0,01574	0,87214	0,91924	0,87781	0,00727	0,87018	0,89295	0,9316	0,00617	0,92476	0,94371	1,09218(1)(1*)	0,04051	1,02941	1,17158
MinMax Scaler 0-10	Default	0,91436(308*)	0,00171	0,90971	0,91781	0,87533(310*)	0,00442	0,8684	0,88573	0,87279(308*)	0,00279	0,86556	0,87818	0,92737	0,00171	0,92305	0,9296	1,18745(17*)	0,04214	1,1385	1,29657
	Optimal_01	0,92752	0,00207	0,92368	0,93051	0,92901	0,00449	0,92005	0,937	0,89593	0,00304	0,89048	0,90022	0,95242	0,00209	0,94843	0,95609	1,15624(10)(10*)	0,04137	1,10455	1,27034
	Optimal_02	0,91809	0,00444	0,91314	0,92606	0,89085	0,01685	0,87303	0,91924	0,87938	0,00772	0,8712	0,89295	0,93299	0,00651	0,92476	0,94371	1,09678(3)(3*)	0,04036	1,02941	1,17158
MinMax Scaler 0-100	Default	0,91465(306*)	0,00193	0,90971	0,91781	0,87639(308*)	0,00463	0,8684	0,88573	0,87333(307*)	0,00313	0,86556	0,87818	0,92762	0,00175	0,9234	0,9296	1,18927(18*)	0,04317	1,15076	1,29657
	Optimal_01	0,92801	0,00213	0,92387	0,93051	0,93011	0,00442	0,92153	0,937	0,89664	0,00315	0,89092	0,90022	0,95301(9)	0,00215	0,94843	0,95609	1,15345(9)(9*)	0,04052	1,10455	1,24972
	Optimal_02	0,91965	0,00464	0,91314	0,92606	0,8973	0,01714	0,87699	0,91924	0,88217	0,00799	0,87219	0,89295	0,93546	0,00659	0,92639	0,94371	1,10673(4)(4*)	0,0336	1,06039	1,16619
MinMax Scaler 0-1000	Default	0,91603	0,00122	0,91375	0,91781	0,88012(305*)	0,00284	0,87543	0,88573	0,87562	0,00195	0,87173	0,87818	0,92841	0,00121	0,9253	0,92959	1,18742(16*)	0,0433	1,15076	1,2653
	Optimal_01	0,92984(3)	0,00069	0,92827	0,93051	0,93372(6)	0,0024	0,92996	0,937	0,89936(3)	0,00104	0,89688	0,90022	0,95488(6)	0,00077	0,95364	0,95609	1,14701(7)(7*)	0,03435	1,10455	1,20103
	Optimal_02	0,92397	0,00139	0,92179	0,92606	0,91373	0,00267	0,90951	0,91924	0,88972	0,00194	0,88723	0,89295	0,94167	0,00169	0,93913	0,94371	1,11778(5)(5*)	0,03277	1,0885	1,16619

Περιγραφή Αποτελεσμάτων Δέντρων Απόφασης

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,93035** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,45128** seconds. Τον χαμηλότερο μέσο όρο precision= **0,90092**, μικρότερο κατά **0,02943** (-3,16%), σε σχέση με το μοντέλο DT με τη μέγιστη precision, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,61526** seconds, που είναι υψηλότερος κατά **0,16398** seconds (+11,30%), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,9369** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,25677** seconds. Τον χαμηλότερο μέσο όρο recall= **0,8342**, μικρότερο κατά **0,1027** (-10,96%), σε σχέση με το μοντέλο DT με τη μέγιστη recall, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,61526** seconds, που είναι υψηλότερος κατά **0,35849** seconds (+28,52%), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Baθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο (mean) F1 score= **0,90032** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler, με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,45128** seconds. Τον χαμηλότερο μέσο όρο F1 score= **0,84959**, μικρότερο κατά **0,05073 (-5,63%)**, σε σχέση με το μοντέλο DT με το μέγιστο F1 Score, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,61526** seconds, που είναι υψηλότερος κατά **0,16398** seconds (**+11,30%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

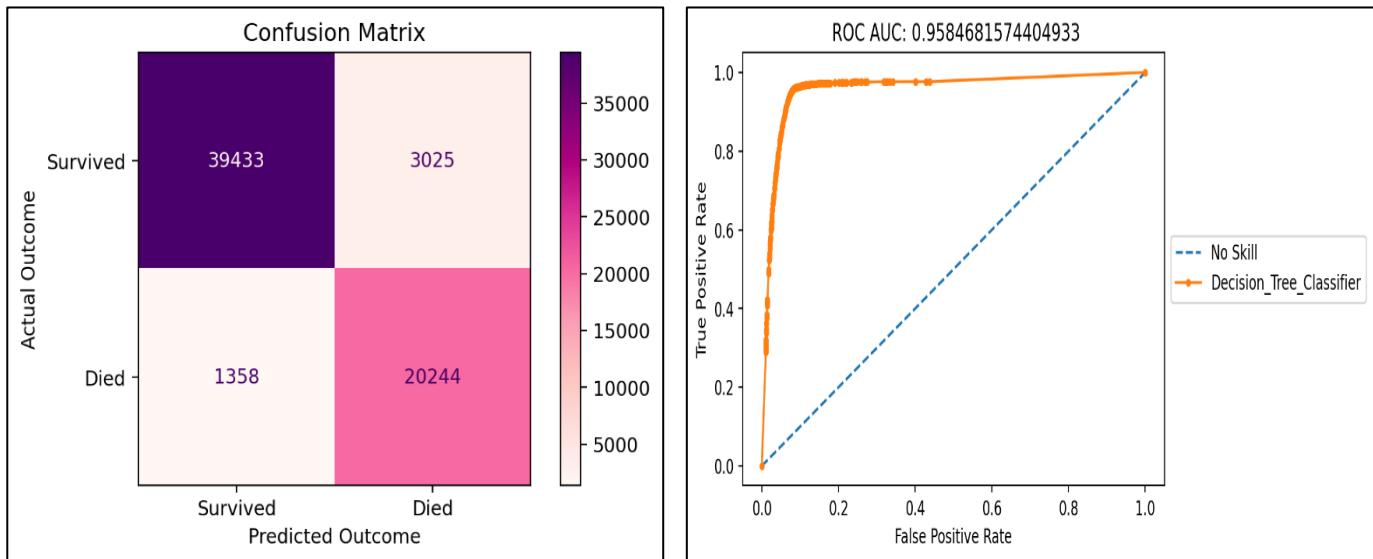
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,95666** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,45128** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,90031**, μικρότερο κατά **0,05635 (-5,89%)**, σε σχέση με το μοντέλο DT με τη μέγιστη τιμή AUC_ROC, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **1,61526** seconds, που είναι υψηλότερος κατά **0,16398** seconds (**+11,30%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

Τον χαμηλότερο μέσο όρο (mean) runtime= **1,09218** seconds σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο precision= **0,91717**. Τον υψηλότερο μέσο όρο runtime= **1,72228** seconds, υψηλότερο κατά **0,63015** seconds (**+57,69%**), σε σχέση με το μοντέλο DT με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,90466**, χαμηλότερο κατά **0,01251** (**+1,36%**), σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα DTs μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

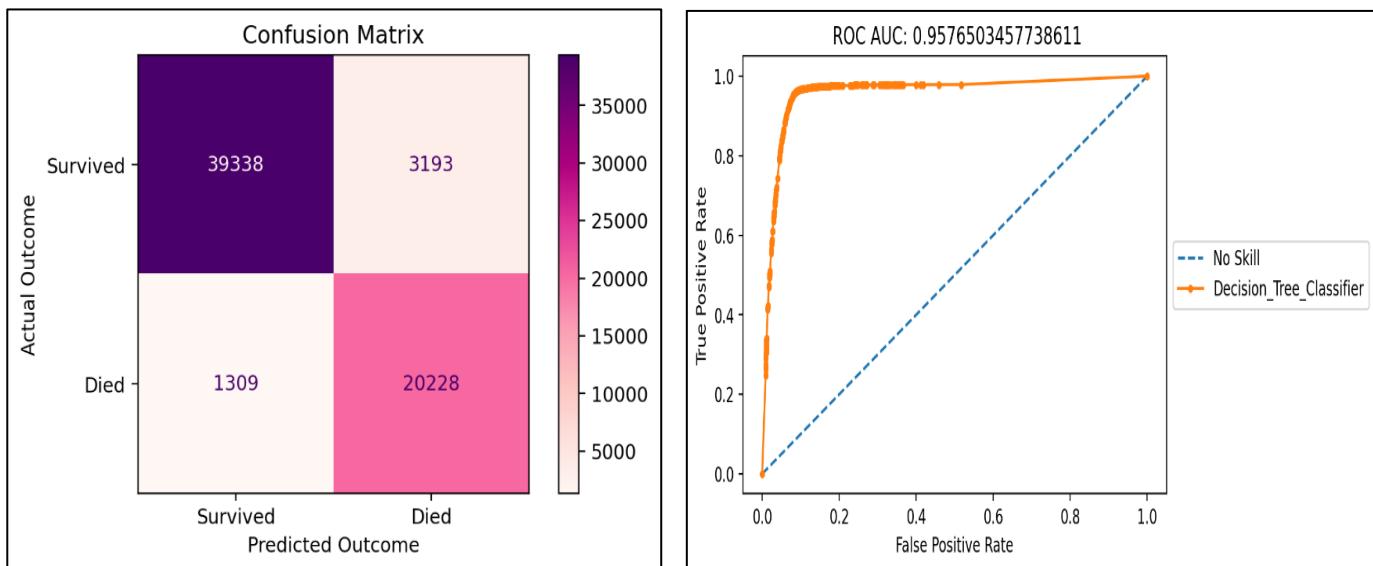
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93035**, Recall: **0,93688**, F1-score: **0,90032**, AUC_ROC: **0,95666**, Runtime: **1,45128** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 74**) και η καμπύλη ROC (**Εικόνα 75**) του DT μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 74: Ο Πίνακας Σύγχυσης του DT μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

Εικόνα 75: Η καμπύλη ROC του DT μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

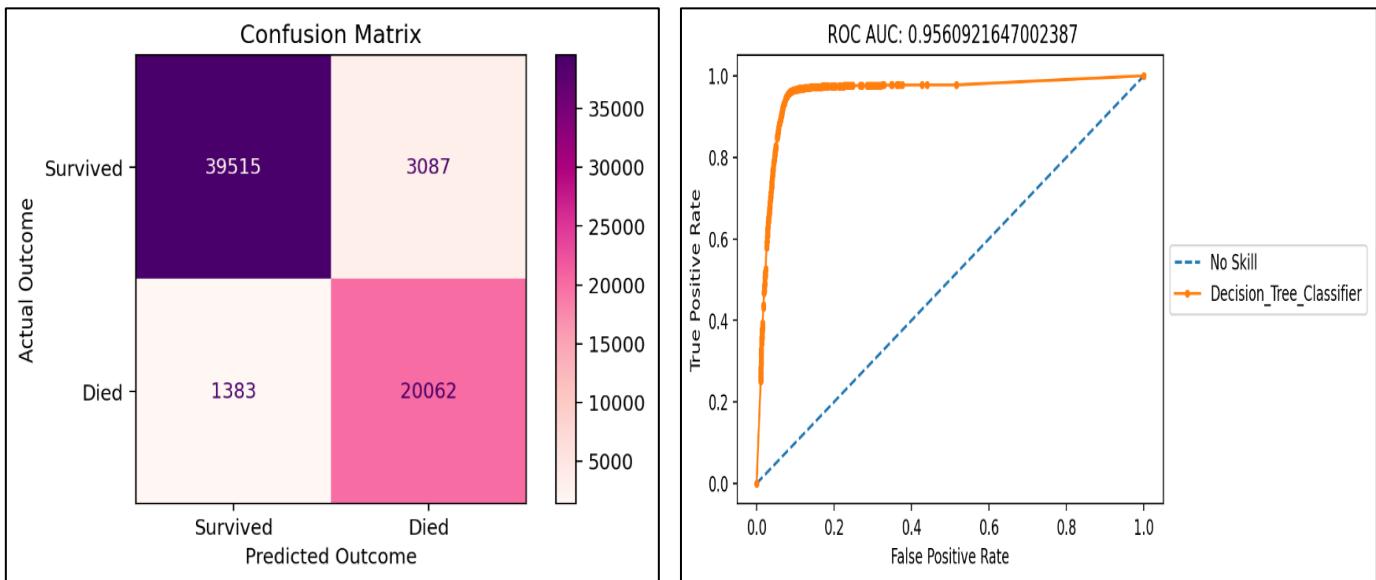
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **15** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93025**, Recall: **0,9369**, F1-score: **0,90022**, AUC_ROC: **0,95632**, Runtime: **1,25677** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 76**) και η καμπύλη ROC (**Εικόνα 77**) του DT μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 76: Ο Πίνακας Σύγχυσης του DT μοντέλου **15_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

Εικόνα 77: Η καμπύλη ROC του DT μοντέλου **15_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **10** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92984**, Recall: **0,93372**, F1-score: **0,89936**, AUC_ROC: **0,95488**, Runtime: **1,14701** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 78**) και η καμπύλη ROC (**Εικόνα 79**) του DT μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 78: Ο Πίνακας Σύγχυσης του DT μοντέλου
10 _Min-Max 0-1000_opt_01 με την μέγιστη AUC_ROC

Εικόνα 79: Η καμπύλη ROC του DT μοντέλου
10 _Min-Max 0-1000_opt_01 με την μέγιστη AUC_ROC

Με βάση τα παραπάνω είναι εμφανές πως για τα DTs, ο βέλτιστος scaler είναι ο **Min-Max**, για το εύρος **0-1000** με τις **optimal_01** υπερπαραμέτρους.

Για τον αλγόριθμο του **Τυχαίο Δάσος** (RF) έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 36)

Random Forest		22 Features																			
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92474(48)	0,00095	0,92262	0,92581	0,91012(52)	0,00164	0,90773	0,91328	0,8904(48)	0,00131	0,88744	0,89162	0,96693	0,00057	0,96586	0,96769	16,79225	0,38048	16,3478	17,4135
	Optimal_01	0,93134	0,00065	0,93032	0,93229	0,94155	0,00132	0,93991	0,94379	0,90202	0,00095	0,90043	0,9034	0,97076	0,00058	0,96994	0,97197	24,4617	0,43547	23,9231	25,3763
	Optimal_02	0,93518(5)	0,00148	0,93295	0,9368	0,9676(3)(3*)	0,00109	0,96528	0,96927	0,90929(5)(15*)	0,00194	0,90631	0,91159	0,97525(3)(17*)	0,00074	0,97418	0,97656	30,22087(54)	1,15381	29,4675	33,4112
Standard Scaler	Default	0,92253(54)	0,00115	0,92102	0,92476	0,90853(53)	0,00212	0,90587	0,91155	0,8873(54)	0,00171	0,88503	0,8907	0,96526	0,00073	0,96426	0,96685	16,67103	0,5214	16,1407	17,8948
	Optimal_01	0,92778	0,00141	0,92537	0,92991	0,9327	0,00167	0,9301	0,93488	0,89672	0,00183	0,89367	0,89955	0,96795	0,00083	0,96671	0,96896	24,80411	1,26947	23,636	28,1608
	Optimal_02	0,93307	0,00083	0,93168	0,93411	0,96246(4*)	0,00108	0,96054	0,96403	0,90621	0,00106	0,90463	0,9077	0,97279(10)	0,00049	0,97204	0,97373	29,51955(49)	0,32479	29,0315	29,9725
MinMax Scaler 0-1	Default	0,92443(50)	0,00145	0,92192	0,92713	0,91134(49)	0,00241	0,90537	0,91727	0,89007(50)	0,00204	0,88662	0,89399	0,96632	0,00132	0,96365	0,96937	16,66155	0,46885	16,0034	18,9838
	Optimal_01	0,92913	0,00195	0,92586	0,93222	0,93531	0,00505	0,92708	0,94472	0,89862	0,00296	0,89341	0,90342	0,96916	0,0017	0,96604	0,97259	24,68037	1,64004	23,704	34,087
	Optimal_02	0,93433(9)	0,00173	0,93049	0,93735	0,9651(9)(9*)	0,00325	0,95996	0,97144	0,90801(9)	0,00243	0,90264	0,91218	0,97428(7)	0,00177	0,97209	0,978	29,66521(50)	0,29253	29,1482	30,2905
MinMax Scaler 0-10	Default	0,92478(47)	0,0014	0,92212	0,92713	0,91184(47)	0,00219	0,9076	0,91727	0,89055(47)	0,00196	0,88662	0,89399	0,96658	0,00138	0,96365	0,96937	16,73792	0,50114	16,1658	18,9838
	Optimal_01	0,92979	0,00174	0,92643	0,93222	0,93695	0,00467	0,92708	0,94472	0,89962	0,00263	0,89458	0,90342	0,96963	0,00168	0,96604	0,97259	24,64248	1,82656	23,704	34,087
	Optimal_02	0,93472(7)	0,00168	0,93095	0,93735	0,96612(6)(6*)	0,00307	0,96156	0,97144	0,90857(7)	0,00236	0,90358	0,91218	0,97476(5)	0,00179	0,97254	0,978	29,67527(51)	0,30249	29,1482	30,2905
MinMax Scaler 0-100	Default	0,92529	0,00139	0,92212	0,92713	0,91206(45)	0,00254	0,9076	0,91727	0,89121(45)	0,00203	0,88662	0,89399	0,96699	0,00146	0,96365	0,96937	16,88867	0,53895	16,3766	18,9838
	Optimal_01	0,9308	0,00089	0,92903	0,93222	0,93985	0,00208	0,93635	0,94472	0,90118	0,00125	0,89826	0,90342	0,97058	0,00109	0,96851	0,97259	24,72417	2,23829	23,7048	34,087
	Optimal_02	0,93544(4)(19*)	0,00145	0,93263	0,93735	0,96765(2)(2*)	0,0025	0,96366	0,97144	0,90962(4)(14*)	0,00201	0,90619	0,91218	0,97554(2)(14*)	0,00167	0,97292	0,978	29,71669(52)	0,32943	29,1482	30,2905
MinMax Scaler 0-1000	Default	0,92604	0,00079	0,92509	0,92713	0,91275	0,00218	0,90966	0,91727	0,89226	0,00118	0,8907	0,89399	0,96805	0,00077	0,96646	0,96937	16,91687	0,7582	16,3766	18,9838
	Optimal_01	0,93091	0,00104	0,92928	0,93222	0,94003	0,00209	0,93673	0,94472	0,90134	0,00139	0,89915	0,90342	0,97149	0,00045	0,97105	0,97259	23,99286	0,2372	23,7048	24,4525
	Optimal_02	0,93661(1)(11*)	0,00055	0,93577	0,93735	0,96999(1)(1*)	0,00084	0,96844	0,97144	0,91127(1)(2*)	0,00073	0,91007	0,91218	0,97708(1)(11*)	0,00053	0,97654	0,978	29,84745(53)	0,32512	29,1482	30,2905

❖ Για τα 15 χαρακτηριστικά (Πίνακας 37)

Random Forest		15 Features																			
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92483(46)	0,00114	0,92335	0,92668	0,91143(48)	0,00274	0,90716	0,91538	0,89068(46)	0,0017	0,88832	0,89323	0,96532	0,00078	0,96398	0,96654	14,03207(9)	0,07525	13,8584	14,1156
	Optimal_01	0,92807	0,00046	0,92738	0,92877	0,9308	0,0019	0,92631	0,93271	0,8967	0,00068	0,89563	0,89759	0,96734	0,00037	0,96664	0,96774	20,9451	0,58172	20,2885	22,4
	Optimal_02	0,93551(3)(17*)	0,00127	0,93332	0,93757	0,96631(5)(5*)	0,00104	0,96464	0,96794	0,90963(3)(13*)	0,00163	0,90663	0,91214	0,97441(6)	0,0006	0,97322	0,97523	26,67779(46)	0,21783	26,426	26,986
Standard Scaler	Default	0,92306(53)	0,001	0,92131	0,92513	0,90832(54)	0,00142	0,90662	0,91055	0,88795(53)	0,00138	0,88572	0,8908	0,96379	0,00081	0,96261	0,96508	13,94055(7)	0,17026	13,7999	14,2839
	Optimal_01	0,92524(45)	0,001	0,92344	0,92724	0,92577	0,00241	0,92293	0,92973	0,89266	0,00133	0,89038	0,89532	0,96499	0,00048	0,96441	0,96598	21,6886	0,45769	20,954	22,2862
	Optimal_02	0,93196	0,001	0,93039	0,93375	0,96213(16*)	0,00163	0,95935	0,96448	0,90468	0,00125	0,90312	0,90723	0,971	0,00045	0,97033	0,97167	26,24805	0,27906	25,8138	26,6569
MinMax Scaler 0-1	Default	0,92375(52)	0,0015	0,92136	0,92774	0,91022(51)	0,00303	0,90469	0,91787	0,88916(52)	0,00218	0,88597	0,89477	0,96462	0,00114	0,96284	0,96717	14,03716(10)	0,28661	13,5257	14,7541
	Optimal_01	0,92697	0,00155	0,92403	0,92999	0,92945	0,00296	0,92437	0,93553	0,8952	0,00223	0,89124	0,89927	0,96601	0,0013	0,96366	0,96837	20,90273	0,31092	20,3703	21,6283
	Optimal_02	0,93388	0,00177	0,93077	0,93797	0,96369(11*)	0,00231	0,9594	0,96853	0,90734	0,00246	0,9032	0,91272	0,97253	0,00171	0,9696	0,97547	26,68521(47)	0,44867	25,3879	27,6299
MinMax Scaler 0-10	Default	0,92397(51)	0,00156	0,92143	0,92774	0,91074(50)	0,00324	0,90469	0,91787	0,88955(51)	0,00228	0,88606	0,89477	0,96482	0,00116	0,96291	0,96717	14,0768	0,23627	13,7067	14,5763
	Optimal_01	0,92729	0,00157	0,92403	0,92999	0,92993	0,0031	0,92437	0,93553	0,89568	0,00224	0,89124	0,89927	0,96634	0,00128	0,96366	0,96837	20,92183	0,33789	20,3703	21,6283
	Optimal_02	0,93431(10)	0,00173	0,93194	0,93797	0,96435(10)(10*)	0,00228	0,9594	0,96853	0,90795(10)	0,00241	0,90445	0,91272	0,97305(9)	0,0016	0,97067	0,97547	26,59301	0,41215	25,3879	27,3778
MinMax Scaler 0-100	Default	0,92454(49)	0,00157	0,92178	0,92774	0,91188(46)	0,00315	0,90469	0,91787	0,89033(49)	0,0023	0,88606	0,89477	0,96519	0,00117	0,96291	0,96717	14,10751	0,22541	13,7352	14,5763
	Optimal_01	0,92796	0,00133	0,92531	0,92999	0,93121	0,00259	0,9259	0,93553	0,89665	0,00191	0,89324	0,89927	0,96701	0,00096	0,96525	0,96837	20,94472	0,38334	20,3703	21,6283
	Optimal_02	0,93508(6)	0,00158	0,93246	0,93797	0,96528(8)(8*)	0,00205	0,9613													

❖ Για τα 10 χαρακτηριστικά (Πίνακας 38)

Random Forest		10 Features																			
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)							
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92623	0,00089	0,92468	0,92745	0,9234	0,00165	0,92163	0,92598	0,89361	0,00131	0,89147	0,89554	0,96232(52)	0,00063	0,96154	0,96307	12,22819(6)	0,17556	11,9693	12,484
	Optimal_01	0,92819	0,00116	0,92649	0,93065	0,94131	0,00196	0,93918	0,94403	0,89797	0,00174	0,89561	0,9018	0,96301(46)	0,00103	0,96135	0,96456	18,19271	0,32126	17,7096	18,7561
	Optimal_02	0,93372	0,00119	0,93195	0,93531	0,96334(12*)	0,00152	0,96149	0,9656	0,90704	0,00154	0,90469	0,90915	0,97011	0,0006	0,9691	0,97095	24,27967	0,26255	23,979	24,9293
Standard Scaler	Default	0,92545	0,00086	0,92441	0,92664	0,92087	0,00213	0,91764	0,9252	0,89251	0,00128	0,89102	0,89467	0,96185(53)	0,00073	0,96095	0,96348	12,03434(2)	0,24121	11,7422	12,5112
	Optimal_01	0,92657	0,00105	0,92516	0,92873	0,93686	0,00189	0,93413	0,94022	0,89547	0,00151	0,89343	0,89856	0,96155(54)	0,00088	0,96053	0,96304	18,33744	0,46129	17,5134	18,8525
	Optimal_02	0,93099	0,00095	0,92944	0,93266	0,95997(18*)	0,00169	0,95785	0,96355	0,90323	0,00137	0,90102	0,90546	0,96855	0,00065	0,96751	0,96939	24,19302	1,02583	23,4749	27,0415
MinMax Scaler 0-1	Default	0,92636	0,0014	0,92271	0,92941	0,92258	0,00337	0,91618	0,92899	0,89377	0,00207	0,88848	0,89797	0,96243(51)	0,00106	0,95961	0,96479	12,08051(5)	0,22811	11,7684	12,6365
	Optimal_01	0,92799	0,00146	0,92462	0,93055	0,94028	0,00278	0,9348	0,94579	0,89763	0,00209	0,89251	0,9015	0,96277(49)	0,0009	0,96099	0,96459	18,15152	0,4553	17,6156	20,3745
	Optimal_02	0,93274	0,00139	0,93012	0,93652	0,96185(17*)	0,00271	0,95824	0,96767	0,9057	0,00198	0,90192	0,91103	0,96901	0,001	0,96717	0,97145	24,48712	0,55894	23,4526	26,778
MinMax Scaler 0-10	Default	0,9265	0,00152	0,92271	0,92941	0,92312	0,00344	0,91618	0,92899	0,89398	0,00225	0,88848	0,89797	0,96251(50)	0,00115	0,95961	0,96479	12,06676(4)	0,22322	11,7684	12,5566
	Optimal_01	0,92837	0,00135	0,92462	0,93055	0,94066	0,00299	0,9348	0,94579	0,89813	0,002	0,89251	0,9015	0,96292(47)	0,00094	0,96103	0,96459	18,10995	0,49657	17,6156	20,3745
	Optimal_02	0,93278	0,00152	0,93012	0,93652	0,96235(15*)	0,0029	0,95824	0,96767	0,90579	0,00218	0,90192	0,91103	0,96927	0,00095	0,96737	0,97145	24,28707	0,36631	23,4526	25,0864
MinMax Scaler 0-100	Default	0,92696	0,00136	0,92387	0,92941	0,92396	0,00346	0,91701	0,92899	0,89467	0,00202	0,89006	0,89797	0,96289(48)	0,00097	0,96133	0,96479	12,05541(3)	0,22274	11,7684	12,5566
	Optimal_01	0,9288	0,00108	0,92708	0,93055	0,94186	0,0027	0,93565	0,94579	0,89882	0,00159	0,89625	0,9015	0,9632(45)	0,00089	0,96173	0,96459	18,12889	0,59327	17,6156	20,3745
	Optimal_02	0,93317	0,00168	0,93012	0,93652	0,96306(13*)	0,00316	0,95824	0,96767	0,90633	0,0024	0,90192	0,91103	0,96969	0,00079	0,96829	0,97145	24,19112	0,36018	23,4526	25,0132
MinMax Scaler 0-1000	Default	0,9278	0,00086	0,92669	0,92941	0,92637	0,00261	0,92265	0,92899	0,89592	0,0012	0,89438	0,89797	0,96336	0,00061	0,96272	0,96466	11,95152(1)	0,15956	11,7684	12,2558
	Optimal_01	0,9295	0,00093	0,92785	0,93055	0,94393	0,00123	0,94192	0,94579	0,89996	0,00114	0,89793	0,9015	0,96371	0,00064	0,96276	0,96459	17,89339	0,29947	17,6156	18,4357
	Optimal_02	0,93439(8)	0,00135	0,93259	0,93652	0,96583(7*)	0,00123	0,96411	0,96767	0,90814(8)	0,00185	0,90567	0,91103	0,97022	0,00069	0,96938	0,97145	24,44122	0,25539	24,1252	25,0132

Περιγραφή Αποτελεσμάτων Τυχαίου Δάσους

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,93661** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 29,84745 seconds**. Τον χαμηλότερο μέσο όρο precision= **0,92253**, μικρότερο κατά **0,01408 (-1,50%)**, σε σχέση με το μοντέλο RF με τη μέγιστη precision, σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 16,67103 seconds**, που είναι χαμηλότερος κατά **13,17642 seconds (-44,14%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,9699** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 29,84745 seconds**. Τον χαμηλότερο μέσο όρο recall= **0,90832**, μικρότερο κατά **0,06158 (-6,35%)**, σε σχέση με το μοντέλο RF με τη μέγιστη recall, σημείωσε το μοντέλο RF που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 13,94055 seconds**, που είναι χαμηλότερος κατά **15,9069 seconds (+53,29%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Baθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο (mean) F1 score= **0,91127** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler, με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **29,84745** seconds. Τον χαμηλότερο μέσο όρο F1 score= **0,8873**, μικρότερο κατά **0,02397 (-2,63%)**, σε σχέση με το μοντέλο RF με το μέγιστο F1 Score, σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **16,67103** seconds, που είναι χαμηλότερος κατά **13,17642** seconds (**-44,15%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

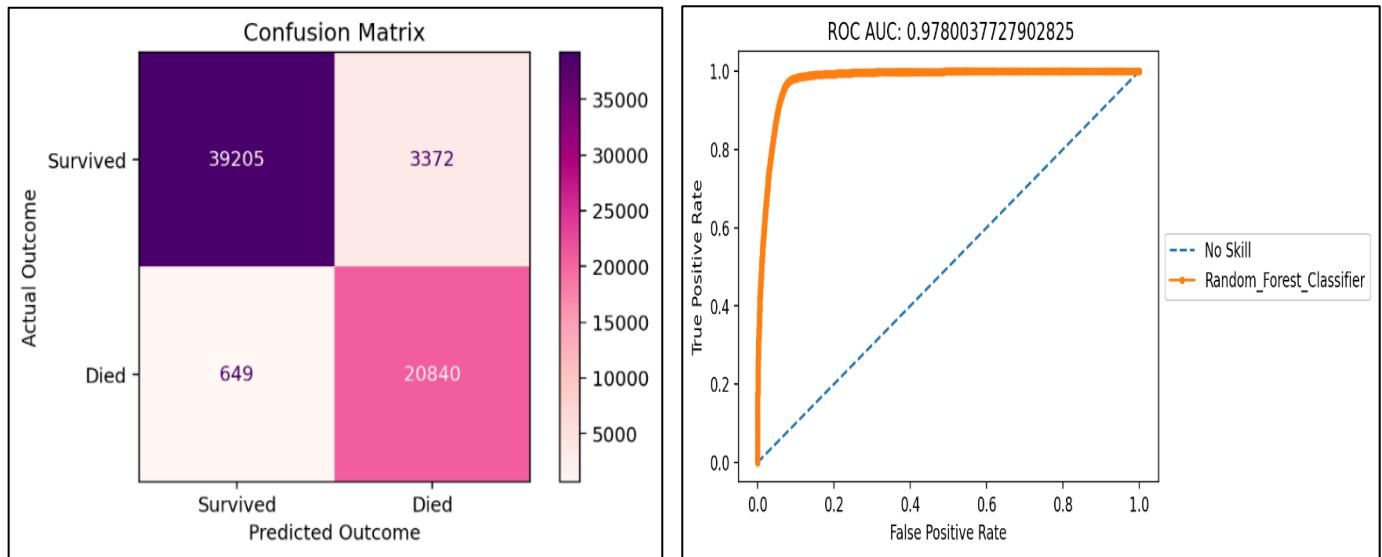
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,97708** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler, με εύρος **0-1000**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **29,84745** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,96155**, μικρότερο κατά **0,01553 (-2,63%)**, σε σχέση με το μοντέλο RF με τη μέγιστη τιμή AUC_ROC, σημείωσε το μοντέλο RF που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Standard** scaler, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **18,33744** seconds, που είναι χαμηλότερος κατά **11,51001** seconds (**-38,56%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

Τον χαμηλότερο μέσο όρο (mean) runtime= **11,95152** seconds σημείωσε το μοντέλο RF που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο precision= **0,9278**. Τον υψηλότερο μέσο όρο runtime= **30,22087** seconds, υψηλότερο κατά **18,26935** seconds (**+152,86%**), σε σχέση με το μοντέλο RF με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,93518**, υψηλότερο κατά **0,00738 (+0,80%)**, σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα RF μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

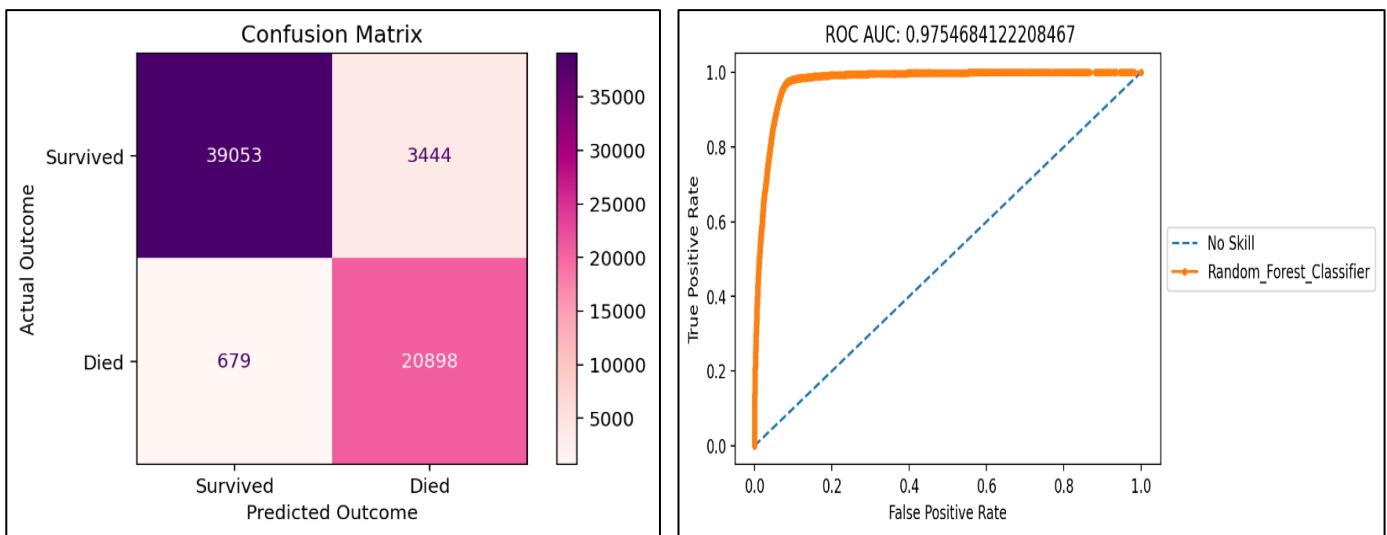
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_02** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93661**, Recall: **0,9699**, F1-score: **0,91127**, AUC_ROC: **0,97708**, Runtime: **29,84745** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 80**) και η καμπύλη ROC (**Εικόνα 81**) του RF μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 80: Ο Πίνακας Σύγχυσης του RF μοντέλου **22_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

Εικόνα 81: Η καμπύλη ROC του RF μοντέλου **22_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

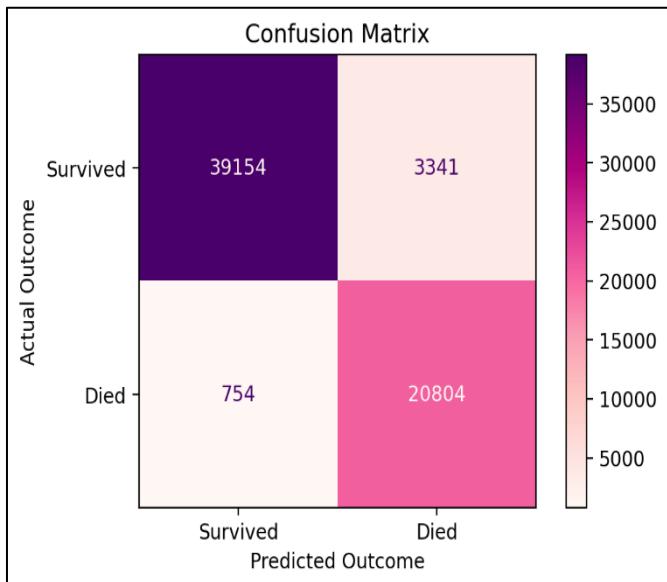
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **15** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_02** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93635**, Recall: **0,96672**, F1-score: **0,9108**, AUC_ROC: **0,97506**, Runtime: **26,96082** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 82**) και η καμπύλη ROC (**Εικόνα 83**) του RF μοντέλου με την μέγιστη AUC_ROC.



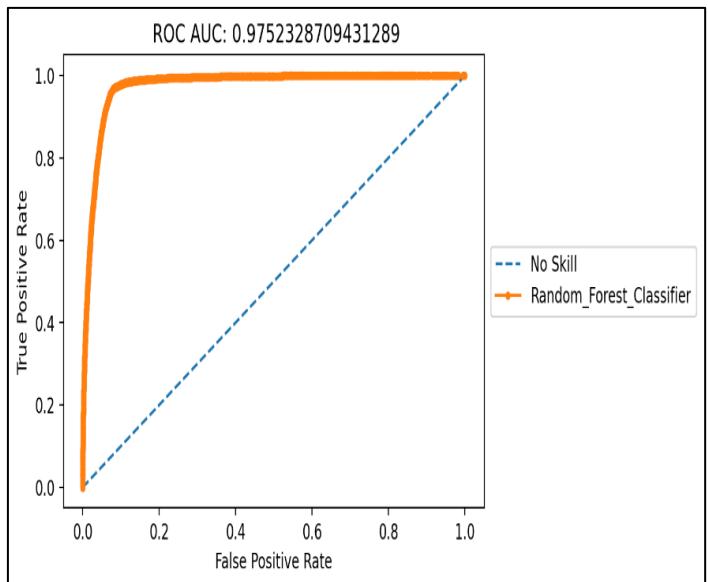
Εικόνα 82: Ο Πίνακας Σύγχυσης του RF μοντέλου **15_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

Εικόνα 83: Η καμπύλη ROC του RF μοντέλου **15_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **15** χαρακτηριστικά, χωρίς τη χρήση scaler(**no_scaler**) και με τις **optimal_02** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93551**, Recall: **0,96631**, F1-score: **0,90963**, AUC_ROC: **0,97441**, Runtime: **26,67779** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 91**) και η καμπύλη ROC (**Εικόνα 92**) του RF μοντέλου με την μέγιστη AUC_ROC. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 84**) και η καμπύλη ROC (**Εικόνα 85**) του RF μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 84: Ο Πίνακας Σύγχυσης του RF μοντέλου **15_none_opt_02** με την μέγιστη **AUC_ROC**



Εικόνα 85: Η καμπύλη ROC του RF μοντέλου **15_none_opt_02** με την μέγιστη **AUC_ROC**

Σύμφωνα με τα παραπάνω για το RF, οι βέλτιστοι τρόποι scaling είναι ο **Min-Max** scaler, για το εύρος **0-1000** και απουσία scaler (**no_scaler**), με τις **optimal_02** υπερπαραμέτρους.

Για την Ακραία Ενίσχυση Κλίσης (XGB) έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 39)

eXtreme Gradient Boosting			22 Features																		
Scaler Type	Parameter Type	Precision			Recall			F1-score				ROC AUC			Runtime(secs)						
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0.93696(9*)	0.00084	0.93583	0.93838	0.95404	0.00201	0.95002	0.95661	0.91037(9)(11*)	0.00118	0.90889	0.91228	0.97766(10)(10*)	0.00041	0.97704	0.97829	7,35696	0,47233	6,98276	8,23502
	Optimal_01	0.93736(3)(3*)	0.00065	0.93651	0.93847	0.9561(1)(20*)	0.00124	0.95416	0.95794	0.91118(2)(3*)	0.00093	0.91002	0.91254	0.97784(9)(9*)	0.00027	0.97732	0.97815	6,12796	0,69462	4,89179	7,3447
	Optimal_02	0.93347	0.00082	0.93202	0.93478	0.94115(49)	0.0017	0.93915	0.94486	0.90479	0.00119	0.90313	0.90674	0.9754(15*)	0.00065	0.97417	0.97613	16,99483(49)	1,44257	16,0646	20,8238
Standard Scaler	Default	0.93583(14*)	0.00084	0.93429	0.93688	0.95286	0.00139	0.94997	0.95481	0.90889(18*)	0.0012	0.90714	0.91041	0.97677(12*)	0.00045	0.97614	0.97736	7,84059	0,23756	7,65569	8,40704
	Optimal_01	0.93582(15*)	0.00079	0.93433	0.93681	0.95339	0.00127	0.95129	0.95631	0.90885(19*)	0.00095	0.90677	0.90992	0.97662(13*)	0.00049	0.97576	0.97742	6,68503	0,22121	6,4985	7,26612
	Optimal_02	0.93337	0.00112	0.93226	0.93547	0.94022(50)	0.00183	0.9381	0.94442	0.90442	0.00155	0.90288	0.90741	0.97513(18*)	0.00059	0.97448	0.97638	17,02819(50)	0,36113	16,6131	17,7593
MinMax Scaler 0-1	Default	0.93709(8)(8*)	0.00097	0.93468	0.93924	0.95419	0.00171	0.95047	0.95724	0.91058(8)(10*)	0.00133	0.90743	0.91368	0.97803(8)(8*)	0.00065	0.97671	0.97911	7,86454	0,34893	7,28328	8,86462
	Optimal_01	0.93723(5)(5*)	0.00091	0.93554	0.93939	0.95452(8)	0.00176	0.94993	0.9584	0.91078(5)(7*)	0.0012	0.90855	0.91377	0.97823(4)(4*)	0.00058	0.97731	0.97939	6,76577	0,28052	6,31202	7,49651
	Optimal_02	0.93302	0.00093	0.93125	0.93519	0.93972(52)	0.00191	0.93633	0.94393	0.90404(50)	0.0013	0.90179	0.90683	0.97508(19*)	0.00051	0.97428	0.97623	17,71116(54)	0,71532	16,2658	20,2194
MinMax Scaler 0-10	Default	0.93714(7)(7*)	0.00108	0.93468	0.93924	0.95447(9)	0.00167	0.95138	0.95724	0.91068(7)(9*)	0.00148	0.90743	0.91368	0.97809(6)(6*)	0.00071	0.97671	0.97911	7,7813	0,31739	7,28328	8,50047
	Optimal_01	0.93731(4)(4*)	0.00097	0.93568	0.93939	0.95453(7)	0.00186	0.94993	0.9584	0.91086(4)(5*)	0.00128	0.90855	0.91377	0.97833(3)(3*)	0.0006	0.97731	0.97876	6,74094	0,31123	6,31202	7,49651
	Optimal_02	0.93294	0.00089	0.93125	0.935	0.9397(53)	0.0019	0.93633	0.94393	0.90394(51)	0.00124	0.90179	0.90662	0.97507(20*)	0.00052	0.97428	0.97623	17,60409(53)	0,77033	16,2658	20,2194
MinMax Scaler 0-100	Default	0.93722(6)(6*)	0.00116	0.93468	0.93924	0.95417	0.00145	0.95165	0.95676	0.91076(6)(8*)	0.0016	0.90743	0.91368	0.97818(5)(5*)	0.00071	0.97671	0.97911	7,69996	0,32437	7,28328	8,45383
	Optimal_01	0.93764(1)(1*)	0.00088	0.93589	0.93939	0.95472(5)	0.00166	0.95198	0.9584	0.9113(1)(1*)	0.00116	0.9089	0.91377	0.97855(2)(2*)	0.00055	0.97757	0.97939	6,67306	0,30961	6,31202	7,47422
	Optimal_02	0.93282(47)	0.00079	0.93125	0.93448	0.93958(54)	0.00196	0.93633	0.94298	0.90377(53)	0.0011	0.90179	0.90616	0.97505(16*)	0.00053	0.97428	0.97623	17,39536(52)	0,83383	16,2658	20,2194
MinMax Scaler 0-1000	Default	0.93686(10)(10*)	0.00125	0.93468	0.93851	0.95365	0.00145	0.95165	0.95609	0.91022(10)(12*)	0.00166	0.90743	0.91272	0.97808(7)(7*)	0.00073	0.97671	0.97906	7,6795	0,3423	7,29255	8,45383
	Optimal_01	0.93754(2)(2*)	0.00094	0.93589	0.93939	0.95403	0.00179	0.95198	0.95715	0.9111(3)(4*)	0.0012	0.9089	0.91336	0.978849(1)(1*)	0.00047	0.97784	0.97939	6,63992	0,33544	6,31202	7,47422
	Optimal_02	0.93312	0.00081	0.93208	0.93448	0.94007(51)	0.00235	0.93633	0.94298	0.90402(48)	0.00114	0.90246	0.90616	0.97526	0.00049	0.97463	0.97623	17,38894(51)	0,55348	16,2658	20,0997

❖ Για τα 15 χαρακτηριστικά (Πίνακας 40)

eXtreme Gradient Boosting			15 Features																		
Scaler Type	Parameter Type	Precision			Recall			F1-score				ROC AUC			Runtime(secs)						
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0.93529	0.00076	0.93421	0.93669	0.9547(6)	0.00152	0.95249	0.95775	0.90835	0.00099	0.90697	0.91019	0.97443	0.00048	0.97362	0.97517	6,29781	0,24079	6,0575	6,94418
	Optimal_01	0.93508	0.00064	0.93423	0.93559	0.95389	0.00163	0.95163	0.9562	0.90798	0.00093	0.90668	0.90917	0.97452	0.0004	0.97394	0.97533	5,53247	0,07826	5,42123	5,66509
	Optimal_02	0.9321(54)	0.00094	0.93061	0.9333	0.94458(48)	0.00118	0.94318	0.94743	0.90332(54)	0.00127	0.90132	0.90484	0.97178(48)	0.00045	0.97098	0.97231	14,24933(47)	0,33066	13,9633	14,8427
Standard Scaler	Default	0.93396	0.00076	0.93295	0.93496	0.95218	0.00132	0.95041	0.95437	0.90641	0.00095	0.90512	0.90777	0.97374	0.00057	0.97264	0.97461	5,99228	0,09271	5,87312	6,17008
	Optimal_01	0.93383	0.00093	0.93256	0.93588	0.95142	0.00159	0.94874	0.95381	0.9062	0.00126	0.90465	0.90904	0.97347	0.00068	0.97258	0.97481	5,32465(7)	0,14494	5,19661	5,66058
	Optimal_02	0.93282(48)	0.00105	0.93141	0.93444	0.94462(47)	0.00239	0.94021	0.94859	0.90421(46)	0.00149	0.90192	0.90653	0.97204	0.00067	0.97073	0.97321	13,45364	0,10992	13,2853	13,6294
MinMax Scaler 0-1	Default	0.93468	0.00119	0.93193	0.93704	0.9529	0.00232	0.94929	0.95841	0.90735	0.00171	0.90307	0.91079	0.97405	0.00075	0.97239	0.97543	6,0945	0,35683	5,55081	7,34375
	Optimal_01	0.93504	0.00097	0.93235	0.93689	0.95287	0.00237	0.94559	0.95782	0.90786	0.00136	0.90441	0.91033	0.97417	0.00062	0.97247	0.97546	5,49868(8)	0,35653	5,19662	6,80531
	Optimal_02	0.93278(49)	0.00085	0.93069	0.9343	0.94516(46)	0.00193	0.9407	0.94841	0.90421(47)	0.00115	0.90096	0.90624	0.97195(46)	0.00064	0.97071	0.97322	13,83423	0,59031	12,9446	15,1558
MinMax Scaler 0-10	Default	0.93504	0.00108	0.93292	0.93704	0.95353	0.0022	0.94996	0.95841	0.90786	0.00157	0.90489	0.91079	0.97428	0.00064	0.97286	0.97543	6,1544	0,37883	5,64419	7,34375
	Optimal_01	0.93515	0.00103	0.93235	0.93689	0.95323	0.00224	0.94841	0.95782	0.90802	0.00142	0.90441	0.91033	0.97428	0.00066	0.97247	0.97546	5,52203	0,3801	5,19662	6,80531
	Optimal_02	0.93292	0.00088	0.9313	0.9343	0.94517(45)	0.00179	0.94081	0.94805	0.90439	0.00105	0.90234	0.90624	0.97199(45)	0.00064	0.97074	0.97322	14,02718(45)	0,54368	13,1855	15,1558
MinMax Scaler 0-100	Default	0.93532(20*)	0.00103	0.93339	0.93704	0.95384	0.00214	0.94996	0.95841	0.90826	0.00154	0.9054	0.91079	0.97455	0.00057	0.97339	0.97543	6,23783	0,3711	5,87856	7,34375
	Optimal_01	0.93554(16*)	0.00086	0.93351	0.93689	0.95415	0.00199	0.94967	0.95782	0.9086(20*)	0.00114	0.90575	0.91033	0.97452	0.00067	0.97247	0.97546	5,51039(9)	0,26033	5,19662	6,08652
	Optimal_02	0.93291(45)	0.00088	0.9313	0.9343	0.9452	0.00197	0.94081	0.												

❖ Για τα 10 χαρακτηριστικά (Πίνακας 41)

Extreme Gradient Boosting			10 Features																		
Scaler Type	Parameter Type	Precision				Recall				F1-score				ROC AUC				Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0.93452	0.00084	0.93319	0.93582	0.95479(4)	0.00241	0.95147	0.95891	0.90733	0.00118	0.9056	0.90939	0.973	0.00061	0.97207	0.97369	5,57817	0,1383	5,36429	5,79774
	Optimal_01	0.9345	0.00037	0.9337	0.93482	0.9544	0.00223	0.95006	0.95713	0.90725	0.00053	0.90611	0.90784	0.97322	0.0005	0.97255	0.97403	5,00691(3)	0,17371	4,78493	5,45962
	Optimal_02	0.93265(50)	0.00117	0.92946	0.93378	0.9497	0.00174	0.94556	0.95184	0.90445	0.00172	0.89981	0.90611	0.97106(49)	0.00052	0.9699	0.9718	12,31829	0,44696	11,8745	13,4047
Standard Scaler	Default	0.93355	0.00123	0.93184	0.93592	0.95222	0.00207	0.94794	0.95543	0.90587	0.00163	0.90392	0.90916	0.97229	0.00089	0.97099	0.97396	5,68466	0,28835	5,37804	6,37479
	Optimal_01	0.93318	0.00087	0.93205	0.93504	0.9508	0.00179	0.94876	0.95373	0.90522	0.00116	0.90354	0.90784	0.97255	0.00052	0.97172	0.9731	4,96984(1)	0,16883	4,74405	5,2898
	Optimal_02	0.93224(53)	0.00052	0.93134	0.93283	0.94786	0.00162	0.94487	0.95009	0.90378(52)	0.00071	0.90271	0.90476	0.97096(50)	0.00031	0.97041	0.97143	12,30304	0,4726	11,672	13,1635
MinMax Scaler 0-1	Default	0.93412	0.00108	0.9316	0.93601	0.95311	0.0022	0.94859	0.95787	0.90666	0.00152	0.90319	0.90946	0.97272	0.00065	0.97135	0.97452	5,62511	0,25232	5,32906	6,45157
	Optimal_01	0.93384	0.00131	0.92986	0.93622	0.9533	0.00219	0.94759	0.95769	0.90633	0.00185	0.90047	0.90968	0.97268	0.00075	0.97113	0.97443	5,0141(5)	0,19234	4,7478	5,61718
	Optimal_02	0.93254(52)	0.00123	0.92984	0.93497	0.94826	0.00216	0.94315	0.95259	0.90418(49)	0.00173	0.90035	0.90757	0.97074(53)	0.0007	0.96937	0.97252	12,10837	0,41026	11,4982	13,0472
MinMax Scaler 0-10	Default	0.93423	0.00108	0.93216	0.93601	0.95343	0.00225	0.94859	0.95787	0.90682	0.00154	0.90425	0.90946	0.97277	0.00069	0.97135	0.97452	5,59054	0,22371	5,3603	6,45157
	Optimal_01	0.93409	0.00108	0.9322	0.93622	0.9539	0.00192	0.95013	0.95769	0.90674	0.00148	0.90422	0.90968	0.97281	0.00072	0.9715	0.97443	5,01337(4)	0,19413	4,7478	5,61718
	Optimal_02	0.93258(51)	0.00122	0.92984	0.93497	0.94843	0.0022	0.94337	0.95259	0.90424(45)	0.00168	0.90035	0.90757	0.97073(54)	0.00072	0.96937	0.97252	12,17782	0,42524	11,5451	13,0472
MinMax Scaler 0-100	Default	0.93455	0.00102	0.9328	0.93601	0.95442(2)	0.00177	0.95122	0.95787	0.90731	0.00144	0.90499	0.90946	0.97297	0.00068	0.97151	0.97452	5,58199	0,18182	5,3603	6,09413
	Optimal_01	0.93446	0.00103	0.9324	0.93622	0.95445(10)	0.00175	0.95013	0.95769	0.90726	0.00141	0.90439	0.90968	0.97304	0.00071	0.97205	0.97443	5,03736(6)	0,21204	4,76558	5,61718
	Optimal_02	0.93299(46)	0.00097	0.93093	0.93497	0.94863	0.00178	0.94523	0.95208	0.90463	0.00134	0.90176	0.90737	0.97084(52)	0.00067	0.96937	0.97252	12,17046	0,41397	11,5451	12,906
MinMax Scaler 0-1000	Default	0.93521	0.00064	0.93425	0.93601	0.95524	0.00147	0.9526	0.95787	0.90828	0.00094	0.90689	0.90946	0.97344	0.00048	0.97282	0.97452	5,59556	0,19612	5,43843	6,09413
	Optimal_01	0.93508	0.00064	0.93379	0.93622	0.95517(3)	0.0015	0.95243	0.95769	0.90809	0.00096	0.90599	0.90968	0.97353	0.00062	0.97253	0.97443	4,98007(2)	0,10178	4,79062	5,11086
	Optimal_02	0.9331	0.00072	0.93255	0.93497	0.94912	0.00142	0.9466	0.95139	0.90495	0.00101	0.90397	0.90737	0.97087(51)	0.00045	0.97002	0.97158	12,2197	0,44242	11,7262	12,8595

Περιγραφή Αποτελεσμάτων της Ακραίας Ενίσχυσης Κλίσης

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,93764** σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 6,67306 seconds**. Τον χαμηλότερο μέσο όρο precision= **0,9321**, μικρότερο κατά **0,00554 (-0,59%)**, σε σχέση με το μοντέλο XGB με τη μέγιστη precision, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **15** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 14,24933 seconds**, που είναι υψηλότερος κατά **7,57627 seconds (+113,54%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,9561** σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 6,12796 seconds**. Τον χαμηλότερο μέσο όρο recall= **0,93958**, μικρότερο κατά **0,01652 (-1,73%)**, σε σχέση με το μοντέλο XGB με τη μέγιστη recall, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης **τα 17,39536 seconds**, που είναι υψηλότερος κατά **11,2674 seconds (+183,87%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Βαθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο (mean) F1 score= **0,9113** σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,67306** seconds. Τον χαμηλότερο μέσο όρο F1 score= **0,90332**, μικρότερο κατά **0,00798** (-**0,88%**), σε σχέση με το μοντέλο XGB με το μέγιστο F1 Score, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **15** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **14,24933** seconds, που είναι υψηλότερος κατά **7,57627** seconds (+**113,54%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

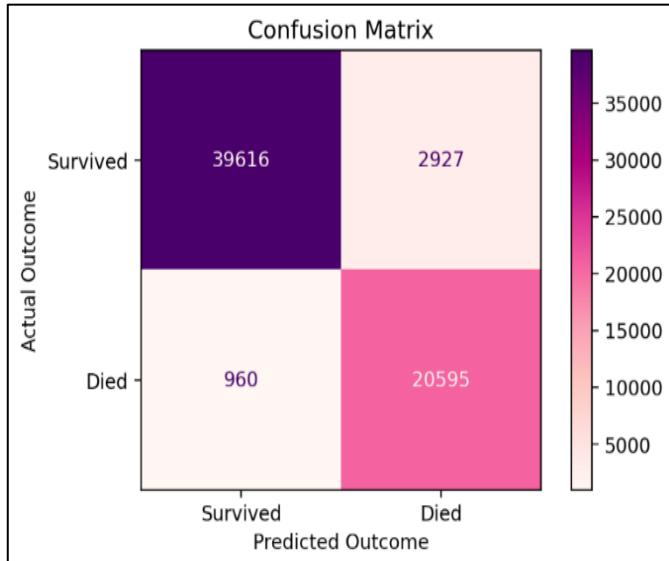
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,978849** σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,63992** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,97073**, ελάχιστα μικρότερο κατά **0,008119** (-**0,83%**), σε σχέση με το μοντέλο XGB με τη μέγιστη τιμή AUC_ROC, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-10**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **12,17782** seconds, που είναι υψηλότερος κατά **5,5379** seconds (-**83,40%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

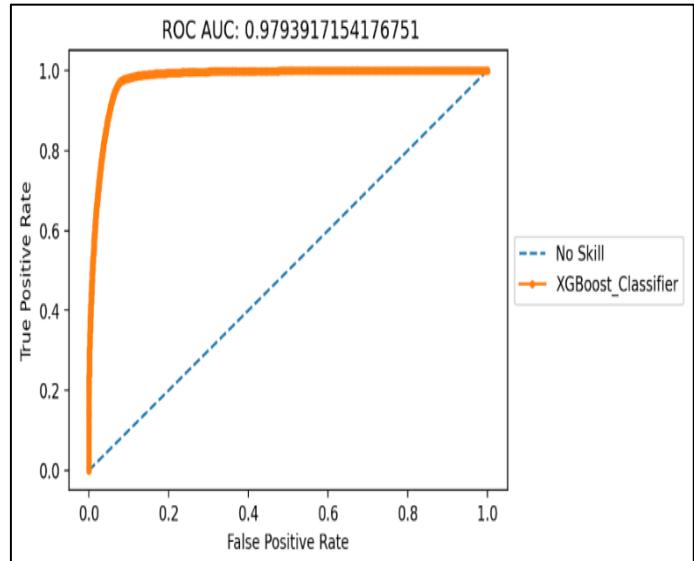
Τον χαμηλότερο μέσο όρο (mean) runtime= **4,96984** seconds σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Standard** scaler, με τις **optimal_01** τιμές των υπερπαραμέτρων και με μέσο όρο precision= **0,93318**. Τον υψηλότερο μέσο όρο runtime= **17,71116** seconds, υψηλότερο κατά **12,74132** seconds (+**256,37%**), σε σχέση με το μοντέλο XGB με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1**, με τις **optimal_02** τιμές των υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,93302**, ελάχιστα χαμηλότερο κατά **0,00016** (+**0,017%**), σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα **XGB** μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-100** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93764**, Recall: **0,95472**, F1-score: **0,9113**, AUC_ROC: **0,97855**, Runtime: **6,67306** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 86**) και η καμπύλη ROC (**Εικόνα 87**) του XGB μοντέλου με την μέγιστη AUC_ROC.

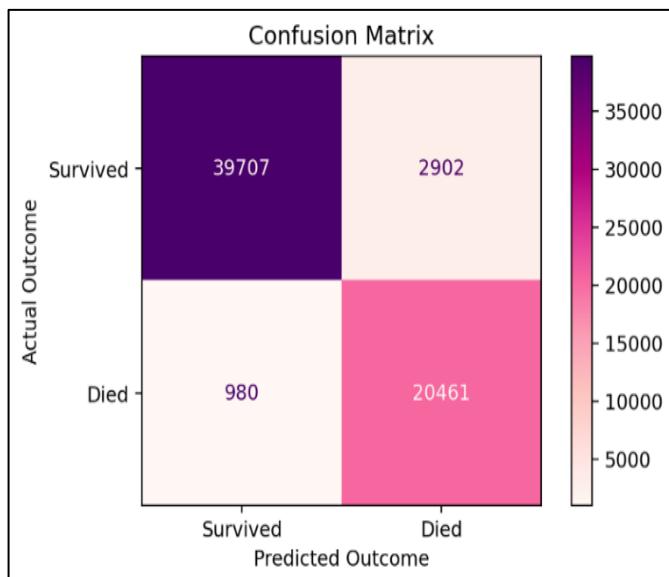


Εικόνα 86: Ο Πίνακας Σύγχυσης του **XGB** μοντέλου **22_Min-Max 0-100_opt_01** με την μέγιστη **AUC_ROC**

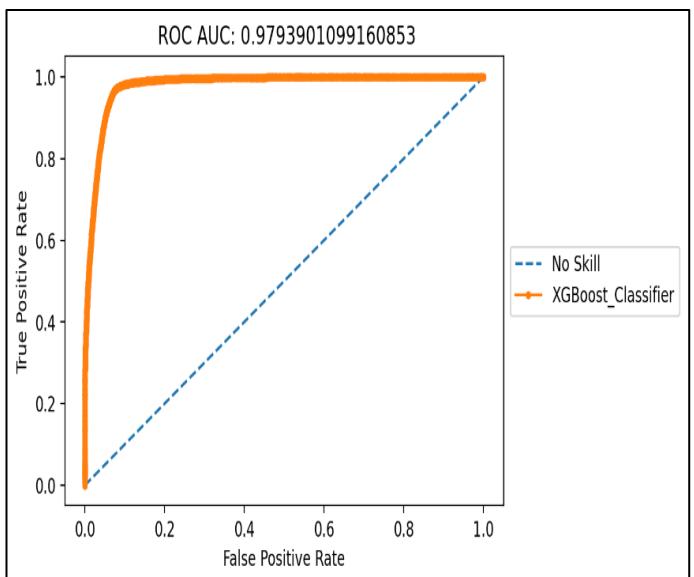


Εικόνα 87: Η καμπύλη ROC του **XGB** μοντέλου **22_Min-Max 0-100_opt_01** με την μέγιστη **AUC_ROC**

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93754**, Recall: **0,95403**, F1-score: **0,9111**, AUC_ROC: **0,978849**, Runtime: **6,63992** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 88**) και η καμπύλη ROC (**Εικόνα 89**) του XGB μοντέλου με την μέγιστη AUC_ROC.

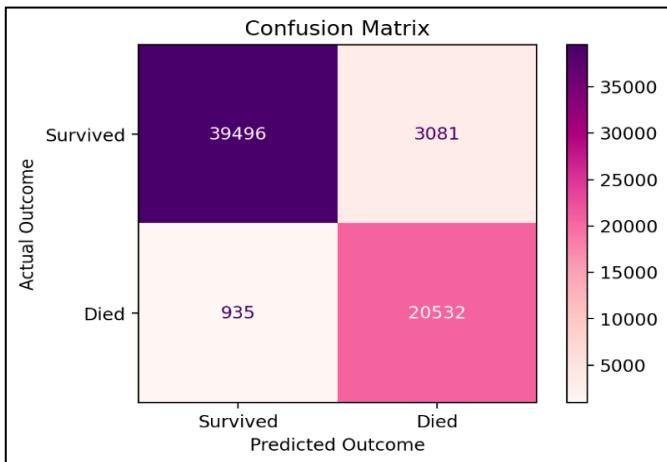


Εικόνα 88: Ο Πίνακας Σύγχυσης του **XGB** μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

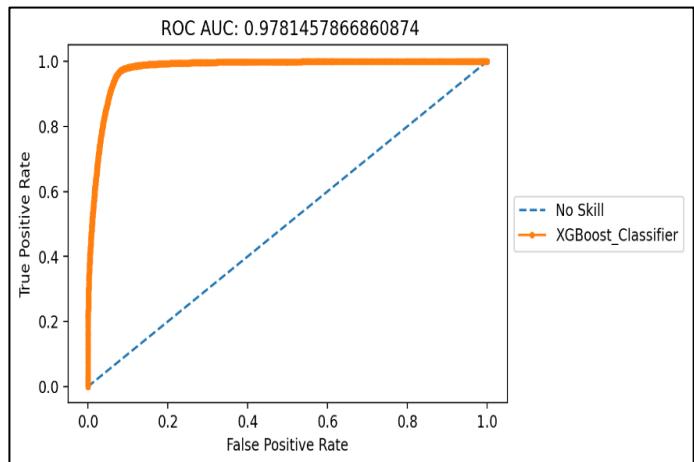


Εικόνα 89: Η καμπύλη ROC του **XGB** μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, χωρίς τη χρήση scaler(**no_scaler**) και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93736**, Recall: **0,9561**, F1-score: **0,91116**, AUC_ROC: **0,97784**, Runtime: **6,12796** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 90**) και η καμπύλη ROC (**Εικόνα 91**) του XGB μοντέλου με την μέγιστη AUC_ROC.

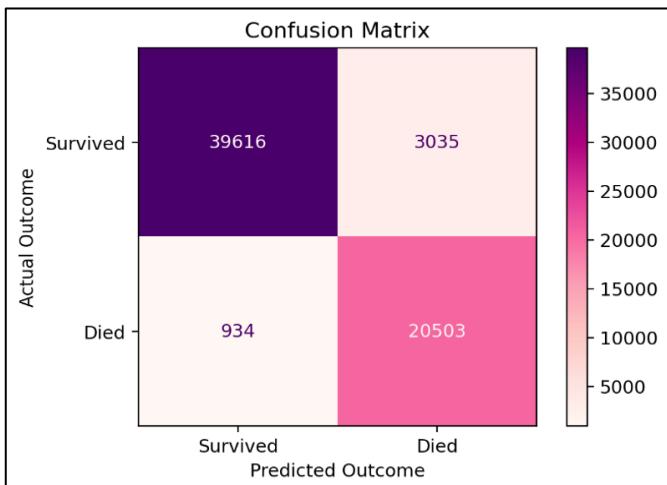


Εικόνα 90: Ο Πίνακας Σύγχυσης του XGB μοντέλου **22_No_Scaler_opt_01** με την μέγιστη **AUC_ROC**

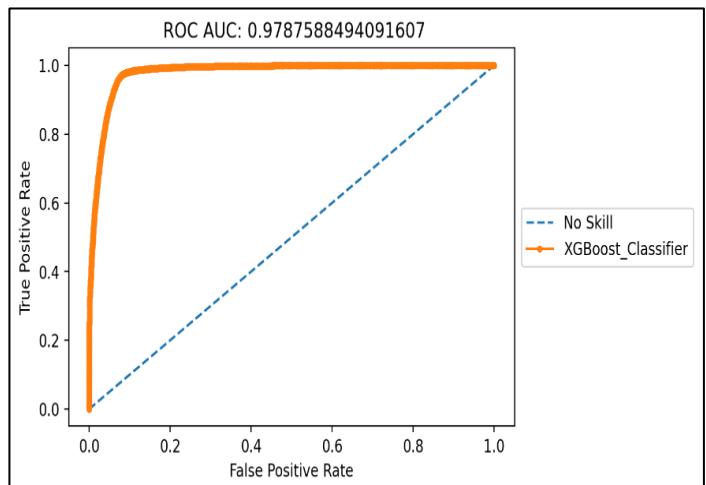


Εικόνα 91: Η καμπύλη ROC του XGB μοντέλου **22_No_Scaler_opt_01** με την μέγιστη **AUC_ROC**

4. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-10** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93731**, Recall: **0,95453**, F1-score: **0,91086**, AUC_ROC: **0,97833**, Runtime: **6,74094** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 92**) και η καμπύλη ROC (**Εικόνα 93**) του XGB μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 92: Ο Πίνακας Σύγχυσης του XGB μοντέλου **22_Min-Max_0-10_opt_01** με την μέγιστη **AUC_ROC**



Εικόνα 93: Η καμπύλη ROC του XGB μοντέλου **22_Min-Max_0-10_opt_01** με την μέγιστη **AUC_ROC**

Σύμφωνα με τα παραπάνω για την XGB, οι βέλτιστοι τρόποι scaling είναι ο **Min-Max** scaler, για τα εύρη **0-10**, **0-100** και **0-1000** και απουσία scaler (**no_scaler**), με τις **optimal_01** υπερπαραμέτρους.

Για τα Τεχνητά Νευρωνικά Δίκτυα Πολυστρωματικών Αντιλήπτρων (MLPs) έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 42)

Multi-Layer Perceptrons		22 Features															
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,93339	0,00155	0,93066	0,936	0,94976	0,00588	0,93644	0,95892	0,90548	0,00213	0,90239	0,90931	0,9722	0,00078	0,97139	0,97351
	Optimal_01	0,93377(7)	0,00119	0,9321	0,93543	0,95034	0,00777	0,93629	0,96101	0,90592	0,00175	0,90337	0,90772	0,97281(7)	0,00087	0,97151	0,97431
	Optimal_02	0,93376(8)	0,00129	0,93181	0,93568	0,94988	0,00742	0,93727	0,96036	0,9058	0,00213	0,90285	0,90924	0,97334(2)	0,00071	0,97211	0,97402
Standard Scaler	Default	0,93222	0,00079	0,93129	0,93388	0,95011	0,00417	0,94275	0,95692	0,90393	0,00122	0,90252	0,90659	0,97156	0,0005	0,97075	0,97245
	Optimal_01	0,93234	0,00124	0,93015	0,93507	0,94931(45)	0,00425	0,94019	0,95386	0,90406	0,00186	0,90028	0,90773	0,97064(54)	0,00074	0,96973	0,97218
	Optimal_02	0,93272	0,0008	0,93184	0,93474	0,95109	0,00351	0,94682	0,95761	0,90473	0,00119	0,90358	0,90778	0,97146	0,00048	0,97083	0,97253
MinMax Scaler 0-1	Default	0,9333	0,00087	0,93157	0,93508	0,95214	0,00597	0,93866	0,96635	0,90554	0,00128	0,90308	0,90848	0,97243	0,00071	0,97098	0,97423
	Optimal_01	0,9336	0,00136	0,93047	0,93699	0,95253	0,00547	0,93693	0,96311	0,90594(9)	0,002	0,90094	0,91062	0,9722	0,00084	0,97075	0,97392
	Optimal_02	0,93352	0,00118	0,93073	0,9357	0,95217	0,00583	0,93831	0,96408	0,90574	0,00169	0,90084	0,90914	0,97232	0,00075	0,97074	0,97412
MinMax Scaler 0-10	Default	0,9335	0,00088	0,93157	0,93508	0,95217	0,0066	0,93866	0,96635	0,90579	0,00137	0,90308	0,90848	0,97262	0,00066	0,97125	0,97423
	Optimal_01	0,9337(10)	0,00134	0,93103	0,93699	0,9531(8)	0,0058	0,93693	0,96311	0,90614(7)	0,00199	0,90135	0,91062	0,97231	0,00083	0,97089	0,97392
	Optimal_02	0,93381(6)	0,00109	0,93209	0,9357	0,9517	0,00558	0,93951	0,96408	0,90607(8)	0,00153	0,90331	0,90914	0,97247	0,00072	0,97134	0,97412
MinMax Scaler 0-100	Default	0,9335	0,0009	0,93198	0,93508	0,95392(4)	0,00657	0,93866	0,96635	0,90593(10)	0,00154	0,90308	0,90848	0,97267(9)	0,00068	0,97125	0,97423
	Optimal_01	0,93401(3)	0,00143	0,93103	0,93699	0,95406(3)	0,00608	0,93693	0,96311	0,90656(3)	0,00205	0,90135	0,91062	0,97262	0,00078	0,97089	0,97295
	Optimal_02	0,93395(4)	0,00103	0,93214	0,9357	0,95118	0,00637	0,93951	0,96408	0,90622(5)	0,00152	0,90331	0,90914	0,97268(8)	0,00064	0,97174	0,97412
MinMax Scaler 0-1000	Default	0,93375(9)	0,00105	0,93198	0,93508	0,95322(5)	0,00918	0,93866	0,96635	0,90627(4)	0,00204	0,90308	0,90848	0,97305(4)	0,00055	0,97203	0,97423
	Optimal_01	0,93456(1)	0,00147	0,93216	0,93699	0,95623(1)(19*)	0,00573	0,94464	0,96311	0,90756(1)	0,00184	0,90473	0,91062	0,97294(5)	0,00088	0,97089	0,97392
	Optimal_02	0,93442(2)	0,00104	0,93264	0,9357	0,95279	0,00713	0,94014	0,96408	0,90667(2)	0,00139	0,90456	0,90914	0,97292(6)	0,00065	0,97219	0,97412

❖ Για τα 15 χαρακτηριστικά (Πίνακας 43)

Multi-Layer Perceptrons		15 Features															
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,93285	0,00137	0,92998	0,93455	0,95281	0,00332	0,94352	0,95476	0,90496	0,00178	0,90119	0,90704	0,97213	0,00082	0,97075	0,97331
	Optimal_01	0,93338	0,00107	0,93189	0,93584	0,95078	0,00795	0,93666	0,95938	0,90559	0,00171	0,90245	0,90875	0,97309(3)	0,00083	0,97164	0,97431
	Optimal_02	0,93394(5)	0,00064	0,93268	0,93514	0,95031	0,00546	0,94206	0,95761	0,90618(6)	0,0012	0,90388	0,90802	0,97352(1)	0,00043	0,97301	0,9743
Standard Scaler	Default	0,93311	0,00065	0,93214	0,93422	0,94923(47)	0,00287	0,94412	0,95268	0,90496	0,00091	0,90326	0,90631	0,9723	0,00043	0,97162	0,97286
	Optimal_01	0,93258	0,0013	0,9299	0,93407	0,95302(10)	0,00372	0,94587	0,9583	0,90471	0,00174	0,90107	0,90641	0,97131(47)	0,0006	0,97063	0,9721
	Optimal_02	0,9332	0,00103	0,93178	0,93483	0,95139	0,00367	0,94551	0,9553	0,90535	0,00152	0,90328	0,90785	0,97219	0,00069	0,97118	0,97318
MinMax Scaler 0-1	Default	0,93261	0,00114	0,92904	0,9354	0,94971	0,00734	0,92006	0,96253	0,90444	0,00188	0,89689	0,9077	0,97223	0,00055	0,97112	0,97349
	Optimal_01	0,93291	0,00105	0,93115	0,93528	0,95071	0,00589	0,94137	0,96585	0,90492	0,00154	0,90252	0,90822	0,97219	0,00062	0,97094	0,97346
	Optimal_02	0,93278	0,00107	0,93053	0,93502	0,95135	0,00477	0,93991	0,9603	0,9048	0,00151	0,90158	0,90782	0,97216	0,00061	0,97095	0,9735
MinMax Scaler 0-10	Default	0,93266	0,00117	0,92904	0,9354	0,94927(46)	0,00813	0,92006	0,96253	0,90445	0,00199	0,89689	0,9077	0,97234	0,00051	0,97113	0,97349
	Optimal_01	0,93278	0,00093	0,93115	0,93474	0,95135	0,00593	0,94164	0,96585	0,90483	0,00147	0,90252	0,90822	0,97219	0,00057	0,97113	0,97317
	Optimal_02	0,93287	0,00116	0,93053	0,93502	0,95116	0,00515	0,93991	0,9603	0,90491	0,00166	0,90158	0,90782	0,97217	0,00058	0,9712	0,9735
MinMax Scaler 0-100	Default	0,93264	0,00135	0,92904	0,9354	0,94859(49)	0,00977	0,92006	0,96253	0,90433	0,00231	0,89689	0,9077	0,97241	0,00057	0,97113	0,97349
	Optimal_01	0,93302	0,00103	0,93115	0,93474	0,95169	0,00621	0,94164	0,96585	0,90513	0,00158	0,90262	0,90822	0,97224	0,0006	0,97113	0,97317
	Optimal_02	0,93307	0,00128	0,93053	0,93502	0,95196	0,00496	0,93991	0,9603	0,90526	0,00174	0,90158	0,90782	0,97224	0,00064	0,97127	0,9735
MinMax Scaler 0-1000	Default	0,93258	0,00176	0,92904	0,9354	0,94626(51)	0,0133	0,92006	0,96253	0,90404	0,00307	0,89689	0,9077	0,97258	0,00055	0,97162	0,97349
	Optimal_01	0,93342	0,00087	0,932	0,93474	0,95225	0,0074	0,94243	0,96585	0,9057	0,00151	0,90334	0,90822	0,97265(10)	0,00041	0,97201	0,97317
	Optimal_02	0,93335	0,00116	0,93118	0,93502	0,95214	0,00533	0,93991	0,95988	0,90566	0,00149	0,90295	0,90747	0,97232	0,00078	0,97141	0,9735

❖ Για τα 10 χαρακτηριστικά (Πίνακας 44)

Multi-Layer Perceptrons		10 Features															
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,93182(49)	0,00076	0,93045	0,93275	0,95275	0,00414	0,94666	0,95958	0,90391(45)	0,00133	0,90145	0,90543	0,97123(49)	0,00037	0,97068	0,97204
	Optimal_01	0,93202	0,00095	0,93055	0,93358	0,95319(7)	0,00478	0,94255	0,96068	0,90394	0,00153	0,90193	0,9067	0,97114(52)	0,00057	0,97055	0,97208
	Optimal_02	0,93241	0,00125	0,93023	0,93463	0,95171	0,00414	0,94498	0,95695	0,90441	0,00188	0,90125	0,90776	0,97153	0,00051	0,97091	0,9723
Standard Scaler	Default	0,93179(50)	0,0011	0,92951	0,93331	0,95097	0,00271	0,94628	0,95332	0,90358(50)	0,00158	0,90016	0,90542	0,97113(53)	0,00049	0,97052	0,97203
	Optimal_01	0,93273	0,0009	0,93144	0,93413	0,95201	0,00421	0,9449	0,95909	0,90476	0,00145	0,90297	0,90673	0,97172	0,00063	0,97053	0,97281
	Optimal_02	0,93286	0,001	0,93075	0,93396	0,9532(6)	0,00195	0,95038	0,95608	0,90512	0,00122	0,90244	0,90626	0,97165	0,00085	0,97024	0,97267
MinMax Scaler 0-1	Default	0,93038(51)	0,00532	0,90315	0,93351	0,94662(50)	0,02834	0,8044	0,97266	0,9011(51)	0,01001	0,84803	0,90632	0,97121(50)	0,0007	0,96973	0,97266
	Optimal_01	0,93198(46)	0,00122	0,92851	0,93469	0,9514	0,0069	0,92331	0,96003	0,90379(46)	0,00201	0,89677	0,90728	0,97135(46)	0,00083	0,96956	0,97253
	Optimal_02	0,93245	0,0009	0,93053	0,93491	0,95264	0,00408	0,9442	0,96515	0,90451	0,00133	0,90186	0,90842	0,9716	0,00054	0,97071	0,97312
MinMax Scaler 0-10	Default	0,92988(52)	0,00605	0,90315	0,93351	0,94532(52)	0,0327	0,8044	0,97266	0,90003(52)	0,01146	0,84803	0,90632	0,97118(51)	0,00071	0,96973	0,97266
	Optimal_01	0,93197(47)	0,00133	0,92851	0,93469	0,9509	0,00783	0,92331	0,96003	0,90374(47)	0,00223	0,89677	0,90728	0,97139	0,00083	0,96956	0,97253
	Optimal_02	0,93263	0,00089	0,93122	0,93491	0,95303(9)	0,00443	0,9442	0,96515	0,90482	0,0013	0,90305	0,90842	0,97167	0,00057	0,97071	0,97312
MinMax Scaler 0-100	Default	0,92889(53)	0,00727	0,90315	0,93351	0,94292(53)	0,04013	0,8044	0,97266	0,89881(53)	0,01389	0,84803	0,90632	0,9713(48)	0,0007	0,96973	0,97266
	Optimal_01	0,93191(48)	0,00141	0,92851	0,93451	0,95008	0,00942	0,92331	0,96003	0,90361(48)	0,00254	0,89677	0,90728	0,97136(45)	0,00087	0,96956	0,97253
	Optimal_02	0,93283	0,00089	0,93153	0,93491	0,95283	0,00524	0,9442	0,96515	0,90503	0,00139	0,90329	0,90842	0,97174	0,00056	0,97094	0,97312
MinMax Scaler 0-1000	Default	0,92598(54)	0,00946	0,90315	0,93218	0,93434(54)	0,05648	0,8044	0,97266	0,89386(54)	0,01865	0,84803	0,90446	0,97146	0,00063	0,97042	0,97238
	Optimal_01	0,93201(45)	0,00182	0,92851	0,93451	0,94873(48)	0,01181	0,92331	0,96003	0,9036(49)	0,00326	0,89677	0,90728	0,97167	0,00082	0,96983	0,97253
	Optimal_02	0,93309	0,0011	0,93166	0,93491	0,95429(2)	0,00633	0,94687	0,96515	0,90549	0,00171	0,90353	0,90842	0,97184	0,00073	0,97094	0,97312

Περιγραφή Αποτελεσμάτων των Τεχνητών Νευρωνικών Δικτυών Πολυστρωματικών Αντιλήπτρων

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,93456** σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **133,76172** seconds. Τον χαμηλότερο μέσο όρο precision= **0,92598**, μικρότερο κατά **0,00858** (-**0,92%**), σε σχέση με το μοντέλο MLPs με τη μέγιστη precision, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **28,3394** seconds, που είναι χαμηλότερος κατά **105,42232** seconds (**-78,81%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,95623** σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **133,76172** seconds. Τον χαμηλότερο μέσο όρο recall= **0,93434**, μικρότερο κατά **0,02189** (-**2,29%**), σε σχέση με το μοντέλο MLPs με τη μέγιστη recall, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **28,3394**

seconds, που είναι χαμηλότερος κατά **105,42232** seconds (**-78,81%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Bαθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο(mean) F1 score= **0,90756** σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **133,76172** seconds. Τον χαμηλότερο μέσο όρο F1 score= **0,89386**, μικρότερο κατά **0,0137** (**-1,51%**), σε σχέση με το μοντέλο MLPs με το μέγιστο F1 Score, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **28,3394** seconds, που είναι χαμηλότερος κατά **105,42232** seconds (**-78,81%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

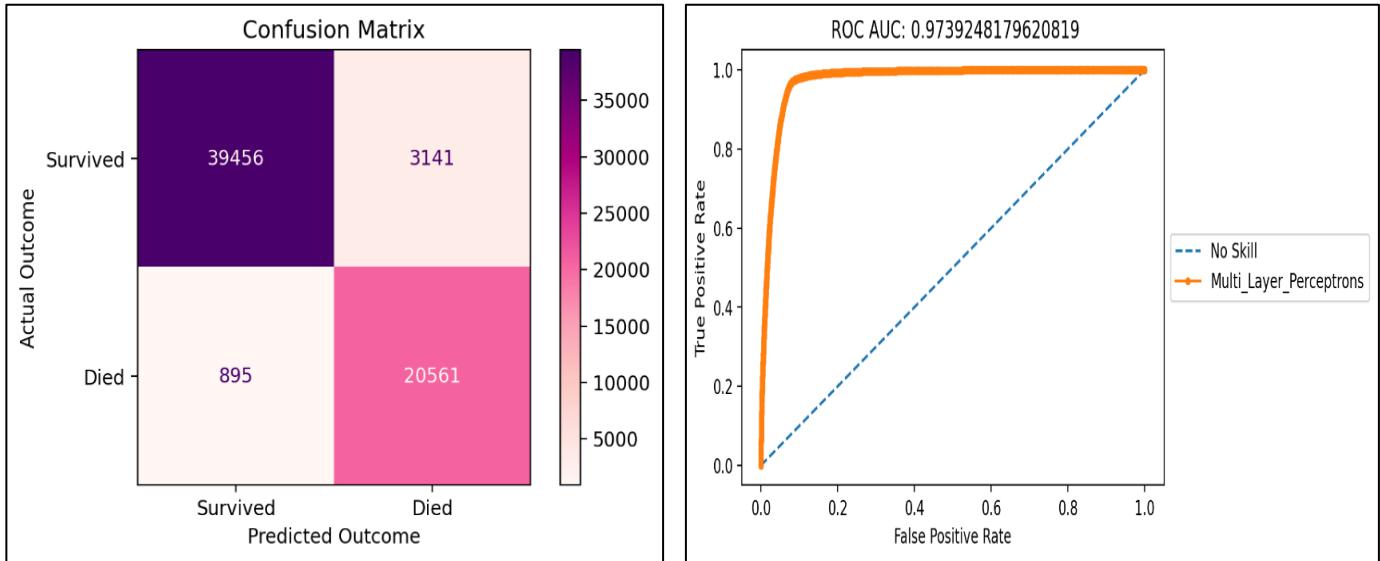
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,97352** σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **15** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **267,55604** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,97064**, ελάχιστα μικρότερο κατά **0,00288** (**-0,296%**), σε σχέση με το μοντέλο MLPs με τη μέγιστη τιμή AUC_ROC, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **199,77529** seconds, που είναι χαμηλότερος κατά **67,78075** seconds (**-78,81%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

Τον χαμηλότερο μέσο όρο (mean) runtime= **18,14386** seconds σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,93179**. Τον υψηλότερο μέσο όρο runtime= **362,46571** seconds, πολύ υψηλότερο κατά **344,32185** seconds (**+1.897,73%**), σε σχέση με το μοντέλο MLPs με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,93377**, ελάχιστα υψηλότερο κατά **0,00198** (**+0,21%**), σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα MLPs μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

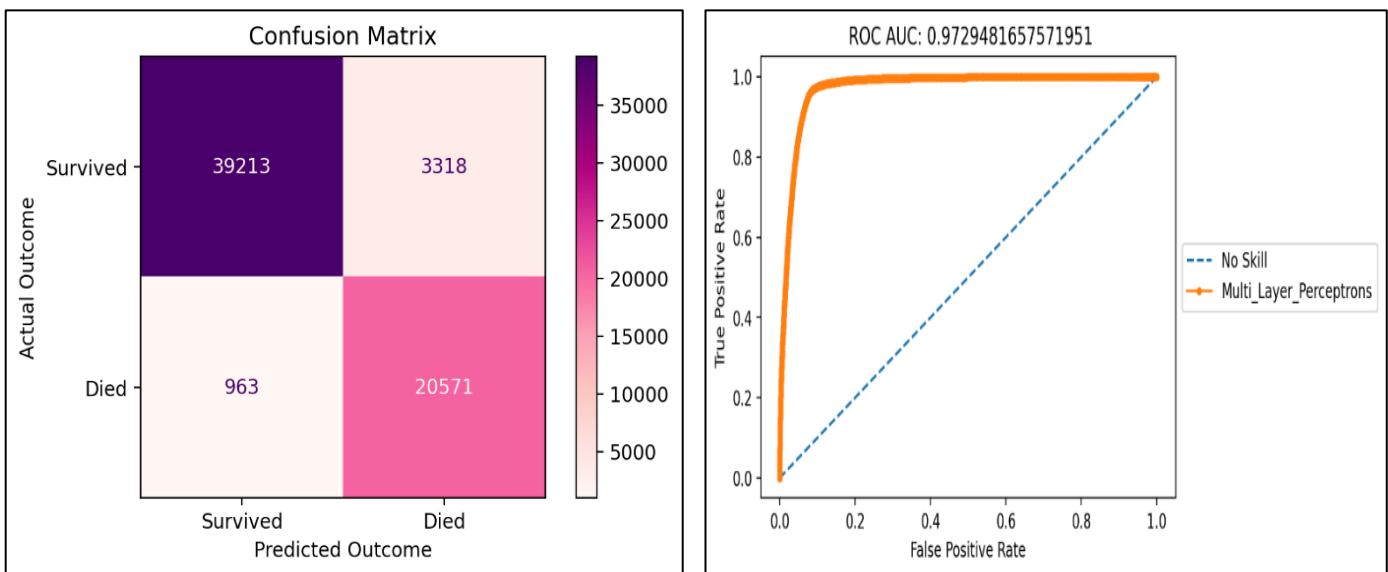
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93456**, Recall: **0,95623**, F1-score: **0,90756**, AUC_ROC: **0,97294**, Runtime: **133,76172** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 94**) και η καμπύλη ROC (**Εικόνα 95**) του MLP μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 94: Ο Πίνακας Σύγχυσης του MLPs μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

Εικόνα 95: Η καμπύλη ROC του MLPs μοντέλου **22_Min-Max 0-1000_opt_01** με την μέγιστη **AUC_ROC**

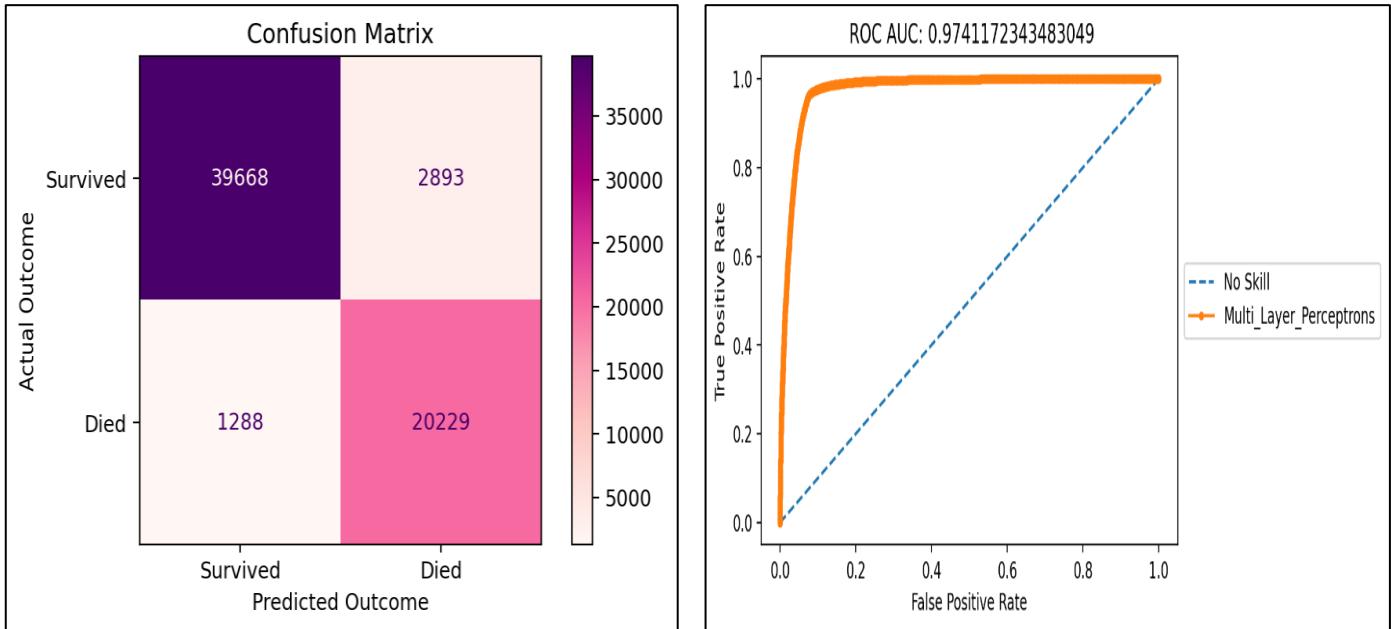
- Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-100** και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,93401**, Recall: **0,95406**, F1-score: **0,90666**, AUC_ROC: **0,97262**, Runtime: **117,22776** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 96**) και η καμπύλη ROC (**Εικόνα 97**) του MLP μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 96: Ο Πίνακας Σύγχυσης του MLPs μοντέλου **22_Min-Max 0-100_opt_01** με την μέγιστη **AUC_ROC**

Εικόνα 97: Η καμπύλη ROC του MLPs μοντέλου **22_Min-Max 0-100_opt_01** με την μέγιστη **AUC_ROC**

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρος **0-1000** και με τις **optimal_02** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,9342**, Recall: **0,95279**, F1-score: **0,90667**, AUC_ROC: **0,97292**, Runtime: **181,84466** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 98**) και η καμπύλη ROC (**Εικόνα 99**) του MLP μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 98: Ο Πίνακας Σύγχυσης του **MLPs** μοντέλου **22_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

Εικόνα 99: Η καμπύλη ROC του **MLPs** μοντέλου **22_Min-Max 0-1000_opt_02** με την μέγιστη **AUC_ROC**

Σύμφωνα με τα παραπάνω για τα MLPs, οι βέλτιστοι τρόποι scaling είναι ο **Min-Max** scaler, για το εύρη **0-100** και **0-1000**, με τις **optimal_01** και **optimal_02** υπερπαραμέτρους.

Για τον αλγόριθμο των **K Εγγύτερων Γειτόνων** (KNN) έχουμε τους παρακάτω 3 πίνακες:

❖ Για τα 22 χαρακτηριστικά (Πίνακας 45)

K-Nearest		22 Features				Precision				Recall				F1-score				ROC AUC				Runtime(secs)			
Scaler Type	Parameter Type	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max												
No Scaler	Default	0,92474	0,00151	0,92287	0,92782	0,91707	0,00316	0,91253	0,92403	0,89106	0,00212	0,88806	0,89562	0,95785(5)	0,00072	0,95673	0,95905	539,20312(313*)	18,3398	513,316	563,551				
	Optimal_01	0,92463	0,00075	0,92348	0,9257	0,91208	0,00185	0,90947	0,91479	0,89051	0,00115	0,88865	0,89216	0,95758(7)	0,00072	0,95609	0,95849	876,82187(49)(319*)	16,0999	853,919	906,503				
	Optimal_02	0,91984(46)	0,00105	0,9182	0,9218	0,89208(48)	0,00228	0,88958	0,89608	0,88199(47)	0,00162	0,87958	0,88479	0,9471	0,0008	0,94602	0,94854	155,39049	8,57598	139,968	169,284				
Standard Scaler	Default	0,92854(1)	0,00081	0,92736	0,93007	0,92947(1)	0,00117	0,92783	0,93134	0,89737(1)	0,00105	0,89566	0,89931	0,9593(1)	0,00068	0,95837	0,9605	549,69155(314*)	16,6064	506,608	567,082				
	Optimal_01	0,92717(2)	0,00126	0,92545	0,9297	0,92555(4)	0,00148	0,92307	0,92776	0,89515(3)	0,00157	0,89297	0,89822	0,95843(2)	0,00093	0,95713	0,96018	882,94407(50)(320*)	18,3219	853,045	902,281				
	Optimal_02	0,9218	0,00109	0,92051	0,92349	0,90199	0,00135	0,8892	0,90377	0,88566	0,00151	0,88372	0,88825	0,94737	0,00048	0,94681	0,94817	530,22712(312*)	9,14194	514,618	546,34				
MinMax Scaler 0-1	Default	0,9254(6)	0,00332	0,91721	0,92978	0,92219(7)	0,00735	0,90621	0,93093	0,89249(6)	0,00504	0,88011	0,89876	0,95739(10)	0,00237	0,95209	0,96015	551,17774(48)(318*)	15,1267	520,41	578,977				
	Optimal_01	0,92537(7)	0,00286	0,91924	0,92905	0,91901	0,00917	0,90168	0,93063	0,89207(8)	0,00457	0,88249	0,89787	0,95723	0,00225	0,95231	0,96009	892,2192(51)(321*)	19,2438	856,857	939,854				
	Optimal_02	0,92043	0,00243	0,91508	0,92449	0,89766	0,00635	0,88469	0,90599	0,88346	0,0038	0,87493	0,88979	0,94707(45)	0,00191	0,94252	0,95002	359,71877(309*)	236,263	29,229	611,874				
MinMax Scaler 0-10	Default	0,9248	0,00363	0,91721	0,92978	0,92015(10)	0,0074	0,90621	0,93024	0,8915	0,00545	0,88011	0,89876	0,95708	0,00261	0,95209	0,96015	551,02913(47)(317*)	14,6717	521,345	578,977				
	Optimal_01	0,92472	0,00299	0,91924	0,92905	0,91643	0,00921	0,90168	0,93063	0,89097	0,00475	0,88249	0,89787	0,95686	0,00244	0,95231	0,96009	894,3036(52)(322*)	20,203	859,675	939,854				
	Optimal_02	0,9201(45)	0,00264	0,91508	0,92449	0,89598(45)	0,00639	0,88469	0,90599	0,88288(45)	0,00411	0,87493	0,88979	0,94679(46)	0,00202	0,94252	0,95001	281,23095(307*)	222,47	29,229	610,807				
MinMax Scaler 0-100	Default	0,92328	0,00343	0,91721	0,92764	0,91651	0,00621	0,90621	0,92436	0,88915	0,00508	0,88011	0,89545	0,95616	0,00266	0,95209	0,95958	550,69042(46)(316*)	14,8917	521,345	578,977				
	Optimal_01	0,9233	0,00261	0,91924	0,92737	0,91133	0,00661	0,90168	0,92177	0,88863	0,00401	0,88249	0,89426	0,95594	0,00245	0,95231	0,95981	900,16639(53)(323*)	20,4708	862,008	939,854				
	Optimal_02	0,91902(48)	0,00241	0,91508	0,92281	0,89258(47)	0,00483	0,88469	0,9002	0,88105(48)	0,00361	0,87493	0,88627	0,94612(47)	0,0021	0,94252	0,95001	136,51624	96,295	29,229	256,369				
MinMax Scaler 0-1000	Default	0,92023	0,00188	0,91721	0,92324	0,91103	0,00338	0,90621	0,91848	0,88464	0,0028	0,88011	0,88937	0,95377	0,00133	0,95209	0,95557	550,40072(45)(315*)	13,3464	521,345	569,311				
	Optimal_01	0,92091	0,00096	0,91924	0,92277	0,90535	0,00283	0,90168	0,91027	0,88494	0,00146	0,88249	0,88772	0,95372	0,00092	0,95231	0,95507	910,1731(54)(324*)	19,7955	885,441	939,854				
	Optimal_02	0,91689(53)	0,00123	0,91508	0,91885	0,88858(51)	0,00321	0,88469	0,89472	0,87786(52)	0,0019	0,87493	0,88103	0,94455(48)	0,00158	0,94252	0,94688	43,65846	6,1696	29,229	50,9896				

❖ Για τα 15 χαρακτηριστικά (Πίνακας 46)

K-Nearest		15 Features				Precision				Recall				F1-score				ROC AUC				Runtime(secs)			
Scaler Type	Parameter Type	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max												
No Scaler	Default	0,92379	0,00096	0,9224	0,92499	0,91481	0,00337	0,91027	0,92217	0,88963	0,00145	0,88724	0,89145	0,95765(6)	0,0009	0,95642	0,9593	14,7616	0,2656	14,5019	15,3586				
	Optimal_01	0,92397	0,00106	0,92241	0,92574	0,91136	0,00224	0,90852	0,91491	0,88941	0,00149	0,88725	0,89201	0,95708	0,00095	0,95597	0,95862	15,76458	1,14649	14,7891	18,5974				
	Optimal_02	0,91782(51)	0,00107	0,9161	0,91945	0,88452(53)	0,00268	0,88229	0,89017	0,87846(51)	0,00162	0,87602	0,88116	0,94349(51)	0,00064	0,94222	0,94423	95,28238	8,82372	84,9395	117,004				
Standard Scaler	Default	0,92711(3)	0,00161	0,92526	0,93024	0,92768(2)	0,00382	0,92103	0,93522	0,8952(2)	0,00246	0,89199	0,8996	0,95831(3)	0,0015	0,95633	0,96111	49,37343	3,9002	45,397	59,532				
	Optimal_01	0,92707(4)	0,00064	0,92601	0,92819	0,9256(3)	0,00182	0,92259	0,92827	0,89503(4)	0,00099	0,89339	0,89671	0,95828(4)	0,00067	0,95696	0,95922	48,21549	1,83774	45,5659	51,6752				
	Optimal_02	0,91958(47)	0,0012	0,9174	0,92091	0,89576(46)	0,00248	0,89314	0,9004	0,88206(46)	0,00178	0,8789	0,88408	0,9443(49)	0,00077	0,94331	0,94581	510,15542(311*)	156,119	368,592	758,403				
MinMax Scaler 0-1	Default	0,92545(5)	0,00249	0,92019	0,92927	0,92228(6)	0,00674	0,9089	0,93194	0,89255(5)	0,00384	0,8843	0,89795	0,95757(8)	0,00177	0,95309	0,96072	86,15592	99,9605	9,56589	266,629				
	Optimal_01	0,9252(8)	0,00224	0,92056	0,92813	0,91909	0,00729	0,90656	0,9283	0,89188	0,00355	0,88452	0,8964	0,95723	0,00149	0,95432	0,95897	85,77902	97,2027	9,96641	258,211				
	Optimal_02	0,91854(49)	0,00222	0,91378	0,92227	0,89084(49)	0,00656	0,87955	0,90164	0,88016(49)	0,00359	0,87269	0,88565	0,94374(50)	0,00162	0,94032	0,94646	235,9961	167,063	27,6615	613,87				
MinMax Scaler 0-10	Default	0,92509(10)	0,00274	0,92019	0,92927	0,92074(9)	0,00703	0,9089	0,93183	0,89191(10)	0,00418	0,8843	0,89795	0,95746(9)	0,00201	0,95309	0,96072	30,59614	24,6077	9,56589	80,5739				
	Optimal_01	0,9246	0,00226	0,92056	0,92813	0,91701	0,00722	0,90656	0,92747	0,89087	0,00352	0,88452	0,89624	0,95686	0,00153	0,95432	0,95893	31,9289	25,6451	9,96641	79,2615				
	Optimal_02	0,91802(50)	0,00227	0,91378	0,9219	0,88905(50)	0,00649	0,87955	0,90164	0,87926(50)	0,00363	0,87269	0,88546	0,94344(52)	0,00175	0,94032	0,94646	199,17438	164,654	27,6615	454,837				
MinMax Scaler 0-100	Default	0,92402	0,00266	0,92019	0,92785	0,91725	0,00559	0,9089	0,9258	0,89021	0,00402	0,8843	0,89591	0,95691	0,00216	0,95309	0,96072	14,72767(9)	4,74127	9,56589	19,6182				
	Optimal_01	0,92377	0,00231	0,92056	0,92813	0,91322	0,00571	0,90656	0,9247																

❖ Για τα 10 χαρακτηριστικά (Πίνακας 47)

K-Nearest		10 Features															
Scaler Type	Parameter Type	Precision			Recall			F1-score			ROC AUC			Runtime(secs)			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
No Scaler	Default	0,92392	0,00088	0,92253	0,92519	0,91575	0,0028	0,909	0,91854	0,88995	0,00138	0,88765	0,89151	0,95478	0,00064	0,95371	0,95579
	Optimal_01	0,92432	0,0011	0,92216	0,92545	0,91633	0,00326	0,91169	0,92086	0,89057	0,00167	0,88733	0,89213	0,9551	0,00075	0,95398	0,95646
	Optimal_02	0,92404	0,00159	0,92154	0,92684	0,91121	0,00334	0,90734	0,91834	0,88954	0,0023	0,88642	0,89358	0,9516	0,00064	0,9507	0,953
Standard Scaler	Default	0,92498	0,00103	0,92293	0,92615	0,92178(8)	0,00284	0,91649	0,92558	0,89193(9)	0,00158	0,88903	0,89392	0,95587	0,00102	0,95421	0,95707
	Optimal_01	0,92514(9)	0,00099	0,9236	0,92659	0,92326(5)	0,00239	0,91922	0,92752	0,89232(7)	0,00163	0,88971	0,89464	0,95569	0,00091	0,95456	0,95698
	Optimal_02	0,9244	0,00088	0,92321	0,92576	0,91311	0,00283	0,90923	0,91832	0,89019	0,00124	0,88841	0,89216	0,9518	0,00064	0,95089	0,95272
MinMax Scaler 0-1	Default	0,92398	0,00242	0,91895	0,92788	0,91887	0,0063	0,90601	0,92988	0,89027	0,00369	0,8826	0,8964	0,95496	0,00136	0,95165	0,95704
	Optimal_01	0,9245	0,00171	0,92042	0,92804	0,91942	0,00551	0,90773	0,92736	0,89106	0,00275	0,88448	0,89653	0,95525	0,00113	0,95262	0,958
	Optimal_02	0,92458	0,00146	0,92018	0,92665	0,91265	0,00379	0,90345	0,91916	0,89041	0,00226	0,88364	0,8938	0,95187	0,00112	0,9494	0,95444
MinMax Scaler 0-10	Default	0,92349	0,0025	0,91895	0,92697	0,91708	0,00607	0,90601	0,92614	0,88947	0,0038	0,8826	0,89445	0,95478	0,00144	0,95165	0,95704
	Optimal_01	0,92435	0,00189	0,92042	0,92804	0,91857	0,00599	0,90773	0,92736	0,89076	0,00303	0,88448	0,89653	0,95515	0,00124	0,95262	0,958
	Optimal_02	0,92436	0,00157	0,92018	0,92665	0,91194	0,0041	0,90345	0,91916	0,89002	0,0024	0,88364	0,8938	0,95181	0,00118	0,94957	0,95444
MinMax Scaler 0-100	Default	0,9226	0,00249	0,91895	0,92697	0,9149	0,0062	0,90601	0,92614	0,88809	0,0038	0,8826	0,89435	0,95437	0,0015	0,95165	0,95697
	Optimal_01	0,92349	0,00153	0,92042	0,92558	0,91588	0,00527	0,90773	0,92455	0,88933	0,00245	0,88448	0,89302	0,95455	0,00088	0,95262	0,95605
	Optimal_02	0,92394	0,00165	0,92018	0,92623	0,91073	0,00413	0,90345	0,91916	0,88935	0,00248	0,88364	0,89328	0,95158	0,00105	0,94972	0,95343
MinMax Scaler 0-1000	Default	0,92047	0,0011	0,91895	0,92275	0,90948	0,00257	0,90601	0,91442	0,88478	0,00157	0,8826	0,88795	0,95323	0,00089	0,95165	0,95451
	Optimal_01	0,92253	0,00141	0,92042	0,92479	0,91149	0,00243	0,90773	0,91582	0,88764	0,002	0,88448	0,89081	0,95395	0,00075	0,95262	0,95515
	Optimal_02	0,92288	0,00147	0,92018	0,9249	0,90778	0,00247	0,90345	0,91131	0,8877	0,00213	0,88364	0,89086	0,95121	0,001	0,94972	0,95241

Περιγραφή Αποτελεσμάτων των K Εγγύτερων Γειτόνων

Ακρίβεια (Precision)

Τον υψηλότερο μέσο όρο (mean) precision= **0,92854** σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **549,69155** seconds. Τον χαμηλότερο μέσο όρο precision= **0,91542**, μικρότερο κατά **0,01312 (-1,41%)**, σε σχέση με το μοντέλο KNN με τη μέγιστη precision, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **30,64266** seconds, που είναι πολύ χαμηλότερος κατά **519,04889** seconds (**-94,43%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Ευαισθησία-Ανάκληση (Recall)

Τον υψηλότερο μέσο όρο (mean) recall= **0,92947** σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **549,69155** seconds. Τον χαμηλότερο μέσο όρο recall = **0,88231**, μικρότερο κατά **0,04716 (-5,07%)**, σε σχέση με το μοντέλο KNN με τη μέγιστη recall, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **30,64266** seconds, που είναι πολύ χαμηλότερος κατά **519,04889** seconds (**-94,43%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Βαθμολογία F1 (F1 Score)

Τον υψηλότερο μέσο όρο (mean) F1 score= **0,89737** σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **549,69155** seconds. Τον χαμηλότερο μέσο όρο F1 score = **0,87515**, μικρότερο κατά **0,02222 (-2,48%)**, σε σχέση με το μοντέλο KNN με το μέγιστο F1 score, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **30,64266** seconds, που είναι πολύ χαμηλότερος κατά **519,04889 sec (-94,43%)**, σε σχέση με το μοντέλο που σημείωσε το μέγιστο F1 Score.

AUC_ROC

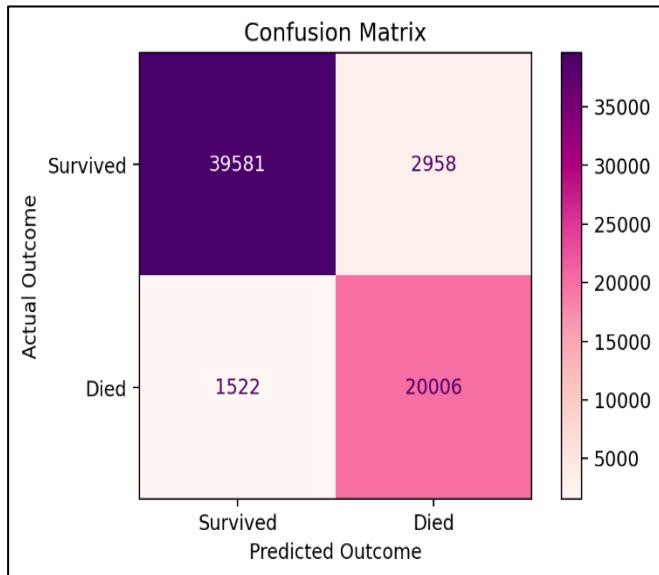
Τον υψηλότερο μέσο όρο (mean) AUC_ROC= **0,9593** σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **549,69155** seconds. Τον χαμηλότερο μέσο όρο AUC_ROC= **0,94146**, μικρότερο κατά **0,01784 (-1,86%)**, σε σχέση με το μοντέλο KNN με τη μέγιστη AUC_ROC, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **30,64266** seconds, που είναι πολύ χαμηλότερος κατά **519,04889 sec (-94,43%)**, σε σχέση με το μοντέλο που σημείωσε τη μέγιστη τιμή AUC_ROC.

Διάρκεια (Runtime)

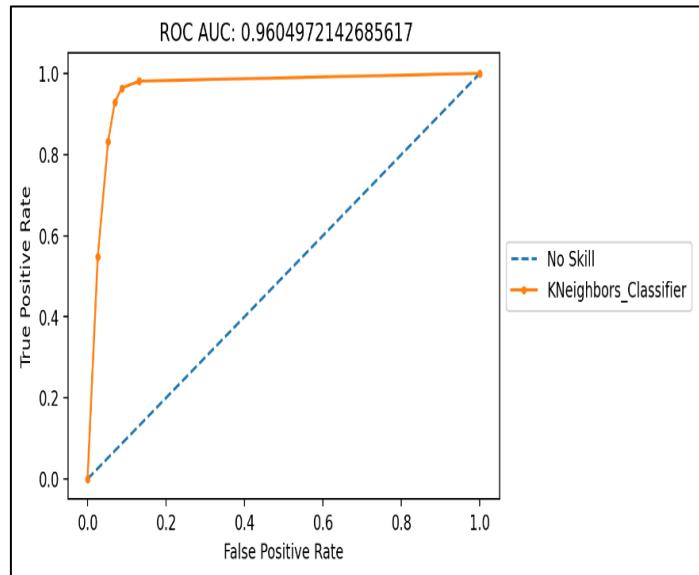
Τον χαμηλότερο μέσο όρο(means) runtime= **8,06013** seconds σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,92047**. Τον υψηλότερο μέσο όρο runtime= **910,1731** seconds, πολύ υψηλότερο κατά **902,11297** seconds (**+11.192,29%**), σε σχέση με το μοντέλο KNN με τον χαμηλότερο μέσο όρο runtime, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο precision της κάθε επανάληψης **0,92091**, ελάχιστα υψηλότερο κατά **0,00044 (+0,047%)**, σε σχέση με το μοντέλο που σημείωσε τη χαμηλότερο μέσο όρο runtime.

Τα KNN μοντέλα με την καλύτερη απόδοση με βάση τα 5 μετρικά είναι τα παρακάτω:

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα 22, με τη χρήση του Standard scaler και με τις default τιμές υπερπαραμέτρων, με μετρικά Precision: 0,92854, Recall: 0,92947, F1-score: 0,89737, AUC_ROC: 0,9593, Runtime: 549,69155 sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (Εικόνα 100) και η καμπύλη ROC (Εικόνα 101) του KNN μοντέλου με την μέγιστη AUC_ROC.

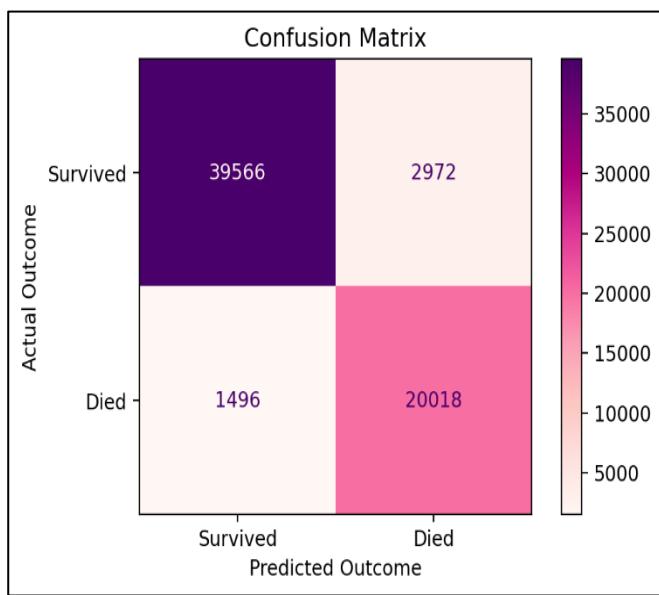


Εικόνα 100: Ο Πίνακας Σύγχυσης του KNN μοντέλου 22_Std_default με την μέγιστη AUC_ROC

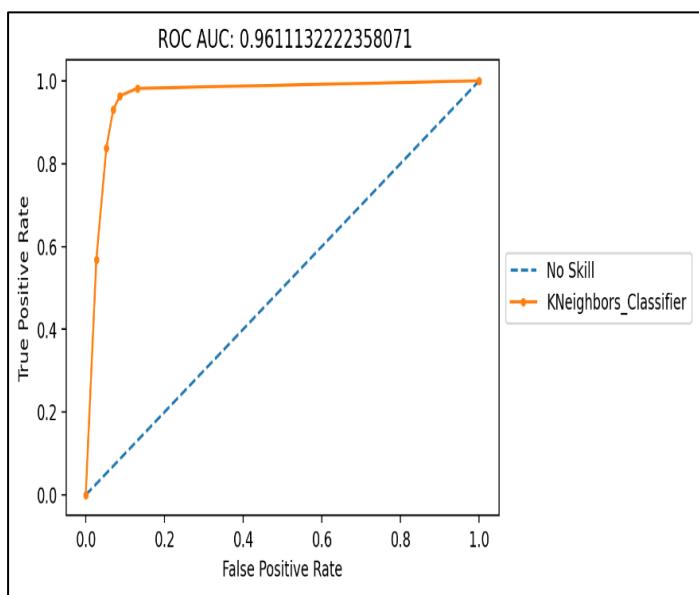


Εικόνα 101: Η καμπύλη ROC του KNN μοντέλου 22_Std_default με την μέγιστη AUC_ROC

- Τα μοντέλα τα οποία χρησιμοποίησαν και τα 15 χαρακτηριστικά, με τη χρήση του Standard scaler και με τις default τιμές υπερπαραμέτρων, με μετρικά Precision: 0,92711, Recall: 0,92768, F1-score: 0,8952, AUC_ROC: 0,95831, Runtime: 49,37343 sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (Εικόνα 102) και η καμπύλη ROC (Εικόνα 103) του KNN μοντέλου με την μέγιστη AUC_ROC.

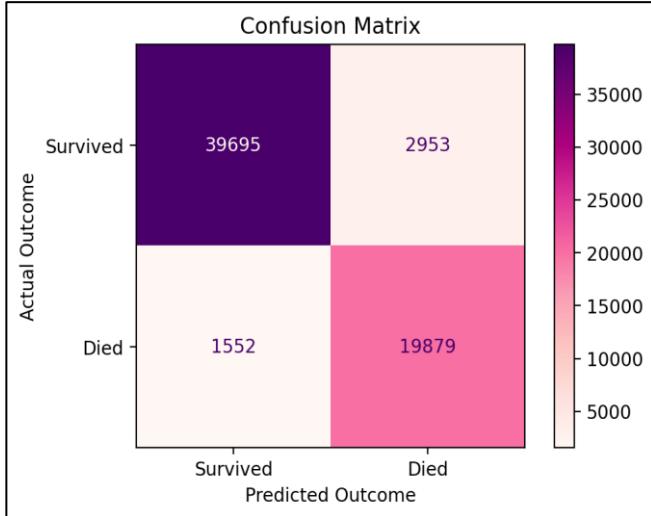


Εικόνα 102: Ο Πίνακας Σύγχυσης του KNN μοντέλου 15_Std_default με την μέγιστη AUC_ROC

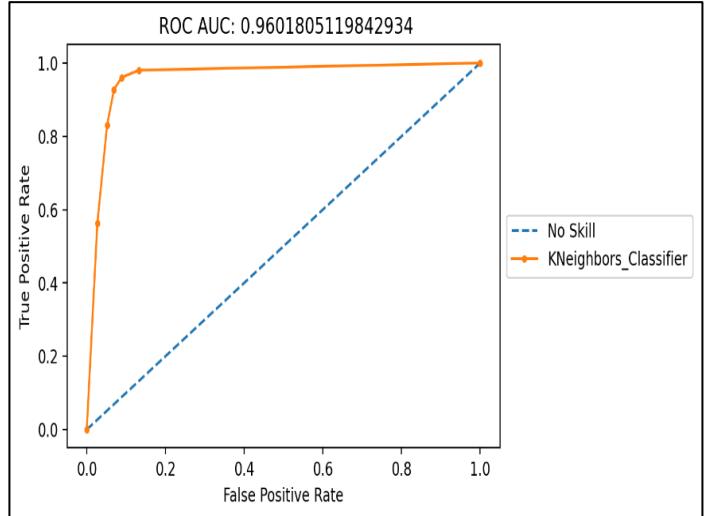


Εικόνα 103: Η καμπύλη ROC του KNN μοντέλου 15_Std_default με την μέγιστη AUC_ROC

3. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **22** χαρακτηριστικά, με τη χρήση του **Standard** scaler και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92717**, Recall: **0,92555**, F1-score: **0,89515**, AUC_ROC: **0,95843**, Runtime: **882,94407** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 104**) και η καμπύλη ROC (**Εικόνα 105**) του KNN μοντέλου με την μέγιστη AUC_ROC.

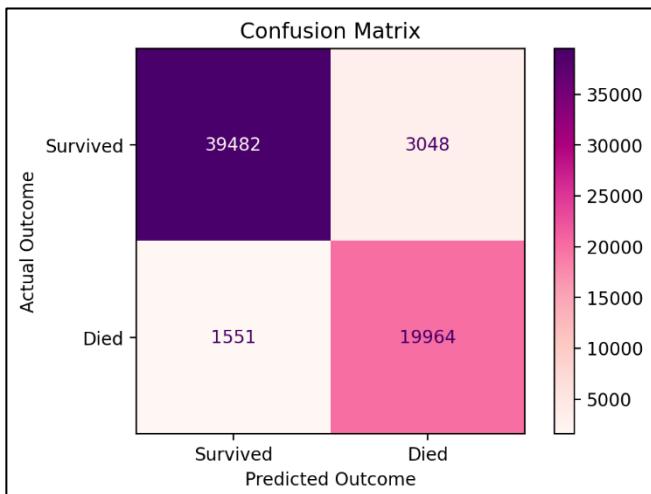


Εικόνα 104: Ο Πίνακας Σύγχυσης του KNN μοντέλου **22_Std_opt_01** με την μέγιστη **AUC_ROC**

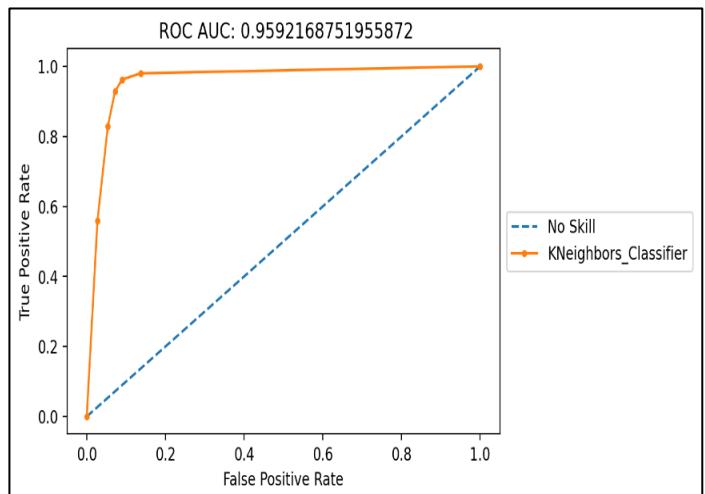


Εικόνα 105: Η καμπύλη ROC του KNN μοντέλου **22_Std_opt_01** με την μέγιστη **AUC_ROC**

4. Τα μοντέλα τα οποία χρησιμοποίησαν και τα **15** χαρακτηριστικά, με τη χρήση του **Standard** scaler και με τις **optimal_01** τιμές υπερπαραμέτρων, με μετρικά Precision: **0,92707**, Recall: **0,9256**, F1-score: **0,89503**, AUC_ROC: **0,95828**, Runtime: **48,21549** sec. Παρακάτω απεικονίζονται ο πίνακας σύγχυσης (**Εικόνα 106**) και η καμπύλη ROC (**Εικόνα 107**) του KNN μοντέλου με την μέγιστη AUC_ROC.



Εικόνα 106: Ο Πίνακας Σύγχυσης του KNN μοντέλου **15_Std_opt_01** με την μέγιστη **AUC_ROC**

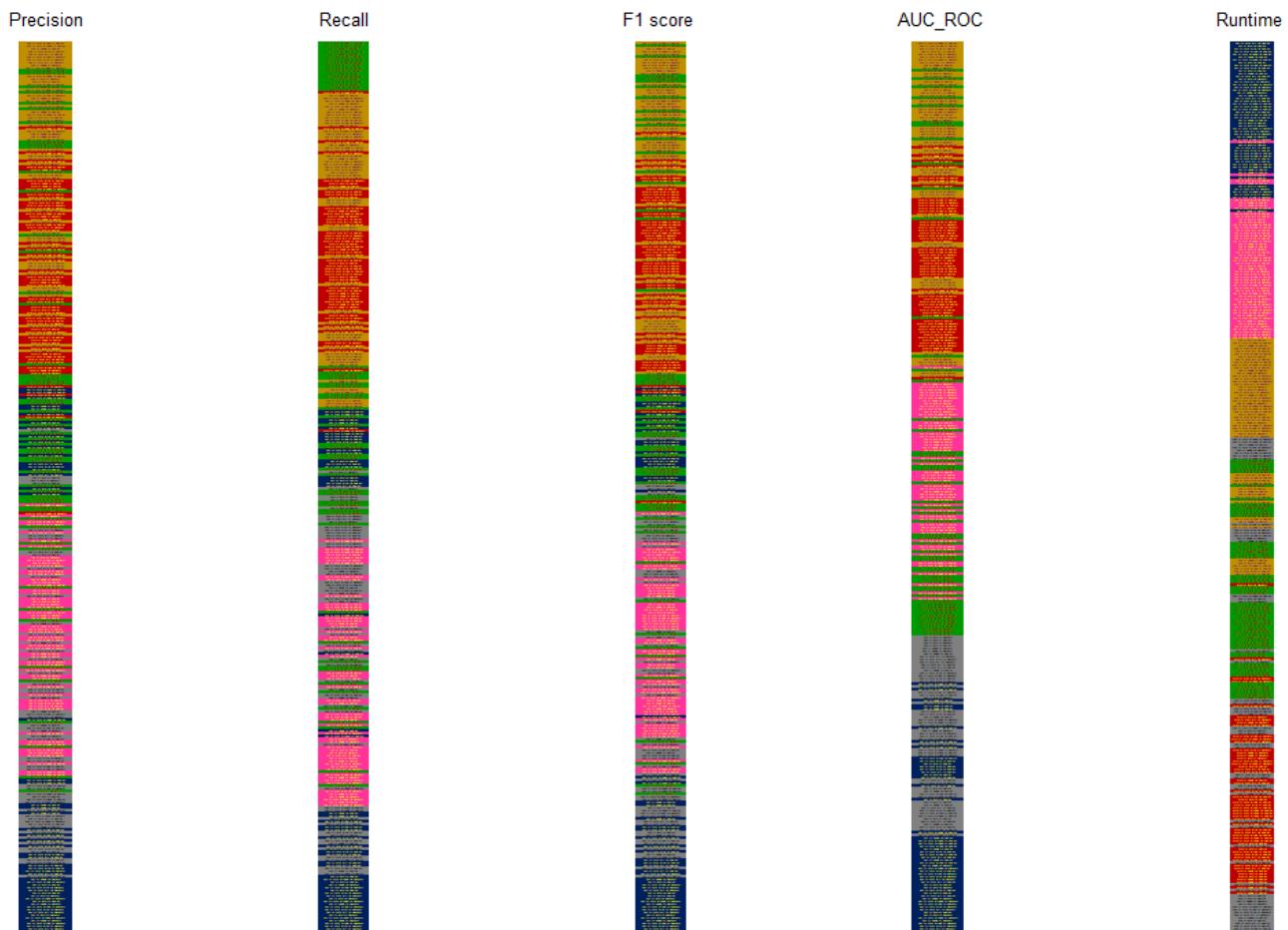


Εικόνα 107: Η καμπύλη ROC του KNN μοντέλου **15_Std_opt_01** με την μέγιστη **AUC_ROC**

Σύμφωνα με τα παραπάνω για τον αλγόριθμό KNN, ο βέλτιστος τρόπος scaling είναι ο **Standard**, με τις **default** και **optimal_01** υπερπαραμέτρους.

5.3 Σύγκριση Αποτελεσμάτων των αλγορίθμων

Η παρακάτω εικόνα (**Εικόνα 108**) εμφανίζει την διάταξη των μέσων όρων όλων των μοντέλων των αλγορίθμων.



Εικόνα 108: Απεικόνιση της διάταξης των μέσων όρων (mean) όλων των 324 διαφορετικών μοντέλων των αλγορίθμων (Λογιστική Παλινδρόμηση: █, Δέντρα Αποφάσεων: █, Τυχαίο Δάσος: █, Ακραία Ενίσχυση Κλίσης: █, ΤΝΔ Πολυστρωματικών Αντιλήπτρων: █, Κ Εγγύτεροι Γείτονες: █), για το κάθε μετρικό(από αριστερά προς τα δεξιά: **Precision**, **Recall**, **F1-score**, **AUC_ROC**, **Runtime**), με την τιμή του κάθε μετρικού, εκτός από το μετρικό της **Runtime** (Διάρκειας), να ακολουθεί φθίνουσα πορεία από πάνω προς τα κάτω, με την **Runtime** να ακολουθεί αύξουσα πορεία από πάνω προς τα κάτω.

Περιγραφή Συγκεντρωτικών Αποτελεσμάτων

Ακρίβεια (Precision)

Τα μοντέλα **XGB** είναι εκείνα τα οποία εμφανίζουν τη υψηλότερες τιμές mean για την Precision, με αυτές να κυμαίνονται από **0,93764** στην **1^η** θέση (runtime= **6,67306** seconds), έως **0,9321** στην **113^η** θέση (runtime= **14,24933** seconds), με το **50%** αυτών (28/54) να εντοπίζεται έως και την **54^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** χαρακτηριστικά, με **Min-Max** scaler για τα εύρη **0-100** και **0-1000** και τις **optimal _01** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα **RF** μοντέλα, με τις τιμές mean για την Precision να κυμαίνονται από **0,93661** στην **11^η** θέση (runtime= **29,84745** seconds), έως **0,92253** στην **273^η** θέση (runtime= **16,67103** seconds), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη

διασπορά με το **50%** αυτών (28/54) να εντοπίζεται έως και την **145^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-100** και **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _02** τιμές υπερπαραμέτρων.

Στην τρίτη θέση ακολουθούν τα μοντέλα των **MLPs**, με τις τιμές mean για την Precision να κυμαίνονται από **0,93456** στην **32^η** θέση (runtime= **133,76172** seconds), έως **0,92598** στην **172^η** θέση (runtime= **28,3394** seconds), με το **50%** αυτών (28/54) να εντοπίζεται έως και την **88^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** χαρακτηριστικά, με scaler τον **Min-Max** για τα εύρη **0-100** και **0-1000** και τις **optimal _01** και **optimal _02** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα μοντέλα των **DTs**, με τις τιμές mean για την Precision να κυμαίνονται από **0,93035** στην **127^η** θέση (runtime= **1,45128** seconds), έως **0,90092** στην **324^η** θέση (runtime= **1,61526** seconds), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη διασπορά με το **64,8%** αυτών (35/54) να εντοπίζεται από την **269^η** έως και την **324^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _01** τιμές υπερπαραμέτρων.

Στη συνέχεια ακολουθούν τα μοντέλα των **KNN**, με τις τιμές mean για την Precision να κυμαίνονται από **0,92854** στην **142^η** θέση (runtime= **549,69155** seconds), έως **0,91542** στην **304^η** θέση (runtime= **30,64266** seconds), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη διασπορά με το **66,6%** αυτών (36/54) να εντοπίζεται από την **234^η** έως και την **304^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με τον **Standard** scaler και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Τελευταία έχουμε τα μοντέλα των **LR**, με τις τιμές mean για την Precision να κυμαίνονται από **0,92626** στην **169^η** θέση (runtime= **3,17691** seconds), έως **0,92322** στην **267^η** θέση (runtime= **2,12615** seconds), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει διασπορά με το **70,3%** αυτών (38/54) να εντοπίζεται από την **188^η** έως και την **243^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με τον **Min-Max** scaler με εύρη **0-1000** και **0-100** και με τον **Standard** scaler και τις **default** τιμές υπερπαραμέτρων.

Κορυφαία 20 μοντέλα

Τα πρώτα **20** σε κατάταξη μοντέλα περιλαμβάνουν **16** μοντέλα **XGB** και **4** μοντέλα **RF**, με τα κορυφαία **10** να είναι μόνο μοντέλα **XGB**. Από τα **16** μοντέλα **XGB**, τον υψηλότερο μέσο όρο(mean) precision= **0,93764**, σημείωσε αυτό που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal _01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,67306** seconds. Επίσης από τα **20** πρώτα τον χαμηλότερο μέσο όρο precision= **0,93532**, μικρότερο κατά **0,00232 (-0,24%)**, σε σχέση με το μοντέλο XGB με τη μέγιστη precision, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **default** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,23783** seconds, που είναι χαμηλότερος κατά **0,43523** seconds (**-6,52%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Από τα **4** μοντέλα **RF** τον υψηλότερο μέσο όρο (mean) precision= **0,93661** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal _02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **29,84745** seconds. Το μοντέλο RF με τον χαμηλότερο μέσο όρο precision= **0,93544**, ελάχιστα μικρότερο κατά **0,00117 (-0,12%)**, σε σχέση με το μοντέλο RF με τη μέγιστη

precision, σημείωσε το μοντέλο που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **29,71669** seconds, που είναι ελάχιστα χαμηλότερος κατά **0,13076** seconds (**-0,44%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη precision.

Τελευταία 20 μοντέλα

Τα τελευταία **20** σε κατάταξη μοντέλα περιλαμβάνουν μόνο μοντέλα **DTs**. Τον υψηλότερο μέσο όρο (mean) precision= **0,91531** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, χωρίς scaler(**no_scaler**), με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,19258** seconds. Τον χαμηλότερο μέσο όρο precision= **0,90092**, μικρότερο κατά **0,01439** (**-1,57%**), σε σχέση με το προηγούμενο μοντέλο DT, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,61526** seconds, που είναι υψηλότερος κατά **0,42268** seconds (**+35,44%**), σε σχέση με το προηγούμενο μοντέλο DT. Στις παρακάτω εικόνες απεικονίζονται τα κορυφαία 20 (Εικόνα 109) και τα τελευταία 20 (Εικόνα 110) μοντέλα για την Precision.

1	xbc_fc_mm_0-100_22_opt-01
2	xbc_fc_mm_0-1000_22_opt-01
3	xbc_fc_none_22_opt-01
4	xbc_fc_mm_0-10_22_opt-01
5	xbc_fc_mm_0-1_22_opt-01
6	xbc_fc_mm_0-100_22_default
7	xbc_fc_mm_0-10_22_default
8	xbc_fc_mm_0-1_22_default
9	xbc_fc_none_22_default
10	xbc_fc_mm_0-1000_22_default
11	rfc_fc_mm_0-1000_22_opt-02
12	rfc_fc_mm_0-1000_15_opt-02
13	xbc_fc_mm_0-1000_15_opt-01
14	xbc_fc_std_22_default
15	xbc_fc_std_22_opt-01
16	xbc_fc_mm_0-100_15_opt-01
17	rfc_fc_none_15_opt-02
18	xbc_fc_mm_0-1000_15_default
19	rfc_fc_mm_0-100_22_opt-02
20	xbc_fc_mm_0-100_15_default

Εικόνα 109: Τα κορυφαία 20 μοντέλα για την Precision

305	dtc_fc_none_10_default
306	dtc_fc_mm_0-100_10_default
307	dtc_fc_std_15_opt-02
308	dtc_fc_mm_0-10_10_default
309	dtc_fc_mm_0-1_10_default
310	dtc_fc_std_10_opt-02
311	dtc_fc_std_10_default
312	dtc_fc_mm_0-1000_15_default
313	dtc_fc_std_22_opt-02
314	dtc_fc_none_15_default
315	dtc_fc_mm_0-100_15_default
316	dtc_fc_mm_0-10_15_default
317	dtc_fc_mm_0-1_15_default
318	dtc_fc_std_15_default
319	dtc_fc_mm_0-1000_22_default
320	dtc_fc_none_22_default
321	dtc_fc_mm_0-100_22_default
322	dtc_fc_mm_0-10_22_default
323	dtc_fc_mm_0-1_22_default
324	dtc_fc_std_22_default

Εικόνα 110: Τα τελευταία 20 μοντέλα για την Precision

Ευαισθησία-Ανάκληση (Recall)

Τα **RF** είναι εκείνα τα οποία εμφανίζουν τη υψηλότερες τιμές mean για την Recall, με αυτές να κυμαίνονται από **0,9699** στην **1^η θέση** (runtime= **29,84745 seconds**), έως **0,90832** στην **271^η θέση** (runtime= **13,94055 seconds**), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη διασπορά με το **33,3%** αυτών (18/54) να εντοπίζεται έως και την **18^η θέση** και το **44,4%** (24/54) να εντοπίζεται μεταξύ της **119^{ης}** και **176^{ης} θέσης**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-100** και **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _02** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα **MLPs** μοντέλα, με τις τιμές mean για την Recall να κυμαίνονται από **0,95623** στην **19^η θέση** (runtime= **133,76172 seconds**), έως **0,93434** στην **142^η θέση** (runtime= **28,3394 seconds**), με το **88,9%** αυτών (48/54) να εντοπίζεται από την **51^η έως και την **113^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και τα **10** χαρακτηριστικά, με scaler τον **Min-Max** για τα εύρη **0-100** και **0-1000** και τις **optimal _01** και **optimal _02** τιμές υπερπαραμέτρων.**

Στην τρίτη θέση ακολουθούν τα μοντέλα των **XGB**, με τις τιμές mean για την Recall να κυμαίνονται από **0,9561** στην **20^η θέση** (runtime= **6,12796 seconds**), έως **0,93958** στην **133^η θέση** (runtime= **17,39536 seconds**), με το **63%** αυτών (34/54) να εντοπίζεται έως και την **69^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και τα **10** χαρακτηριστικά με scaler είτε τον **Min-Max** για τα εύρη **0-1000** και **0-100** είτε χωρίς scaler (**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα μοντέλα των **DTs** με τις τιμές mean για την Recall να κυμαίνονται από **0,9369** στην **135^η θέση** (runtime= **1,25677 seconds**), έως **0,8342** στην **324^η θέση** (runtime= **1,61526 seconds**), με το **33,3%** αυτών (18/54) να εντοπίζεται από την **135^η έως και την **162^η θέση** και το **59,2%** (42/54) να εντοπίζεται από την **281^η έως και την **324^η**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _01** τιμές υπερπαραμέτρων.****

Στη συνέχεια ακολουθούν τα μοντέλα των **KNN**, με τις τιμές mean για την Recall να κυμαίνονται, από **0,92947** στην **157^η θέση** (runtime= **549,69155 seconds**), έως **0,88231** στην **303^η θέση** (runtime= **30,64266 seconds**), με το **50%** αυτών (28/54) να εντοπίζεται από την **157^η έως και την **216^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με τον **Standard** scaler και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.**

Τέλος έχουμε τα μοντέλα των **LR**, με τις τιμές mean για την Recall να κυμαίνονται από **0,91899** στην **182^η θέση** (runtime= **1,79217 seconds**), έως **0,90785** στην **278^η θέση** (runtime= **2,72038 seconds**), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει μεγάλη διασπορά, με το **50%** αυτών (28/54) να εντοπίζεται από την **240^η έως και την **278^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** και **22** χαρακτηριστικά, με τον **Min-Max** scaler με εύρη **0-1000** και **0-1** και τις **optimal _02** και **optimal _01** τιμές υπερπαραμέτρων.**

Κορυφαία 20 μοντέλα

Τα πρώτα **20** σε κατάταξη μοντέλα περιλαμβάνουν **18** μοντέλα **RF**, **1** μοντέλο **MLPs** και **1** μοντέλο **XGB** με τα κορυφαία **18** να είναι μόνο μοντέλα **RF**. Από τα **18** μοντέλα **RF**, τον υψηλότερο μέσο όρο (mean) recall = **0,9699**, σημείωσε αυτό που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal _02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **29,84745 seconds**.

Επίσης για τα μοντέλα **RF**, τον χαμηλότερο μέσο όρο recall = **0,95997**, μικρότερο κατά **0,00993 (-1,08%)**, σε σχέση με το μοντέλο με τη μέγιστη recall, σημείωσε το μοντέλο που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Standard** scaler, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **24,19302** seconds, που είναι χαμηλότερος κατά **5,65443** seconds (**-18,94%**), σε σχέση με το μοντέλο που σημείωσε τη μέγιστη recall.

Το μοντέλο των **MLPs** που κατατάσσεται στην **19^η** θέση έχει μέσο όρο (mean) recall = **0,95623**, χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **133,76172** seconds.

Το μοντέλο των **XGB** που κατατάσσεται στην **20^η** θέση έχει μέσο όρο (mean) recall = **0,95384**, χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,23783** seconds.

Τελευταία 20 μοντέλα

Εδώ όπως και για την Precision τα τελευταία **20** σε κατάταξη μοντέλα περιλαμβάνουν μόνο μοντέλα **DTs**. Τον υψηλότερο μέσο όρο (mean) recall = **0,88012** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,18742** seconds. Τον χαμηλότερο μέσο όρο recall = **0,8342**, μικρότερο κατά **0,04592 (-5,22%)**, σε σχέση με το προηγούμενο μοντέλο DT, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,61526** seconds, που είναι υψηλότερος κατά **0,42784** seconds (**+36,03%**), σε σχέση με το προηγούμενο μοντέλο DT. Στις παρακάτω εικόνες απεικονίζονται τα κορυφαία 20 (**Εικόνα 111**) και τα τελευταία 20 (**Εικόνα 112**) μοντέλα για την Recall.

1	rfc_fc_mm_0-1000_22_opt-02
2	rfc_fc_mm_0-100_22_opt-02
3	rfc_fc_none_22_opt-02
4	rfc_fc_mm_0-1000_15_opt-02
5	rfc_fc_none_15_opt-02
6	rfc_fc_mm_0-10_22_opt-02
7	rfc_fc_min_0-1000_10_opt-02
8	rfc_fc_mm_0-100_15_opt-02
9	rfc_fc_mm_0-1_22_opt-02
10	rfc_fc_mm_0-10_15_opt-02
11	rfc_fc_mm_0-1_15_opt-02
12	rfc_fc_none_10_opt-02
13	rfc_fc_mm_0-100_10_opt-02
14	rfc_fc_std_22_opt-02
15	rfc_fc_mm_0-10_10_opt-02
16	rfc_fc_std_15_opt-02
17	rfc_fc_mm_0-1_10_opt-02
18	rfc_fc_std_10_opt-02
19	mlp_fc_mm_0-1000_22_opt-01
20	xbc_fc_none_22_opt-01

Εικόνα 111: Τα κορυφαία 20 μοντέλα για την Recall

305	dtc_fc_mm_0-1000_10_default
306	dtc_fc_std_22_opt-02
307	dtc_fc_none_10_default
308	dtc_fc_mm_0-100_10_default
309	dtc_fc_std_10_opt-02
310	dtc_fc_mm_0-10_10_default
311	dtc_fc_mm_0-1_10_default
312	dtc_fc_std_10_default
313	dtc_fc_mm_0-1000_15_default
314	dtc_fc_none_15_default
315	dtc_fc_mm_0-100_15_default
316	dtc_fc_mm_0-10_15_default
317	dtc_fc_mm_0-1_15_default
318	dtc_fc_std_15_default
319	dtc_fc_mm_0-1000_22_default
320	dtc_fc_none_22_default
321	dtc_fc_mm_0-100_22_default
322	dtc_fc_mm_0-10_22_default
323	dtc_fc_mm_0-1_22_default
324	dtc_fc_std_22_default

Εικόνα 112: Τα τελευταία 20 μοντέλα για την Recall

Βαθμολογία F1 (F1 Score)

Τα **XGB** είναι εκείνα τα οποία εμφανίζουν τη υψηλότερες τιμές mean για το F1-score, με αυτές να κυμαίνονται από **0,9113** στην **1^η θέση** (runtime= **6,67306** seconds), έως **0,90332** στην **121^η θέση** (runtime= **14,24933** seconds), με το **63%** αυτών (34/54) να εντοπίζεται έως και την **53^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** χαρακτηριστικά με scaler είτε τον **Min-Max** για τα εύρη **0-1000**, **0-100** και **0-10** είτε χωρίς scaler (**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα **RF** μοντέλα, με τις τιμές mean για το F1-score να κυμαίνονται από **0,91127** στην **2^η θέση** (runtime= **29,84745** seconds), έως **0,8873** στην **274^η θέση** (runtime= **16,67103** seconds), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη διασπορά με το **50%** αυτών (28/54) να εντοπίζεται από την **122^η** έως και την **190^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-100** και **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _02** τιμές υπερπαραμέτρων.

Στην τρίτη θέση ακολουθούν τα μοντέλα των **MLPs**, με τις τιμές mean για το F1-score να κυμαίνονται από **0,90756** στην **34^η θέση** (runtime= **133,76172** seconds), έως **0,89386** στην **168^η θέση** (runtime= **28,3394** seconds), με το **88,9%** αυτών (48/54) να εντοπίζεται από την **50^η** έως και την **120^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και τα **15** χαρακτηριστικά, με scaler τον **Min-Max** για τα εύρη **0-100** και **0-1000** και τις **optimal _01**, **optimal _02** και **default** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα μοντέλα των **DTs**, με τις τιμές mean για το F1-score να κυμαίνονται από **0,90032** στην **127^η θέση** (runtime= **1,45128** seconds), έως **0,84959** στην **324^η θέση** (runtime= **1,61526** seconds), με το **33,3%** αυτών (18/54) να εντοπίζεται από την **127^η** έως και την **164^η θέση** και το **64,8%** (35/54) να εντοπίζεται από την **268^η έως και την 324^η**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _01** τιμές υπερπαραμέτρων.

Στη συνέχεια ακολουθούν τα μοντέλα των **KNN**, με τις τιμές mean για το F1-score να κυμαίνονται από **0,89737** στην **145^η θέση** (runtime= **549,69155** seconds), έως **0,87515** στην **304^η θέση** (runtime= **30,64266** seconds), με το **31,5%** αυτών (17/54) να εντοπίζεται από την **145^η** έως και την **203^η θέση** και το **50%** αυτών (28/54) να εντοπίζεται από την **254^η έως και την 304^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με τον **Standard** scaler και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Τέλος έχουμε τα μοντέλα των **LR**, με τις τιμές mean για το F1-score να κυμαίνονται από **0,89262** στην **172^η θέση** (runtime= **3,17691** seconds), έως **0,88879** στην **266^η θέση** (runtime= **1,83556** seconds), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει μεγάλη διασπορά με το **90,7%** αυτών (49/54) να εντοπίζεται από την **183^η έως και την 253^η θέση**. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** χαρακτηριστικά, με scaler είτε τον **Min-Max** scaler με εύρη **0-1000** και **0-100** είτε τον **Standard** scaler και τις **default** και **optimal _02** τιμές υπερπαραμέτρων.

Κορυφαία 20 μοντέλα

Τα πρώτα **20** σε κατάταξη μοντέλα περιλαμβάνουν **14** μοντέλα **XGB** και **6** μοντέλα **RF**. Από τα **14** μοντέλα **XGB**, τον υψηλότερο μέσο όρο (mean) f1-score= **0,9113**, σημείωσε αυτό που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal _01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,67306** seconds. Επίσης για τα μοντέλα **XGB**, τον χαμηλότερο μέσο όρο f1-score= **0,9086**,

μικρότερο κατά **0,0027** (-0,30%), σε σχέση με το προηγούμενο μοντέλο, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **5,51039** seconds, που είναι χαμηλότερος κατά **1,16267** seconds (-17,42%), σε σχέση με το μοντέλο που σημείωσε το μέγιστο f1-score.

Από τα **6** μοντέλα **RF** τον υψηλότερο μέσο όρο (mean) f1-score= **0,91127** σημείωσε το μοντέλο RF που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **29,84745** seconds. Το μοντέλο **RF** με τον χαμηλότερο μέσο όρο f1-score= **0,90903**, ελάχιστα μικρότερο κατά **0,00224** (-0,25%), σε σχέση με το προηγούμενο **RF** μοντέλο, σημείωσε το μοντέλο που χρησιμοποίησε τα **15** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-100**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **26,6195** seconds, που είναι χαμηλότερος κατά **3,22795** seconds (-10,81%), σε σχέση με το προηγούμενο **RF** μοντέλο.

Τελευταία 20 μοντέλα

Εδώ τα τελευταία **20** σε κατάταξη μοντέλα, όπως και στα προηγούμενα μετρικά, περιλαμβάνουν μόνο μοντέλα **DTs**. Τον υψηλότερο μέσο όρο (mean) f1-score= **0,87437** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,19258** seconds. Τον χαμηλότερο μέσο όρο f1-score= **0,84959**, μικρότερο κατά **0,02478** (-2,83%), σε σχέση με το προηγούμενο μοντέλο DT, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,61526** seconds, που είναι υψηλότερος κατά **0,42268** seconds (+35,44%), σε σχέση με το προηγούμενο μοντέλο DT. Στις παρακάτω εικόνες απεικονίζονται τα κορυφαία 20 (Εικόνα 113) και τα τελευταία 20 (Εικόνα 114) μοντέλα για το F1 σκορ.

1	xbc_fc_mm_0-100_22_opt-01
2	rfc_fc_mm_0-1000_22_opt-02
3	xbc_fc_none_22_opt-01
4	xbc_fc_mm_0-1000_22_opt-01
5	xbc_fc_mm_0-10_22_opt-01
6	rfc_fc_mm_0-1000_15_opt-02
7	xbc_fc_mm_0-1_22_opt-01
8	xbc_fc_mm_0-100_22_default
9	xbc_fc_mm_0-10_22_default
10	xbc_fc_mm_0-1_22_default
11	xbc_fc_none_22_default
12	xbc_fc_mm_0-1000_22_default
13	rfc_fc_none_15_opt-02
14	rfc_fc_mm_0-100_22_opt-02
15	rfc_fc_none_22_opt-02
16	xbc_fc_mm_0-1000_15_opt-01
17	rfc_fc_mm_0-100_15_opt-02
18	xbc_fc_std_22_default
19	xbc_fc_std_22_opt-01
20	xbc_fc_mm_0-100_15_opt-01

Εικόνα 113: Τα κορυφαία 20 μοντέλα για το F1 score

305	dtc_fc_none_10_default
306	dtc_fc_std_15_opt-02
307	dtc_fc_mm_0-100_10_default
308	dtc_fc_mm_0-10_10_default
309	dtc_fc_mm_0-1_10_default
310	dtc_fc_std_10_opt-02
311	dtc_fc_std_22_opt-02
312	dtc_fc_std_10_default
313	dtc_fc_mm_0-1000_15_default
314	dtc_fc_none_15_default
315	dtc_fc_mm_0-100_15_default
316	dtc_fc_mm_0-10_15_default
317	dtc_fc_mm_0-1_15_default
318	dtc_fc_std_15_default
319	dtc_fc_mm_0-1000_22_default
320	dtc_fc_none_22_default
321	dtc_fc_mm_0-100_22_default
322	dtc_fc_mm_0-10_22_default
323	dtc_fc_mm_0-1_22_default
324	dtc_fc_std_22_default

Εικόνα 114: Τα τελευταία 20 μοντέλα για το F1 score

AUC_ROC

Τα **XGB** είναι εκείνα τα οποία εμφανίζουν τη υψηλότερες τιμές mean για την AUC_ROC, με αυτές να κυμαίνονται από **0,978849** στην **1^η** θέση (runtime= **6,63992** seconds), έως **0,97073** στην **122^η** θέση (runtime= **12,17782** seconds), με το **74,1%** αυτών (40/54) να εντοπίζεται έως και την **57^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** χαρακτηριστικά με scaler είτε τον **Min-Max** για τα εύρη **0-1000, 0-100, 0-10** και **0-1** είτε χωρίς scaler (**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα **RF** μοντέλα, με τις τιμές mean για την AUC_ROC να κυμαίνονται από **0,97708** στην **11^η** θέση (runtime= **29,84745** seconds), έως **0,96155** στην **216^η** θέση (runtime= **18,33744** seconds), η κατανομή των θέσεων των μοντέλων εμφανίζει μεγάλη διασπορά με το **50%** αυτών (28/54) να εντοπίζεται από την **175^η** έως και την **216^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-100** και **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _02** τιμές υπερπαραμέτρων.

Στην τρίτη θέση ακολουθούν τα μοντέλα των **MLPs**, με τις τιμές mean για την AUC_ROC να κυμαίνονται από **0,97352** στην **39^η** θέση (runtime= **267,55604** seconds), έως **0,97064** στην **123^η** θέση (runtime= **199,77529** seconds), με το **85,2%** αυτών (46/54) να εντοπίζεται από την **58^η** έως και την **113^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και τα **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για το εύρος **0-1000**, είτε χωρίς scaler (**no scaler**) και τις **optimal _01, optimal _02** και **default** τιμές υπερπαραμέτρων.

Στη συνέχεια ακολουθούν τα μοντέλα των **LR**, με τις τιμές mean για την AUC_ROC να κυμαίνονται από **0,97081** στην **119^η** θέση (runtime= **2,99462** seconds), έως **0,96459** στην **203^η** θέση (runtime= **1,79217** seconds), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει μεγάλη διασπορά με το **79,6%** αυτών (43/54) να εντοπίζεται από την **125^η** έως και την **179^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και τα **15** χαρακτηριστικά, με scaler τον **Min-Max** scaler με εύρη **0-100, 0-10** και **0-1**, τον **Standard** scaler ή χωρίς scaler (**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Στη συνέχεια ακολουθούν τα μοντέλα των **KNN**, με τις τιμές mean για την AUC_ROC να κυμαίνονται από **0,9593** στην **217^η** θέση (runtime= **549,69155** seconds), έως **0,94146** στην **289^η** θέση (runtime= **30,64266** seconds), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει μεγάλη διασπορά. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με τον **Standard** scaler ή χωρίς scaler(**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Τέλος έχουμε τα μοντέλα των **DTs**, με τις τιμές mean για την AUC_ROC να κυμαίνονται από **0,95666** στην **234^η** θέση (runtime= **1,45128** seconds), έως **0,90031** στην **324^η** θέση (runtime= **1,61526** seconds), με το **66,7%** αυτών (36/54) να εντοπίζεται από την **288^η** έως και την **324^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **22** και **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα εύρη **0-1000** είτε χωρίς scaler (**no scaler**) και τις **optimal _01** τιμές υπερπαραμέτρων.

Κορυφαία 20 μοντέλα

Τα πρώτα **20** σε κατάταξη μοντέλα περιλαμβάνουν **17** μοντέλα **XGB** και **3** μοντέλα **RF**, με τα κορυφαία **10** να είναι μόνο μοντέλα **XGB**. Από τα **17** μοντέλα **XGB**, τον υψηλότερο μέσο όρο (mean) auc_roc= **0,978849**, σημείωσε αυτό που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **6,63992** seconds. Επίσης για τα μοντέλα **XGB**, τον χαμηλότερο μέσο όρο auc_roc= **0,97507**, ελάχιστα μικρότερο κατά **0,003779 (-0,39%)**, σε

σχέση με το προηγούμενο μοντέλο, σημείωσε το μοντέλο XGB που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-10**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **17,60409** seconds, που είναι πολύ υψηλότερος κατά **10,96417** seconds (**-165,13%**), σε σχέση με το μοντέλο που σημείωσε το μέγιστο μέσο όρο auc_roc.

Από τα **3** μοντέλα **RF** τον υψηλότερο μέσο όρο (mean) auc_roc= **0,97708** σημείωσε το μοντέλο **RF** που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **29,84745** seconds. Το μοντέλο **RF** με τον χαμηλότερο μέσο όρο auc_roc= **0,97525**, ελάχιστα μικρότερο κατά **0,00183** (**-0,19%**), σε σχέση με το προηγούμενο **RF** μοντέλο, σημείωσε το μοντέλο που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **30,22087** seconds, που είναι ελάχιστα υψηλότερος κατά **0,37342** seconds (**-1,25%**), σε σχέση με το προηγούμενο **RF** μοντέλο.

Τελευταία 20 μοντέλα

Εδώ πάλι τα τελευταία **20** σε κατάταξη μοντέλα, περιλαμβάνουν μόνο μοντέλα **DTs**. Τον υψηλότερο μέσο όρο (mean) auc_roc= **0,92714** σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1**, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,18687** seconds. Τον χαμηλότερο μέσο όρο auc_roc = **0,90031**, μικρότερο κατά **0,02683** (**-2,89%**), σε σχέση με το προηγούμενο μοντέλο DT, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Standard** scaler, με τις **default** τιμές των υπερπαραμέτρων και με μέσο όρο διάρκειας της κάθε επανάληψης τα **1,61526** seconds, που είναι υψηλότερος κατά **0,42839** seconds (**+36,09%**), σε σχέση με το προηγούμενο μοντέλο DT. Στις παρακάτω εικόνες απεικονίζονται τα κορυφαία 20 (Εικόνα 115) και τα τελευταία 20 (Εικόνα 116) μοντέλα για την AUC_ROC.

1	xbc_fc_mm_0-1000_22_opt-01
2	xbc_fc_mm_0-100_22_opt-01
3	xbc_fc_mm_0-10_22_opt-01
4	xbc_fc_mm_0-1_22_opt-01
5	xbc_fc_mm_0-100_22_default
6	xbc_fc_mm_0-10_22_default
7	xbc_fc_mm_0-1000_22_default
8	xbc_fc_mm_0-1_22_default
9	xbc_fc_none_22_opt-01
10	xbc_fc_none_22_default
11	rfc_fc_mm_0-1000_22_opt-02
12	xbc_fc_std_22_default
13	xbc_fc_std_22_opt-01
14	rfc_fc_mm_0-100_22_opt-02
15	xbc_fc_none_22_opt-02
16	xbc_fc_mm_0-1000_22_opt-02
17	rfc_fc_none_22_opt-02
18	xbc_fc_std_22_opt-02
19	xbc_fc_mm_0-1_22_opt-02
20	xbc_fc_mm_0-10_22_opt-02

Εικόνα 115: Τα κορυφαία 20 μοντέλα για την AUC_ROC

305	dtc_fc_mm_0-1_10_default
306	dtc_fc_mm_0-1_15_opt-02
307	dtc_fc_std_10_opt-02
308	dtc_fc_std_10_default
309	dtc_fc_mm_0-10_22_opt-02
310	dtc_fc_std_15_opt-02
311	dtc_fc_mm_0-1_22_opt-02
312	dtc_fc_none_15_default
313	dtc_fc_mm_0-1000_15_default
314	dtc_fc_mm_0-100_15_default
315	dtc_fc_mm_0-10_15_default
316	dtc_fc_mm_0-1_15_default
317	dtc_fc_std_22_opt-02
318	dtc_fc_std_15_default
319	dtc_fc_mm_0-1000_22_default
320	dtc_fc_none_22_default
321	dtc_fc_mm_0-100_22_default
322	dtc_fc_mm_0-10_22_default
323	dtc_fc_mm_0-1_22_default
324	dtc_fc_std_22_default

Εικόνα 116: Τα τελευταία 20 μοντέλα για την AUC_ROC

Διάρκεια (Runtime)

Τα μοντέλα **DTs** είναι εκείνα τα οποία εμφανίζουν τη χαμηλότερες τιμές mean για την Runtime, με αυτές να κυμαίνονται από **1,09218** seconds στην **1^η** θέση (precision= **0,91717**), έως **1,72228** στην **62^η** θέση (precision = **0,90466** seconds), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει διασπορά. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** χαρακτηριστικά, με scaler είτε τον **Min-Max** για όλα τα εύρη είτε τον **Standard** scaler και τις **optimal _02** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα μοντέλα των **LR**, με τις τιμές mean για την Runtime να κυμαίνονται από **1,3429** seconds στην **37^η** θέση (precision = **0,92448**), έως **1,97369** seconds στην **108^η** θέση (precision = **0,92476**), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει διασπορά, με το **92,6%** αυτών (50/54) να εντοπίζεται από την **58^η** έως και την **108^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **15** και τα **10** χαρακτηριστικά, είτε με τον **Min-Max** scaler με εύρη **0-100** και **0-10** είτε χωρίς scaler (**no scaler**) και τις **optimal _01** τιμές υπερπαραμέτρων.

Στην τρίτη θέση ακολουθούν τα μοντέλα των **XGB**, με τις τιμές mean για την Runtime να κυμαίνονται από **4,96984** seconds στην **109^η** θέση (precision = **0,93318**), έως **17,71116** seconds στην **144^η** θέση (precision = **0,93302**). Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** και τα **15** χαρακτηριστικά, **όλους** τους διαφορετικούς των scalers(και το no scaler) και τις **optimal _01** τιμές υπερπαραμέτρων.

Έπειτα ακολουθούν τα **RF** μοντέλα, με τις τιμές mean για την Runtime να κυμαίνονται από **11,95152** seconds στην **153^η** θέση (precision = **0,9278**), έως **30,22087** seconds στην **239^η** θέση (precision = **0,93518**), με το **77,8%** αυτών (42/54) να εντοπίζεται από την **183^η** έως και την **239^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα όλα τα εύρη είτε τον **Standard** scaler και τις **default** τιμές υπερπαραμέτρων.

Στη συνέχεια ακολουθούν τα μοντέλα των **MLPs** με τις τιμές mean για την Runtime να κυμαίνονται από **18,14386** seconds στην **198^η** θέση (precision = **0,93179**), έως **362,46571** seconds στην **310^η** θέση (precision = **0,93377**), με το **90,7%** αυτών (49/54) να εντοπίζεται από την **246^η** έως και την **310^η** θέση. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** και τα **15** χαρακτηριστικά, με scaler είτε τον **Min-Max** για τα όλα τα εύρη είτε τον **Standard** scaler και τις **default** τιμές υπερπαραμέτρων.

Τέλος τα μοντέλα των **KNN** είναι εκείνα τα οποία εμφανίζουν τη υψηλότερες τιμές mean για την Runtime, με αυτές να κυμαίνονται από **8,06013** seconds στην **145^η** θέση (precision = **0,92047**), έως **910,1731** seconds στην **324^η** θέση (precision = **0,92091**), η κατανομή των θέσεων των μοντέλων δεν εμφανίζει μεγάλη διασπορά. Στις πρώτες θέσεις, εμφανίζονται τα μοντέλα που χρησιμοποίησαν τα **10** και τα **15** χαρακτηριστικά, είτε με τον **Min-Max** scaler με εύρη **0-1000**, **0-100** και **0-10** είτε χωρίς scaler (**no scaler**) και τις **default** και **optimal _01** τιμές υπερπαραμέτρων.

Κορυφαία 20 μοντέλα

Εδώ τα κορυφαία **20** σε κατάταξη μοντέλα, περιλαμβάνουν μόνο μοντέλα **DTs**. Τον χαμηλότερο μέσο όρο (mean) runtime= **1,09218** seconds σημείωσε το μοντέλο DT που χρησιμοποίησε τα **10** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1**, με τις **optimal _02** τιμές των υπερπαραμέτρων και με μέσο όρο precision= **0,91717**. Τον υψηλότερο μέσο όρο runtime= **1,20965** seconds, μικρότερο κατά **0,11747** seconds (**-10,75%**), σε σχέση με το προηγούμενο μοντέλο DT, σημείωσε το μοντέλο DT που χρησιμοποίησε τα **15** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal _02** τιμές των υπερπαραμέτρων

και με μέσο όρο precision= **0,92137**, που είναι ελάχιστα χαμηλότερος κατά **0,0042 (-0,46%)**, σε σχέση με το μοντέλο DT, με την χαμηλότερη runtime.

Τελευταία 20 μοντέλα

Τα τελευταία **20** σε κατάταξη μοντέλα περιλαμβάνουν **16** μοντέλα **KNN** και **4** μοντέλα **MLPs**. Από τα **4** μοντέλα **MLPs**, τον χαμηλότερο μέσο όρο (mean) runtime= **267,55604** seconds, σημείωσε αυτό που χρησιμοποίησε τα **15** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,93394**. Επίσης για τα μοντέλα **MLPs**, τον υψηλότερο μέσο όρο runtime= **362,46571** seconds, αρκετά υψηλότερο κατά **94,90967** seconds (**+35,47%**), σε σχέση με το προηγούμενο μοντέλο, σημείωσε το μοντέλο MLPs που χρησιμοποίησε τα **22** χαρακτηριστικά, χωρίς scaler (**no_scaler**), με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,93377**, που είναι ελάχιστα χαμηλότερος κατά **0,00017 (-0,018%)**, σε σχέση με το μοντέλο **MLPs** που σημείωσε τον μικρότερο μέσο όρο runtime.

Από τα **16** μοντέλα **KNN**, τον χαμηλότερο μέσο όρο (mean) runtime= **281,23095** seconds, σημείωσε αυτό που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-10**, με τις **optimal_02** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,9201**. Επίσης τον υψηλότερο μέσο όρο runtime= **910,1731** seconds, αρκετά υψηλότερο κατά **628,94215** seconds (**+223,64%**), σε σχέση με το προηγούμενο μοντέλο, σημείωσε το μοντέλο KNN που χρησιμοποίησε τα **22** χαρακτηριστικά, τον **Min-Max** scaler με εύρος **0-1000**, με τις **optimal_01** τιμές υπερπαραμέτρων και με μέσο όρο precision= **0,92091**, που είναι ελάχιστα χαμηλότερος κατά **0,00081 (-0,088%)**, σε σχέση με το μοντέλο **KNN** που σημείωσε τον μικρότερο μέσο όρο runtime. Στις παρακάτω εικόνες απεικονίζονται τα 20 ταχύτερα (Εικόνα 117) και τα 20 πιο αργά (Εικόνα 118) μοντέλα για την Runtime.

1	dtc_fc_mm_0-1_10_opt-02
2	dtc_fc_std_10_opt-02
3	dtc_fc_mm_0-10_10_opt-02
4	dtc_fc_mm_0-100_10_opt-02
5	dtc_fc_mm_0-1000_10_opt-02
6	dtc_fc_none_10_opt-02
7	dtc_fc_mm_0-1000_10_opt-01
8	dtc_fc_std_10_opt-01
9	dtc_fc_mm_0-100_10_opt-01
10	dtc_fc_mm_0-10_10_opt-01
11	dtc_fc_std_15_opt-02
12	dtc_fc_none_10_opt-01
13	dtc_fc_mm_0-1_10_opt-01
14	dtc_fc_std_10_default
15	dtc_fc_mm_0-1_10_default
16	dtc_fc_mm_0-1000_10_default
17	dtc_fc_mm_0-10_10_default
18	dtc_fc_mm_0-100_10_default
19	dtc_fc_none_10_default
20	dtc_fc_none_15_opt-02

Εικόνα 117: Τα κορυφαία 20 μοντέλα για την Runtime

305	mlp_fc_none_15_opt-02
306	mlp_fc_none_22_opt-02
307	knc_fc_mm_0-10_22_opt-02
308	mlp_fc_none_15_opt-01
309	knc_fc_mm_0-1_22_opt-02
310	mlp_fc_none_22_opt-01
311	knc_fc_std_15_opt-02
312	knc_fc_std_22_opt-02
313	knc_fc_none_22_default
314	knc_fc_std_22_default
315	knc_fc_mm_0-1000_22_default
316	knc_fc_mm_0-100_22_default
317	knc_fc_mm_0-10_22_default
318	knc_fc_mm_0-1_22_default
319	knc_fc_none_22_opt-01
320	knc_fc_std_22_opt-01
321	knc_fc_mm_0-1_22_opt-01
322	knc_fc_mm_0-10_22_opt-01
323	knc_fc_mm_0-100_22_opt-01
324	knc_fc_mm_0-1000_22_opt-01

Εικόνα 118: Τα τελευταία 20 μοντέλα για την Runtime

Στους παρακάτω πίνακες απεικονίζονται τα συγκεντρωτικά αποτελέσματα των μετρικών, όλων των μοντέλων (**Πίνακας 42**) και των βέλτιστων μοντέλων (**Πίνακας 43**) των 6 αλγορίθμων.

Πίνακας 48: Συγκεντρωτικά αποτελέσματα απόδοσης των όλων των μοντέλων των 6 αλγορίθμων

Αλγόριθμος	Precision	Recall	F1-score	AU_ROC	Runtime(sec)
LR	0,92322 - 0,92626	0,90785 - 0,91899	0,88879 - 0,89262	0,96459 - 0,97081	1,3429 - 3,31659
DTs	0,90092 - 0,93035	0,8342 - 0,9369	0,84959 - 0,90032	0,90031 - 0,95666	1,09218 - 1,72228
RF	0,92253 - 0,93661	0,90832 - 0,9699	0,8873 - 0,91127	0,96155 - 0,97708	11,95152 - 30,22087
XGB	0,9321 - 0,93764	0,93958 - 0,9561	0,90332 - 0,9113	0,97073 - 0,978849	4,96984 - 17,71116
MLPs	0,92598 - 0,93456	0,93434 - 0,95623	0,89386 - 0,90756	0,97064 - 0,97352	18,14386 - 362,46571
KNN	0,91542 - 0,92854	0,88231 - 0,92947	0,87515 - 0,89737	0,94146 - 0,9593	8,06013 - 910,1731

Πίνακας 49: Αποτελέσματα απόδοσης των βέλτιστων μοντέλων των 6 αλγορίθμων

Αλγόριθμος	Precision	Recall	F1-score	AU_ROC	Runtime(sec)
LR	0,92562 - 0,92626	0,90994 - 0,91287	0,89136 - 0,89262	0,97041 - 0,97081	2,99462 - 3,23525
DTs	0,93035 - 0,92984	0,93372 - 0,9369	0,89936 - 0,90032	0,95488 - 0,95666	1,14701 - 1,45128
RF	0,93551 - 0,93661	0,96631 - 0,9699	0,90963 - 0,91127	0,97441 - 0,97708	26,67779 - 29,84745
XGB	0,93731 - 0,93764	0,95403 - 0,9561	0,91086 - 0,91116	0,97784 - 0,978849	6,12796 - 6,74094
MLPs	0,93401 - 0,93456	0,95279 - 0,95623	0,90756 - 0,90666	0,97262 - 0,97294	117,22776 - 181,84466
KNN	0,92707 - 0,92854	0,92555 - 0,92947	0,89503 - 0,89737	0,95828 - 0,9593	48,21549 - 882,94407

Επεξήγηση Πινάκων 42 και 43 : Με πράσινο σημειώνονται οι υψηλότερες τιμές και με κόκκινο οι χαμηλότερες για όλα τα μετρικά εκτός της Runtime όπου είναι αντίστροφα.

Συζήτηση

Στην παρούσα εργασία μελετήθηκε η δυνατότητα πρόβλεψης της θνησιμότητας ασθενών με COVID-19, με βάση έναν αριθμό από 22 χαρακτηριστικά κυρίως κατηγορικά, με μόνο 2 από αυτά να είναι συνεχή. Στο σύνολο των δεδομένων που χρησιμοποιήθηκε, είχε προηγηθεί προεπεξεργασία που διέφερε μόνο στον μετασχηματισμό των 2 συνεχών δεδομένων στα οποία εφαρμόστηκαν 6 διαφορετικές μέθοδοι.

Για την δημιουργία των μοντέλων χρησιμοποιήθηκαν 6 αλγόριθμοι, αυτοί της Λογιστικής Παλινδρόμησης (LR), των Δέντρων Απόφασης (DTs), του Τυχαίου Δάσους (RF), της Ακραίας Ενίσχυσης Κλίσης (XGB), των TNΔ Πολυστρωματικών Αντιλήπτρων (MLPs) και των K Εγγύτερων Γειτόνων (KNN). Σε αυτούς χρησιμοποιήθηκαν 3 διαφορετικά σετ υπερπαραμέτρων, default-optimal_01, optimal_02, με τα 2 τελευταία σετ να είναι αυτά τα οποία προέκυψαν από την εφαρμογή της μεθόδου της αναζήτησης πλέγματος (GridSearchCV).

Η δοκιμή των μοντέλων των αλγορίθμων έγινε σε 3 διαφορετικά σετ χαρακτηριστικών. Αρχικά με χρήση όλων των 22 χαρακτηριστικών, χρήση μόνο των 15 σημαντικότερων και τέλος χρήση μόνο των 10 σημαντικότερων. Η σημαντικότητα των χαρακτηριστικών υπολογίστηκε ξεχωριστά για τον κάθε αλγόριθμο, με την χρήση των αντίστοιχων python functions.

Έτσι προέκυψαν 54 (6 τρόποι προεπεξεργασίας x 3 σετ χαρακτηριστικών x 3 σετ υπερπαραμέτρων) διαφορετικοί τρόποι εκτέλεσης ανά αλγόριθμο με τον κάθε τρόπο να εκτελείται 10 φόρες ώστε να ληφθεί η μέση τιμή (mean) των μετρικών του, καταλήγοντας έτσι στις 540 εκτελέσεις για κάθε αλγόριθμο, με το συνολικό αριθμό τους να φθάνει τις 3.240 και για τους 6 αλγορίθμους.

Κατόπιν της περάτωσης των εκτελέσεων, προχωρήσαμε στην αξιολόγηση τους με βάση 5 μετρικά την Ακρίβεια (Precision), την Ευαισθησία-Ανάκληση (Recall), την Βαθμολογία F1 (F1 Score), την Περιοχή κάτω από την καμπύλη ROC (Area Under Curve ROC, AUC_ROC) και την Διάρκεια (Runtime).

Λογιστική Παλινδρόμηση (LR)

Η Λογιστική Παλινδρόμηση βρίσκει εφαρμογή κυρίως στις στατιστικές και κοινωνικές επιστήμες, σε αρκετά ιατρικά πεδία και στον τομέα της Μηχανικής Μάθησης μέσω της σύνθεσης μοντέλων παλινδρόμησης. Βασικός τους στόχος της LR είναι η πρόβλεψη για το αν ένα δείγμα θα ανήκει ή όχι σε κάποια κατηγορία.

Τα μοντέλα που δημιουργήθηκαν με τη χρήση της **LR** ήρθαν στην τελευταία (**6^η**) θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,92322 - 0,92626**, Κατανομή Θέσεων: **169^η - 267^η**), της Ευαισθησίας-Ανάκλησης (Recall: **0,90785 - 0,91899**, Κατανομή Θέσεων: **182^η - 278^η**), του F1-score (**0,88879 - 0,89262**, Κατανομή Θέσεων: **172^η - 266^η**), στην **4^η** θέση με βάση την AUC_ROC (**0,96459 - 0,97081**, Κατανομή Θέσεων: **_119^η - 203^η**) και στην **2^η** θέση με βάση την Διάρκεια (Runtime: **1,3429 - 3,31659 sec**, Κατανομή Θέσεων: **37^η - 108^η**).

Τα LR μοντέλα της παρούσας εργασίας, εμφάνισαν παραπλήσια απόδοση με αυτά παρόμοιων μελετών, που και αυτές είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, σε dataset παρόμοιας σύστασης. Πιο συγκεκριμένα η μελέτη του **BO Josephus** και των συνεργατών του [496], με σύνολο δεδομένων που αποτελείτο από **1.085** ασθενείς με COVID-19, με **25** στοιχεία για τον καθένα. Μετά από την καθαρισμό και την προεπεξεργασία

των δεδομένων το LR μοντέλο πέτυχε Ακρίβεια (Precision) **0,86** (86%), F1-score **0,92** και AUC_ROC **1,00**. Επίσης στη μελέτη των **Pourhomayoun & Shakibi [497]**, με σύνολο δεδομένων το οποίο περισσότερους από **2.670.000** ασθενείς με COVID-19, με **32** στοιχεία δεδομένων για τον κάθε ασθενή, συμπεριλαμβανομένων δημογραφικών και φυσιολογικών δεδομένων, αφού αφαιρέθηκαν τα χαρακτηριστικά τα οποία δεν συνδέονται με τον κίνδυνο θανάτου από COVID-19 και χρησιμοποιήθηκε επιλογή χαρακτηριστικών (feature selection), τα LR μοντέλα πέτυχαν Ακρίβεια (Precision) **0,8791** (87,91%) και AUC_ROC **0,92**.

Δέντρα Απόφασης (DTs)

Τα Δέντρα Αποφάσεων, είναι μια μη παραμετρική εποπτευόμενη μέθοδος μηχανικής μάθησης που εφαρμόζεται κυρίως σε προβλήματα ταξινόμησης ή παλινδρόμησης. Κύριος στόχος των DTs είναι η δημιουργία μοντέλων που να προβλέπουν την τιμή μιας μεταβλητής στόχου, με τη χρήση απλών κανόνων απόφασης που συνάγονται από τα χαρακτηριστικά του εκάστοτε συνόλου δεδομένων.

Τα μοντέλα που δημιουργήθηκαν με τη χρήση των **DTs** ήρθαν στην **4^η** θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,90092 - 0,93035**, Κατανομή θέσεων: **127^η - 324^η**), της Ευαισθησίας-Ανάκλησης (Recall: **0,8342 - 0,9369**, Κατανομή θέσεων: **135^η - 324^η**), του F1-score (**0,84959 - 0,90032**, Κατανομή θέσεων: **127^η - 324^η**), στην τελευταία (**6^η**) θέση με βάση την AUC_ROC (**0,90031 - 0,95666**, Κατανομή θέσεων: **234^η - 324^η**) και στην **1^η** θέση με βάση την Διάρκεια (Runtime: **1,09218 - 1,72228 sec**, Κατανομή θέσεων: **1^η - 62^η**).

Τα μοντέλα των DTs της εργασίας αυτής εμφάνισαν παραπλήσια απόδοση με άλλα παρόμοιων μελετών, που και αυτές είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, σε dataset παρόμοιας σύστασης. Συγκεκριμένα στη μελέτη της **Maleeha Naseem** και των συνεργατών της [498], με σύνολο δεδομένων που αποτελείτο από **1.214** ασθενείς με COVID-19 και περιελάμβανε τα κλινικά στοιχεία για του καθένα. Μετά από την καθαρισμό και την προεπεξεργασία των δεδομένων το DT μοντέλο πέτυχε Ακρίβεια (Precision) **0,8521** (85,21%) και AUC_ROC **0,6674**. Επίσης στη μελέτη των **Pourhomayoun & Shakibi [497]**, που έγινε αναφορά και πριν, μετά τον καθαρισμό, την προεπεξεργασία των δεδομένων και την επιλογή χαρακτηριστικών (feature selection), τα DTs μοντέλα πέτυχαν Ακρίβεια (Precision) **0,8687** (86,87%) και AUC_ROC **0,93**.

Τυχαίο Δάσος (RF)

Το Τυχαίο Δάσος αποτελεί ακόμα μια μέθοδο μηχανικής μάθησης που δύναται να εφαρμοστεί σε προβλήματα ταξινόμησης και παλινδρόμησης. Το RF έχει ως βασική αρχή την μάθηση συνόλου (Ensemble), χρησιμοποιώντας έναν αριθμό δέντρων αποφάσεων, με το καθένα να έχει εκπαίδευτεί σε διαφορετικό υποσύνολο του συνόλου εκπαίδευσης και με την τελική απόφαση (πρόβλεψη) να λαμβάνεται είτε από την πλειοψηφία, για προβλέψεις σε κατηγορικές μεταβλητές είτε από το μέσο όρο των τιμών για προβλέψεις σε αριθμητικές μεταβλητές.

Τα μοντέλα που δημιουργήθηκαν με τη χρήση του **RF** ήρθαν στην **2^η** θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,92253 - 0,93661**, Κατανομή θέσεων: **11^η - 273^η**), του F1-score (**0,8873 - 0,91127**, Κατανομή θέσεων: **2^η - 274^η**), της AUC_ROC (**0,96155 - 0,97708**, Κατανομή θέσεων: **11^η - 216^η**), στην **1^η** θέση με βάση την Ευαισθησία-Ανάκληση

(Recall: **0,90832 - 0,9699**, Κατανομή θέσεων: **1^η - 271^η**) και στην **5^η** θέση με βάση την Διάρκεια (Runtime: **11,95152 - 30,22087 sec**, Κατανομή θέσεων: **153^η - 239^η**).

Τα RF μοντέλα της παρούσας εργασίας εμφάνισαν παραπλήσια απόδοση με άλλα παρόμοιων μελετών, που και αυτές είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, σε dataset τα οποία είχαν παρόμοια δομή δεδομένων. Πιο συγκεκριμένα στην μελέτη των **Roberto Bárcenas & Ruth Fuentes-García [499]**, με σύνολο δεδομένων που αποτελείτο από **220.657** ασθενείς με COVID-19, και περιελάμβανε τα κλινικά και γεωγραφικά στοιχεία του καθένα, μετά από την καθαρισμό και την προεπεξεργασία των δεδομένων το RF μοντέλο σημείωσε Ακρίβεια (Precision) **0,9824** (98,24%) για τους ασθενείς Χαμηλού ρίσκου, **0,6737** (67,37%) για τους ασθενείς Μεσαίου ρίσκου και **0,7232** (72,32%) για τους ασθενείς Υψηλού ρίσκου και F1-score **0,9908** για τους ασθενείς Χαμηλού ρίσκου, **0,5184** για τους ασθενείς Μεσαίου ρίσκου και **0,7731** για τους ασθενείς Υψηλού ρίσκου. Επίσης στη μελέτη της **Maleeha Naseem** και των συνεργατών της [498], που έγινε αναφορά στα DTs αφού ολοκληρώθηκε ο καθαρισμός και η προεπεξεργασία των δεδομένων το μοντέλο RF πέτυχε Ακρίβεια (Precision) **0,8575** (85,75%) και AUC_ROC **0,6950**. Ακόμα στη μελέτη των **Pourhomayoun & Shakibi [453]**, που αναφερθήκαμε και πριν, μετά τον καθαρισμό, την προεπεξεργασία και την επιλογή χαρακτηριστικών, τα RF μοντέλα πέτυχαν Ακρίβεια (Precision) **0,8793** (87,93%) και AUC_ROC **0,94**.

Ακραία Ενίσχυση Κλίσης (XGB)

Η μέθοδος της Ακραίας ενίσχυσης Κλίσης αποτελεί μια επιπλέον τεχνική μηχανικής μάθησης, που μπορεί να βρει εφαρμογή σε προβλήματα ταξινόμησης και παλινδρόμησης. Είναι και αυτή μέθοδος μάθησης συνόλου (Ensemble). Στην XGB δημιουργούνται μεμονωμένα DTs με την χρήση πολλαπλών πυρήνων, παράλληλα με την οργάνωση των δεδομένων, έτσι ώστε να μειωθούν στο ελάχιστο οι χρόνοι αναζήτησης με σκοπό τη μείωση του χρόνου εκπαίδευσης των μοντέλων και την αύξηση της απόδοσης [483]. Εδώ η βελτίωση της ακρίβειας των προβλέψεων επιτυγχάνεται με την χρήση τεχνικών βελτιστοποίησης του εκάστοτε συστήματος, καθώς και τεχνικών βελτιστοποίησης των αλγορίθμικών λειτουργειών. Η XGB αποτελεί την πιο διαδεδομένη μέθοδο Μηχανικής Μάθησης (ML), έχοντας αποκομίσει μεγάλη δημοτικότητα σε διαγωνισμούς ML [384].

Τα μοντέλα που δημιουργήθηκαν με τη χρήση της **XGB** ήρθαν στην **1^η** θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,9321 - 0,93764**), Κατανομή θέσεων: **1^η - 113^η**), του F1-score (**0,90332 - 0,9113**, Κατανομή θέσεων: **1^η - 121^η**), της AUC_ROC (**0,97073 - 0,978849**, Κατανομή θέσεων: **1^η - 122^η**), στην **3^η** θέση με βάση την Ευαισθησία-Ανάκληση (Recall: **0,93958 - 0,9561**, Κατανομή θέσεων: **20^η - 133^η**) και την Διάρκεια (Runtime: **4,96984 - 17,71116 sec**, Κατανομή θέσεων: **109^η - 144^η**).

Τα XGB μοντέλα της εργασίας αυτής εμφάνισαν παραπλήσια απόδοση με αυτά παρόμοιων μελετών, που και αυτές είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, σε dataset με παρόμοια δομή δεδομένων. Πιο συγκεκριμένα στην μελέτη των **Li Yan** και των συνεργατών του [500] με σύνολο δεδομένων που αποτελείτο από **375** ασθενείς με COVID-19, και περιελάμβανε τα κλινικά και στοιχεία του καθένα. Στην μελέτη αυτή έπειτα από την καθαρισμό και την προεπεξεργασία των δεδομένων και με την χρήση 3 χαρακτηριστικών, των επιπέδων της γαλακτικής αφυδρογονάσης (LDH), των αριθμό των λεμφοκυττάρων και τον προσδιορισμό της υψηλής ευαισθησίας της C-αντιδρώσας πρωτεΐνης (hsCRP), το XGB μοντέλο σημείωσε Ακρίβεια (Precision) **1,000** (100%), Ανάκληση (Recall) **0,8333** (83,33%) και F1-score **0,9090**. Επιπλέον στη μελέτη των **Roberto Bárcenas & Ruth Fuentes-García [499]**, που έγινε αναφορά και προηγουμένως, ύστερα από την καθαρισμό και την προεπεξεργασία των δεδομένων το XGB μοντέλο σημείωσε Ακρίβεια (Precision) **0,9822**

(98,22%) για τους ασθενείς Χαμηλού ρίσκου, **0,7075** (70,75%) για τους ασθενείς Μεσαίου ρίσκου και **0,6842** (68,42) για τους ασθενείς Υψηλού ρίσκου και F1-score **0,9909** για τους ασθενείς Χαμηλού ρίσκου, **0,3742** για τους ασθενείς Μεσαίου ρίσκου και **0,7713** για τους ασθενείς Υψηλού ρίσκου.

ΤΝΔ Πολυστρωματικών Αντιλήπτρων (MLPs)

Τα ΤΝΔ Πολυστρωματικών Αντιλήπτρων είναι και αυτά αλγόριθμοι εποπτευόμενης μάθησης, οι οποίοι μιμούμενοι την λειτουργία των νευρώνων, εκπαιδεύονται σε ένα σύνολο δεδομένων, «μαθαίνοντας» μια συνάρτηση με βάση τη μεταβλητή στόχο με σκοπό την επίλυση σύνθετων υπολογιστικών προβλημάτων [373,374]. Βασική διαφορά τους από την LR είναι το γεγονός πως μεταξύ εισόδου και εξόδου δύνανται να παρεμβάλλονται ένα ή περισσότερα μη γραμμικά (κρυφά) επίπεδα [484].

Τα μοντέλα που δημιουργήθηκαν με τη χρήση της **MLPs** ήρθαν στην **3^η** θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,92598 - 0,93456**, Κατανομή θέσεων: **32^η - 172^η**), του F1-score (**0,89386 - 0,90756**, Κατανομή θέσεων: **34^η - 168^η**), της AUC_ROC (**0,97064 - 0,97352**, Κατανομή θέσεων: **39^η - 123^η**), στην **2^η** θέση με βάση την Ευαισθησία-Ανάκληση (Recall: **0,93434 - 0,95623**, Κατανομή θέσεων: **19^η - 142^η**) και στην τελευταία (**6^η** θέση με βάση τη Διάρκεια (Runtime: **18,14386 - 362,46571 sec**, Κατανομή θέσεων: **198^η - 310^η**)

Τα MLPs μοντέλα της παρούσας εργασίας εμφάνισαν παραπλήσια απόδοση με άλλα μοντέλα ANN (Artificial Neural Networks) από παρόμοιες μελέτες, που είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, χρησιμοποιώντας dataset τα οποία είχαν παρόμοιο τύπο και δομή δεδομένων. Πιο συγκεκριμένα στη μελέτη της **Maleeha Naseem** και των συνεργατών της [498], που έγινε αναφορά προηγουμένως, αφού ολοκληρώθηκε ο καθαρισμός και η προεπεξεργασία των δεδομένων το ANN (Deep-Neo-V) μοντέλο πέτυχε Ακρίβεια (Precision) **0,8767** (87,67%) και AUC_ROC **0,8850**. Επιπλέον στη μελέτη των **Pourhomayoun & Shakibi** [497], μετά τον καθαρισμό, την προεπεξεργασία και την επιλογή χαρακτηριστικών, το ANN μοντέλο πέτυχε Ακρίβεια (Precision) **0,8998** (89,98%) και AUC_ROC **0,93**.

Κ Εγγύτεροι Γείτονες (KNN)

Ο ταξινομητής των **Κ Εγγύτερων Γείτονων** είναι μια μη παραμετρική μέθοδος επιβλεπόμενης μάθησης που χρησιμοποιείται τόσο σε προβλήματα Κατηγοριοποίησης, για διακριτές μεταβλητές στόχους, όσο και σε προβλήματα Παλινδρόμησης, για συνεχείς μεταβλητές στόχους. Ο KNN εφαρμόζει την μέθοδο της ψηφοφορία k-πλησιέστερων γειτόνων, στηριζόμενος στην απόσταση του κάθε νέου δείγματος, από τις τιμές των γνωστών δειγμάτων, ώστε να μπορέσει να το ταξινομήσει.

Τα μοντέλα που δημιουργήθηκαν με τη χρήση της **KNN** ήρθαν στην **5^η** θέση σε κατάταξη, με βάση τα μετρικά της Ακρίβειας (Precision: **0,91542 - 0,92854**, Κατανομή θέσεων: **142^η - 304^η**), της Ευαισθησίας-Ανάκλησης (Recall: **0,88231 - 0,92947**, Κατανομή θέσεων: **157^η - 303^η**) του F1-score (**0,87515 - 0,89737**, Κατανομή θέσεων: **145^η - 304^η**), της AUC_ROC (**0,94146 - 0,9593**, Κατανομή θέσεων: **217^η - 289^η**) και στην **4^η** θέση με βάση την Διάρκεια (Runtime: **8,06013 - 910,1731 sec**, Κατανομή θέσεων: **145^η - 324^η**)

Τα KNN μοντέλα της εργασίας αυτής εμφάνισαν παραπλήσια απόδοση με αυτά παρόμοιων μελετών, που και αυτές είχαν ως αντικείμενο, την πρόβλεψη της θνησιμότητας, με τη χρήση

dataset που είχαν παρόμοιο τύπο και δομή δεδομένων. Συγκεκριμένα στη μελέτη της **Maleeha Naseem** και των συνεργατών της [498], αφού ολοκληρώθηκε ο καθαρισμός και η προεπεξεργασία των δεδομένων το KNN μοντέλο πέτυχε Ακρίβεια (Precision) **0,7781** (77,81%) και AUC_ROC **0,6650**. Ακόμα στη μελέτη των **Pourhomayoun & Shakibi** [497], μετά τον καθαρισμό, την προεπεξεργασία και την επιλογή χαρακτηριστικών, το ANN μοντέλο πέτυχε Ακρίβεια (Precision) **0,8983** (89,83%) και AUC_ROC **0,90**.

Συμπεράσματα

Σύμφωνα με τα αποτελέσματα της παρούσας εργασίας τα μοντέλα σημείωσαν Ακρίβεια από **0,90092** (90,09%) έως **0,93764** (93,76%), Ανάκληση από **0,8342** (83,42%) έως **0,9699** (96,99%), F1-score από **0,84959** έως **0,9113**, AUC_ROC από **0,90031** έως **0,978849** και Διάρκεια από **1,09218** έως **910,1731** seconds. Τα βέλτιστα μοντέλα εμφάνισαν Ακρίβεια από **0,92562** (92,56%), Ανάκληση από **0,90994** (90,99%), F1-score από **0,89136**, AUC_ROC από **0,95488** και Διάρκεια από **1,14701** έως **882,94407** seconds.

Η **Ακραία Ενίσχυσης Κλίσης** (XGB) κατατάσσεται **1^η** με βάση την απόδοση όλων των μοντέλων της καθώς έχει πέτυχει τις υψηλότερες τιμές για τα μετρικά της Ακρίβειας, του F1-score και της AUC_ROC και τις 3^{ες} υψηλότερες τιμές για το μετρικό της Ανάκλησης. Στην **2^η** θέση έχουμε τα μοντέλα του **Τυχαίου Δάσους** (RF) που σημείωσαν τις 2^{ες} υψηλότερες τιμές στα μετρικά ης Ακρίβειας, του F1-score και της AUC_ROC και τις υψηλότερες τιμές για το μετρικό της Ανάκλησης. Έπειτα στην **3^η** θέση έχουμε τα μοντέλα των **Τεχνητών Νευρωνικών Δικτύων Πολυστρωματικών Αντιλήπτρων** (MLPs) που σημείωσαν τις 3^{ες} υψηλότερες τιμές στα μετρικά ης Ακρίβειας, του F1-score και της AUC_ROC και τις 2^{ες} υψηλότερες τιμές για το μετρικό της Ανάκλησης άλλα έχουν τις 2^{ες} πιο υψηλές τιμές Διάρκειας. Στη συνέχεια στην **4^η** θέση έχουμε τα μοντέλα των **Δέντρων Αποφάσεως** (DTs), που εμφάνισαν τις 4^{ες} υψηλότερες τιμές για τα μετρικά της Ακρίβειας, της Ανάκλησης και του F1-score, τις χαμηλότερες τιμές για τα μετρικά της AUC_ROC και τις Διάρκειας. Στην **5^η** θέση έρχονται τα μοντέλα των **Κ Εγγύτερων Γειτόνων** (KNN) με τις 2^{ες} χαμηλότερες τιμές για τα μετρικά της Ακρίβειας, της Ανάκλησης, του F1-score και της AUC_ROC και τις υψηλότερες τιμές για την Διάρκεια. Τέλος στην τελευταία (**6^η**) θέση κατατάσσονται τα μοντέλα της **Λογιστικής Παλινδρόμησης** (LR), σημειώνοντας τις χαμηλότερες τιμές για τα μετρικά της Ακρίβειας, της Ανάκλησης και του F1-score και τις 3^{ες} χαμηλότερες τιμές για το μετρικό της AUC_ROC και τις 2^{ες} χαμηλότερες τιμές για τη Διάρκεια.

Όμοια είναι και η κατάταξη των αλγορίθμων, με βάση την απόδοση μόνο των βέλτιστων μοντέλων τους. Με **1^η** τα μοντέλα της **Ακραίας Ενίσχυσης Κλίσης** (XGB), **2^η** αυτά του **Τυχαίο Δάσος** (RF), **3^η** εκείνα των **Τεχνητών Νευρωνικών Δικτύων Πολυστρωματικών Αντιλήπτρων** (MLPs), **4^η** αυτά των **Δέντρα Αποφάσεως** (DTs), **5^η** εκείνα των **Κ Εγγύτερων Γειτόνων** (KNN) και **6^η** αυτά της **Λογιστικής Παλινδρόμησης** (LR).

Από τα βέλτιστα μοντέλα θα μπορούσαν να χρησιμοποιηθούν τα **1^η** σε κατάταξη από τους αλγορίθμους της XGB (**22_Min-Max 0-100_opt_01**), του RF (**22_Min-Max 0-1000_opt_02**) και των MLPs (**22_Min-Max 0-1000_opt_01**) σε μοντέλο Ensemble 3 αλγορίθμων, με λήψη αποφάσεων με πλειοψηφίες 3-0 και 2-1. Σε άλλη περίπτωση θα μπορούσαν να χρησιμοποιηθούν τα 2 καλύτερα μοντέλα της XGB (**22_Min-Max 0-100_opt_01**, **22_Min-Max 0-1000_opt_01**), τα **1^η** του RF (**22_Min-Max 0-1000_opt_02**), των MLPs (**22_Min-Max 0-1000_opt_01**) και των DTs (**22_Min-Max 0-1000_opt_01**) σε μοντέλο Ensemble 5 αλγορίθμων, με λήψη αποφάσεων με πλειοψηφίες 5-0, 4-1, 3-2.

Η εργασία πραγματοποιήθηκε χρησιμοποιώντας τα δεδομένα του Εθνικού Συστήματος Υγείας του Μεξικού, με χαρακτηριστικά κυρίως κατηγορικής φύσεως (δυαδικά), με μόνο 2 από αυτά να είναι αριθμητικά (Ηλικία, Ημέρες από την εμφάνιση Συμπτωμάτων έως την εισαγωγή στο Νοσοκομείο) και με το χαρακτηριστικό στόχο να είναι και αυτό κατηγορικό (Επιβίωση [Ναι, Όχι]). Έτσι οι 6 αλγόριθμοι, που μελετήθηκαν εδώ, θα εμφάνιζαν διαφορές στην ακρίβεια των προβλέψεων, συγκριτικά με την ακρίβεια τους στην παρούσα εργασία, σε σύνολα δεδομένων διαφορετικής σύστασης, τα οποία θα περιείχαν περισσότερα αριθμητικά και συνεχή χαρακτηριστικά (ανεξάρτητες μεταβλητές) και χαρακτηριστικό-ά στόχο (εξαρτημένη-ες μεταβλήτη-ες), το-α οποίο-α και αυτό-α να είναι αριθμητικό-ά. Επίσης καθώς τα δεδομένα έχουν ληφθεί στο Μεξικό η απόδοση των αλγορίθμων θα εμφάνιζε κάποια

απόκλιση, σε δεδομένα από άλλες χώρες λόγω των διαφορών σε συνθήκες περίθαλψης, ατομικής υγιεινής κ.α.

Τα αποτελέσματα της παρούσας εργασίας θα μπορούσαν να χρησιμοποιηθούν στον τρόπο αξιολόγησης των δεδομένων ερωτηματολογίων, στα οποία ο κάθε ασθενής με COVID-19 αναφέρει το φύλο, την ηλικία, την ημερομηνία έναρξης των συμπτωμάτων και την ύπαρξη ή όχι κάποιου από τα 14 χαρακτηριστικά, κλινικά ή μη, τα οποία αναφέρθηκαν στον **Πίνακα 23**. Η επεξεργασία των παραπάνω δεδομένων από το βέλτιστο μοντέλο της **XGB** ή από μεθόδους Ensembe 3 ή 5 μοντέλων, όπως αναφέρθηκαν παραπάνω, θα μπορούσαν να δώσουν μια πρόβλεψη υψηλής ακρίβειας (93,76%) για τον κίνδυνο θανάτου του κάθε ασθενή με COVID-19, που σε περιόδους ασφυκτικής πίεσης του Εθνικού Συστήματος Υγείας θα μπορούσε να βιοθήσει, εστιάζοντας στους ασθενείς υψηλού κίνδυνου.

Σε μελλοντικές μελέτες η χρήση κβαντικών υπολογιστών, θα επέτρεπε την εκτέλεση περισσότερων επαναλήψεων (iterations) των μοντέλων στον ίδιο χρόνο, παρέχοντας έτσι πιο γρήγορα αποτελέσματα και για πιο σύνθετα προβλήματα με περισσότερες παραμέτρους.

Βιβλιογραφία

1. Coronavirus disease (COVID-19) [Internet]. WHO; 2021 [cited 2022 Oct 24]; Available from: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19>
2. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>
3. Key Messages and Actions for COVID-19 Prevention and Control in Schools [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: https://www.who.int/docs/default-source/coronavirus/key-messages-and-actions-for-covid-19-prevention-and-control-in-schools-march-2020.pdf?sfvrsn=baf81d52_4
4. How COVID-19 Spreads [Internet]. CDC; 2020 [cited 2022 Oct 24]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>
5. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020 Apr;5(4):536-544. doi: 10.1038/s41564-020-0695-z. Epub 2020 Mar 2. PMID: 32123347; PMCID: PMC7095448.
6. Naming the coronavirus disease (COVID-19) and the virus that causes it [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
7. Coronavirus disease named Covid-19 [Internet]. BBC News; 2020 [cited 2022 Oct 24]; Available from: <https://www.bbc.com/news/world-asia-china-51466362>
8. Surveillance case definitions for human infection with novel coronavirus (nCoV): interim guidance, 11 January 2020 [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: <https://apps.who.int/iris/handle/10665/330376>
9. About Novel Coronavirus (2019-nCoV) [Internet]. CDC; 2020 [cited 2020 Oct 28]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/about/index.html>
10. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, Lloyd-Smith JO, de Wit E, Munster VJ. Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. *medRxiv* [Preprint]. 2020 Mar 13:2020.03.09.20033217. doi: 10.1101/2020.03.09.20033217. Update in: *N Engl J Med.* 2020 Apr 16;382(16):1564-1567. PMID: 32511427; PMCID: PMC7217062.
11. Novel coronavirus named 'Covid-19': WHO [Internet]. TODAY online; 2020 [cited 2022 Oct 24]; Available from: <https://www.todayonline.com/world/wuhan-novel-coronavirus-named-covid-19-who>
12. The coronavirus spreads racism against-and among-ethnic Chinese [Internet]. The Economist; 2020 [cited 2022 Oct 24]; Available from: <https://www.economist.com/china/2020/02/17/the-coronavirus-spreads-racism-against-and-among-ethnic-chinese>

- 13.** Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: [https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
- 14.** WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [Internet]. WHO; 2020 [cited 2022 Oct 24]; Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- 15.** Machhi J, Herskovitz J, Senan AM, Dutta D, Nath B, Oleynikov MD, Blomberg WR, Meigs DD, Hasan M, Patel M, Kline P, Chang RC, Chang L, Gendelman HE, Kevadiya BD. The Natural History, Pathobiology, and Clinical Manifestations of SARS-CoV-2 Infections. *J Neuroimmune Pharmacol.* 2020 Sep;15(3):359-386. doi: 10.1007/s11481-020-09944-5. Epub 2020 Jul 21. PMID: 32696264; PMCID: PMC7373339.
- 16.** Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg Infect Dis.* 2020 Jul;26(7):1470-1477. doi: 10.3201/eid2607.200282. Epub 2020 Jun 21. PMID: 32255761; PMCID: PMC7323562.
- 17.** Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020 Dec;588(7836):E6. doi: 10.1038/s41586-020-2951-z. Erratum in: *Nature.* 2020 Mar;579(7798):270-273. PMID: 33199918.
- 18.** Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol.* 2020 Apr;5(4):562-569. doi: 10.1038/s41564-020-0688-y. Epub 2020 Feb 24. PMID: 32094589; PMCID: PMC7095430.
- 19.** Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020 Apr 16;181(2):271-280.e8. doi: 10.1016/j.cell.2020.02.052. Epub 2020 Mar 5. PMID: 32142651; PMCID: PMC7102627.
- 20.** Lednicky JA, Tagliamonte MS, White SK, Elbadry MA, Alam MM, Stephenson CJ, Bonny TS, Loeb JC, Telisma T, Chavannes S, Ostrov DA, Mavian C, De Rochars VMB, Salemi M, Morris JG. Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution. *medRxiv* [Preprint]. 2021 Mar 25:2021.03.19.21253391. doi: 10.1101/2021.03.19.21253391. PMID: 33791709; PMCID: PMC8010738.
- 21.** Vlasova AN, Diaz A, Damtie D, Xiu L, Toh TH, Lee JS, Saif LJ, Gray GC. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis.* 2022 Feb 11;74(3):446-454. doi: 10.1093/cid/ciab456. PMID: 34013321; PMCID: PMC8194511.

- 22.**Xu RH, He JF, Evans MR, Peng GW, Field HE, Yu DW, Lee CK, Luo HM, Lin WS, Lin P, Li LH, Liang WJ, Lin JY, Schnur A. Epidemiologic clues to SARS origin in China. *Emerg Infect Dis.* 2004 Jun;10(6):1030-7. doi: 10.3201/eid1006.030852. PMID: 15207054; PMCID: PMC3323155.
- 23.**Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science.* 2003 Oct 10;302(5643):276-8. doi: 10.1126/science.1087139. Epub 2003 Sep 4. PMID: 12958366.
- 24.**Xiao X, Newman C, Buesching CD, Macdonald DW, Zhou ZM. Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci Rep.* 2021 Jun 7;11(1):11898. doi: 10.1038/s41598-021-91470-2. PMID: 34099828; PMCID: PMC8184983.
- 25.**Freuling CM, Breithaupt A, Müller T, Sehl J, Balkema-Buschmann A, Rissmann M, Klein A, Wylezich C, Höper D, Wernike K, Aebischer A, Hoffmann D, Friedrichs V, Dorhoi A, Groschup MH, Beer M, Mettenleiter TC. Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerg Infect Dis.* 2020 Dec;26(12):2982-2985. doi: 10.3201/eid2612.203733. Epub 2020 Oct 22. PMID: 33089771; PMCID: PMC7706974.
- 26.**Centers for Disease Control and Prevention (CDC). Prevalence of IgG antibody to SARS-associated coronavirus in animal traders--Guangdong Province, China, 2003. *MMWR Morb Mortal Wkly Rep.* 2003 Oct 17;52(41):986-7. PMID: 14561956.
- 27.**Wang N, Li SY, Yang XL, Huang HM, Zhang YJ, Guo H, Luo CM, Miller M, Zhu G, Chmura AA, Hagan E, Zhou JH, Zhang YZ, Wang LF, Daszak P, Shi ZL. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virol Sin.* 2018 Feb;33(1):104-107. doi: 10.1007/s12250-018-0012-7. Epub 2018 Mar 2. PMID: 29500691; PMCID: PMC6178078.
- 28.**Woo PC, Lau SK, Chu CM, Chan KH, Tsui HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol.* 2005 Jan;79(2):884-95. doi: 10.1128/JVI.79.2.884-895.2005. PMID: 15613317; PMCID: PMC538593.
- 29.**WHO-Convened Global Study of Origins of SARS-CoV-2: China Part [Internet]. WHO; 2021 [cited 2021 Mar 30]; Available from: <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>
- 30.**Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020 Nov;5(11):1403-1407. doi: 10.1038/s41564-020-0770-5. Epub 2020 Jul 15. PMID: 32669681; PMCID: PMC7610519.
- 31.**Yaqing HE. Surveillance of SARS coronavirus among wild animal sold in Dongmen market in Shenzhen city. 疾病监测. 2004 Aug 20;19(8):287-91.
- 32.**Tu C, Crameri G, Kong X, Chen J, Sun Y, Yu M, Xiang H, Xia X, Liu S, Ren T, Yu Y, Eaton BT, Xuan H, Wang LF. Antibodies to SARS coronavirus in civets. *Emerg Infect Dis.* 2004 Dec;10(12):2244-8. doi: 10.3201/eid1012.040520. PMID: 15663874; PMCID: PMC3323399.
- 33.**Wang M, Jing HQ, Xu HF, Jiang XG, Kan B, Liu QY, Wan KL, Cui BY, Zheng H, Cui ZG, Yan MY, Liang WL, Wang HX, Qi XB, Li ZJ, Li MC, Chen K, Zhang EM, Zhang

- SY, Hai R, Yu DZ, Xu JG. [Surveillance on severe acute respiratory syndrome associated coronavirus in animals at a live animal market of Guangzhou in 2004]. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2005 Feb;26(2):84-7. Chinese. PMID: 15921605.
34. Lytras S, Hughes J, Martin D, Swanepoel P, de Klerk A, Lourens R, Kosakovsky Pond SL, Xia W, Jiang X, Robertson DL. Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biol Evol*. 2022 Feb 4;14(2):evac018. doi: 10.1093/gbe/evac018. PMID: 35137080; PMCID: PMC8882382.
35. Zhou H, Ji J, Chen X, Bi Y, Li J, Wang Q, Hu T, Song H, Zhao R, Chen Y, Cui M, Zhang Y, Hughes AC, Holmes EC, Shi W. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell*. 2021 Aug 19;184(17):4380-4391.e14. doi: 10.1016/j.cell.2021.06.008. Epub 2021 Jun 9. PMID: 34147139; PMCID: PMC8188299.
36. Wacharapluesadee S, Tan CW, Maneeorn P, Duengkae P, Zhu F, Joyjinda Y, Kaewpom T, Chia WN, Ampoot W, Lim BL, Worachotsueptrakun K, Chen VC, Sirichan N, Ruchisrisarod C, Rodpan A, Noradechanon K, Phaichana T, Jantarat N, Thongnumchaima B, Tu C, Crameri G, Stokes MM, Hemachudha T, Wang LF. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat Commun*. 2021 Feb 9;12(1):972. doi: 10.1038/s41467-021-21240-1. Erratum in: *Nat Commun*. 2021 Feb 25;12(1):1430. PMID: 33563978; PMCID: PMC7873279.
37. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, Rambaut A, Robertson DL. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol*. 2020 Nov;5(11):1408-1417. doi: 10.1038/s41564-020-0771-4. Epub 2020 Jul 28. PMID: 32724171.
38. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3. Erratum for: *Nature*. 2020 Dec;588(7836):E6. PMID: 32015507; PMCID: PMC7095418.
39. Li LL, Wang JL, Ma XH, Sun XM, Li JS, Yang XF, Shi WF, Duan ZJ. A novel SARS-CoV-2 related coronavirus with complex recombination isolated from bats in Yunnan province, China. *Emerg Microbes Infect*. 2021 Dec;10(1):1683-1690. doi: 10.1080/22221751.2021.1964925. PMID: 34348599; PMCID: PMC8381922.
40. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, Wang P, Liu D, Yang J, Holmes EC, Hughes AC, Bi Y, Shi W. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*. 2020 Jun 8;30(11):2196-2203.e3. doi: 10.1016/j.cub.2020.05.023. Epub 2020 May 11. Erratum in: *Curr Biol*. 2020 Oct 5;30(19):3896. PMID: 32416074; PMCID: PMC7211627.
41. Pekar J, Worobey M, Moshiri N, Scheffler K, Wertheim JO. Timing the SARS-CoV-2 index case in Hubei province. *Science*. 2021 Apr 23;372(6540):412-417. doi: 10.1126/science.abf8003. Epub 2021 Mar 18. PMID: 33737402; PMCID: PMC8139421.
42. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin

- of SARS coronavirus. *PLoS Pathog.* 2017 Nov 30;13(11):e1006698. doi: 10.1371/journal.ppat.1006698. PMID: 29190287; PMCID: PMC5708621.
43. Luk HKH, Li X, Fung J, Lau SKP, Woo PCY. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect Genet Evol.* 2019 Jul;71:21-30. doi: 10.1016/j.meegid.2019.03.001. Epub 2019 Mar 4. PMID: 30844511; PMCID: PMC7106202.
44. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020 Feb 15;395(10223):507-513. doi: 10.1016/S0140-6736(20)30211-7. Epub 2020 Jan 30. PMID: 32007143; PMCID: PMC7135076.
45. He ML, Zheng BJ, Chen Y, Wong KL, Huang JD, Lin MC, Peng Y, Yuen KY, Sung JJ, Kung HF. Kinetics and synergistic effects of siRNAs targeting structural and replicase genes of SARS-associated coronavirus. *FEBS Lett.* 2006 May 1;580(10):2414-20. doi: 10.1016/j.febslet.2006.03.066. Epub 2006 Mar 30. PMID: 16638566; PMCID: PMC7094648.
46. Wu F, Zhao S, Yu B, Chen YM, Wang W, Hu Y, Song ZG, Tao ZW, Tian JH, Pei YY, Yuan ML. Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. *BioRxiv.* 2020 Jan 1. DOI: 10.1101/2020.01.24.919183.
47. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. *Int J Biol Sci.* 2020 Mar 15;16(10):1686-1697. doi: 10.7150/ijbs.45472. PMID: 32226286; PMCID: PMC7098031.
48. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.* 2015;1282:1-23. doi: 10.1007/978-1-4939-2438-7_1. PMID: 25720466; PMCID: PMC4369385.
49. Prentice E, McAuliffe J, Lu X, Subbarao K, Denison MR. Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J Virol.* 2004 Sep;78(18):9977-86. doi: 10.1128/JVI.78.18.9977-9986.2004. PMID: 15331731; PMCID: PMC514967.
50. Sahu AK, Sreepadmanab M, Rai M, Chande A. SARS-CoV-2: phylogenetic origins, pathogenesis, modes of transmission, and the potential role of nanotechnology. *Virusdisease.* 2021 Mar;32(1):1-12. Figure 2, Structure, Genome organization, and Phylogenetic analysis of SARS-CoV-2; p. 3.
51. Beniac DR, Andonov A, Grudeski E, Booth TF. Architecture of the SARS coronavirus prefusion spike. *Nat Struct Mol Biol.* 2006 Aug;13(8):751-2. doi: 10.1038/nsmb1123. Epub 2006 Jul 16. PMID: 16845391; PMCID: PMC7097490.
52. Delmas B, Laude H. Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J Virol.* 1990 Nov;64(11):5367-75. doi: 10.1128/JVI.64.11.5367-5375.1990. PMID: 2170676; PMCID: PMC248586.
53. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol.* 2016 Sep 29;3(1):237-261. doi: 10.1146/annurev-virology-110615-042301. Epub 2016 Aug 25. PMID: 27578435; PMCID: PMC5457962.
54. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol.* 2003 Aug;77(16):8801-11. doi: 10.1128/jvi.77.16.8801-8811.2003. PMID: 12885899; PMCID: PMC167208.

- 55.** Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*. 2005 Sep 16;309(5742):1864-8. doi: 10.1126/science.1116480. PMID: 16166518.
- 56.** Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, Qi F, Bao L, Du L, Liu S, Qin C, Sun F, Shi Z, Zhu Y, Jiang S, Lu L. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res*. 2020 Apr;30(4):343-355. doi: 10.1038/s41422-020-0305-x. Epub 2020 Mar 30. PMID: 32231345; PMCID: PMC7104723.
- 57.** Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, Ying T, Liu S, Shi Z, Jiang S, Lu L. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol*. 2020 Jul;17(7):765-767. doi: 10.1038/s41423-020-0374-2. Epub 2020 Feb 11. PMID: 32047258; PMCID: PMC7075278.
- 58.** Wikipedia. SARS-CoV-2 [report on the Internet]. 2021 Figure 11, SARS-CoV-2 spike homotrimer with one protein subunit highlighted. Available from: <https://en.wikipedia.org/wiki/SARS-CoV-2>
- 59.** Rabaan AA, Al-Ahmed SH, Haque S, Sah R, Tiwari R, Malik YS, Dhama K, Yatoo MI, Bonilla-Aldana DK, Rodriguez-Morales AJ. SARS-CoV-2, SARS-CoV, and MERS-CoV: A comparative overview. *Infez Med*. 2020 Ahead Of Print Jun 1;28(2):174-184. PMID: 32275259.
- 60.** de Haan CA, Vennema H, Rottier PJ. Assembly of the coronavirus envelope: homotypic interactions between the M proteins. *J Virol*. 2000 Jun;74(11):4967-78. doi: 10.1128/jvi.74.11.4967-4978.2000. PMID: 10799570; PMCID: PMC110848.
- 61.** Nieto-Torres JL, DeDiego ML, Verdiá-Báguena C, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, Castaño-Rodriguez C, Alcaraz A, Torres J, Aguilera VM, Enjuanes L. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog*. 2014 May 1;10(5):e1004077. doi: 10.1371/journal.ppat.1004077. PMID: 24788150; PMCID: PMC4006877.
- 62.** Chang CK, Sue SC, Yu TH, Hsieh CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang WL, Lin CH, Huang TH. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci*. 2006 Jan;13(1):59-72. doi: 10.1007/s11373-005-9035-9. Epub 2005 Oct 14. PMID: 16228284; PMCID: PMC7089556.
- 63.** Kuo L, Masters PS. Functional analysis of the murine coronavirus genomic RNA packaging signal. *J Virol*. 2013 May;87(9):5182-92. doi: 10.1128/JVI.00100-13. Epub 2013 Feb 28. PMID: 23449786; PMCID: PMC3624306.
- 64.** Molenkamp R, Spaan WJ. Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. *Virology*. 1997 Dec 8;239(1):78-86. doi: 10.1006/viro.1997.8867. PMID: 9426448; PMCID: PMC7130520.
- 65.** Stohlman SA, Baric RS, Nelson GN, Soe LH, Welter LM, Deans RJ. Specific interaction between coronavirus leader RNA and nucleocapsid protein. *J Virol*. 1988 Nov;62(11):4288-95. doi: 10.1128/JVI.62.11.4288-4295.1988. PMID: 2845141; PMCID: PMC253863.
- 66.** Hurst KR, Koetzner CA, Masters PS. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-

- transcriptase complex. *J Virol.* 2013 Aug;87(16):9159-72. doi: 10.1128/JVI.01275-13. Epub 2013 Jun 12. PMID: 23760243; PMCID: PMC3754073.
67. Sturman LS, Holmes KV, Behnke J. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J Virol.* 1980 Jan;33(1):449-62. doi: 10.1128/JVI.33.1.449-462.1980. PMID: 6245243; PMCID: PMC288560.
68. Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev.* 2005 Dec;69(4):635-64. doi: 10.1128/MMBR.69.4.635-664.2005. PMID: 16339739; PMCID: PMC1306801.
69. Lai MM, Perlman S, Anderson L, Straus SE. Coronaviridae. In *Fields Virology*. Knipe, DM, Howley, PM, Eds.; LippincottWilliams & Wilkins: Philadelphia, PA, USA, 2007; Volume 1, pp. 1305-18
70. Lai MM, Cavanagh D. The molecular biology of coronaviruses. *Adv Virus Res.* 1997;48:1-100. doi: 10.1016/S0065-3527(08)60286-9. PMID: 9233431; PMCID: PMC7130985.
71. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J Clin Med.* 2020 Apr 24;9(4):1225. Table 1, Classification of Coronaviridae according to the International Committee of Taxonomy of Viruses (ICTV), with special emphasis on reservoir host, zoonotic importance, and major epidemics; p. 6.
72. Tyrell DA, Bynoe ML. Cultivation of a novel type of common-cold virus in organ cultures. *Br Med J.* 1965 Jun 5;1(5448):1467-70. doi: 10.1136/bmj.1.5448.1467. PMID: 14288084; PMCID: PMC2166670.
73. Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguière AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med.* 2003 May 15;348(20):1967-76. doi: 10.1056/NEJMoa030747. Epub 2003 Apr 10. PMID: 12690091.
74. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012 Nov 8;367(19):1814-20. doi: 10.1056/NEJMoa1211721. Epub 2012 Oct 17. Erratum in: *N Engl J Med.* 2013 Jul 25;369(4):394. PMID: 23075143.
75. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 2020 Jan 28;9(1):221-236. doi: 10.1080/22221751.2020.1719902. Erratum in: *Emerg Microbes Infect.* 2020 Dec;9(1):540. PMID: 31987001; PMCID: PMC7067204.
76. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020 Feb 15;395(10223):507-513. doi: 10.1016/S0140-6736(20)30211-7. Epub 2020 Jan 30. PMID: 32007143; PMCID: PMC7135076.
77. Zhao Y, Zhang H, Zhao J, Zhong Q, Jin JH, Zhang GZ. Evolution of infectious bronchitis virus in China over the past two decades. *J Gen Virol.* 2016 Jul;97(7):1566-

1574. doi: 10.1099/jgv.0.000464. Epub 2016 Mar 23. PMID: 27008625; PMCID: PMC7079583.
78. Vlasova AN, Halpin R, Wang S, Ghedin E, Spiro DJ, Saif LJ. Molecular characterization of a new species in the genus Alphacoronavirus associated with mink epizootic catarrhal gastroenteritis. *J Gen Virol.* 2011 Jun;92(Pt 6):1369-1379. doi: 10.1099/vir.0.025353-0. Epub 2011 Feb 23. PMID: 21346029; PMCID: PMC3168282.
79. Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, Zhu Y, Zhang YW, Xie QM, Mani S, Zheng XS, Li B, Li JM, Guo H, Pei GQ, An XP, Chen JW, Zhou L, Mai KJ, Wu ZX, Li D, Anderson DE, Zhang LB, Li SY, Mi ZQ, He TT, Cong F, Guo PJ, Huang R, Luo Y, Liu XL, Chen J, Huang Y, Sun Q, Zhang XL, Wang YY, Xing SZ, Chen YS, Sun Y, Li J, Daszak P, Wang LF, Shi ZL, Tong YG, Ma JY. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature.* 2018 Apr;556(7700):255-258. doi: 10.1038/s41586-018-0010-9. Epub 2018 Apr 4. PMID: 29618817; PMCID: PMC7094983.
80. Elena SF, Sanjuán R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol.* 2005 Sep;79(18):11555-8. doi: 10.1128/JVI.79.18.11555-11558.2005. PMID: 16140732; PMCID: PMC1212614.
81. Ferron F, Subissi L, Silveira De Morais AT, Le NTT, Sevajol M, Gluais L, Decroly E, Vonrhein C, Bricogne G, Canard B, Imbert I. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci U S A.* 2018 Jan 9;115(2):E162-E171. doi: 10.1073/pnas.1718806115. Epub 2017 Dec 26. PMID: 29279395; PMCID: PMC5777078.
82. Lau SK, Lee P, Tsang AK, Yip CC, Tse H, Lee RA, So LY, Lau YL, Chan KH, Woo PC, Yuen KY. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol.* 2011 Nov;85(21):11325-37. doi: 10.1128/JVI.05512-11. Epub 2011 Aug 17. PMID: 21849456; PMCID: PMC3194943.
83. Zhang XW, Yap YL, Danchin A. Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch Virol.* 2005 Jan;150(1):1-20. doi: 10.1007/s00705-004-0413-9. Epub 2004 Oct 11. PMID: 15480857; PMCID: PMC7087341.
84. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3. Erratum for: *Nature.* 2020 Dec;588(7836):E6. PMID: 32015507; PMCID: PMC7095418.
85. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect Genet Evol.* 2020 Apr;79:104212. doi: 10.1016/j.meegid.2020.104212. Epub 2020 Jan 29. PMID: 32004758; PMCID: PMC7106301.
86. Lassmann T, Mathelier A. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics.* 2019 Vol. 36(6):1928-1929. DOI: 10.1093/bioinformatics/btz795 Epub 2019 Oct 26
87. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017

- Jun;14(6):587-589. doi: 10.1038/nmeth.4285. Epub 2017 May 8. PMID: 28481363; PMCID: PMC5453245.
88. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015 Jan;32(1):268-74. doi: 10.1093/molbev/msu300. Epub 2014 Nov 3. PMID: 25371430; PMCID: PMC4271533.
89. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018 Jun 1;35(6):1547-1549. doi: 10.1093/molbev/msy096. PMID: 29722887; PMCID: PMC5967553.
90. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*. 2020 Mar 11;27(3):325-328. doi: 10.1016/j.chom.2020.02.001. Epub 2020 Feb 7. PMID: 32035028; PMCID: PMC7154514.
91. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J Clin Med*. 2020 Apr 24;9(4):1225. Figure 2, Phylogenetic tree based on the complete genome sequences of 45 selected coronaviruses from 18 countries including the SARS-CoV-2, SARS-CoV, HCoV, bat SARS, SARS-like CoV, and MERS-CoV; p. 11.
92. Grubaugh ND, Petrone ME, Holmes EC. We shouldn't worry when a virus mutates during disease outbreaks. *Nat Microbiol*. 2020 Apr;5(4):529-530. doi: 10.1038/s41564-020-0690-4. PMID: 32071422; PMCID: PMC7095397.
93. Sanyaolu A, Okorie C, Marinkovic A, Haider N, Abbasi AF, Jaferi U, Prakash S, Balendra V. The emerging SARS-CoV-2 variants of concern. *Ther Adv Infect Dis*. 2021 Jun 18;8:20499361211024372. doi: 10.1177/20499361211024372. PMID: 34211709; PMCID: PMC8216402.
94. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill*. 2017 Mar 30;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PMID: 28382917; PMCID: PMC5388101.
95. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018 Dec 1;34(23):4121-4123. doi: 10.1093/bioinformatics/bty407. PMID: 29790939; PMCID: PMC6247931.
96. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol*. 2021 Jul 30;7(2):veab064. doi: 10.1093/ve/veab064. PMID: 34527285; PMCID: PMC8344591.
97. Tracking SARS-CoV-2 variants [Internet]. WHO; 2021 [cited 2022 Oct 24]; Available from: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>
98. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, Roberts DJ, Chandra A, Temperton N; CITIID-NIHR BioResource COVID-19 Collaboration; COVID-19 Genomics UK (COG-UK) Consortium, Sharrocks K, Blane E, Modis Y, Leigh KE, Briggs JAG, van Gils MJ, Smith KGC, Bradley JR, Smith C, Doffinger R, Ceron-Gutierrez L, Barcenas-Morales G, Pollock DD, Goldstein RA, Smielewska A, Skittrall JP, Gouliouris T, Goodfellow IG,

- Gkrania-Klotsas E, Illingworth CJR, McCoy LE, Gupta RK. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021 Apr;592(7853):277-282. doi: 10.1038/s41586-021-03291-y. Epub 2021 Feb 5. Erratum in: *Nature*. 2022 Aug;608(7922):E23. PMID: 33545711; PMCID: PMC7610568.
- 99.** Haynes WA, Kamath K, Lucas C, Shon J, Iwasaki A. Impact of B. 1.1. 7 variant mutations on antibody recognition of linear SARS-CoV-2 epitopes. *MedRxiv*. 2021 Jan 1. doi: 10.1101/2021.01.06.20248960.
- 100.** Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med*. 2021 Apr;27(4):620-621. doi: 10.1038/s41591-021-01270-4. Epub 2021 Feb 8. PMID: 33558724.
- 101.** Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. 2020 Sep 3;182(5):1295-1310.e20. doi: 10.1016/j.cell.2020.08.012. Epub 2020 Aug 11. PMID: 32841599; PMCID: PMC7418704.
- 102.** Mylonakis E. SARS-CoV-2 Variants and their Clinical Implications. *R I Med J* (2013). 2021 Sep 1;104(7):59-60. PMID: 34339484.
- 103.** Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, Ku Z, An Z, Scharton D, Schindewolf C, Menachery VD, Shi PY, Weaver SC. The N501Y spike substitution enhances SARS-CoV-2 transmission. *bioRxiv* [Preprint]. 2021 Mar 9:2021.03.08.434499. doi: 10.1101/2021.03.08.434499. Update in: *Nature*. 2021 Nov 24;; PMID: 33758836; PMCID: PMC7986995.
- 104.** Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD; CMMID COVID-19 Working Group; COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. 2021 Apr 9;372(6538):eabg3055. doi: 10.1126/science.abg3055. Epub 2021 Mar 3. PMID: 33658326; PMCID: PMC8128288.
- 105.** Davies NG, Jarvis CI; CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *medRxiv* [Preprint]. 2021 Mar 5:2021.02.01.21250959. doi: 10.1101/2021.02.01.21250959. Update in: *Nature*. 2021 May;593(7858):270-274. PMID: 33564794; PMCID: PMC7872389.
- 106.** Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021 May;593(7857):130-135. doi: 10.1038/s41586-021-03398-2. Epub 2021 Mar 8. PMID: 33684923.
- 107.** Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi

A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021 Apr;592(7854):438-443. doi: 10.1038/s41586-021-03402-9. Epub 2021 Mar 9. PMID: 33690265.

- 108.** Cele S, Gazy I, Jackson L, Hwa SH, Tegally H, Lustig G, Giandhari J, Pillay S, Wilkinson E, Naidoo Y, Karim F, Ganga Y, Khan K, Bernstein M, Balazs AB, Gosnell BI, Hanekom W, Moosa MS; Network for Genomic Surveillance in South Africa; COMMIT-KZN Team, Lessells RJ, de Oliveira T, Sigal A. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature*. 2021 May;593(7857):142-146. doi: 10.1038/s41586-021-03471-w. Epub 2021 Mar 29. PMID: 33780970.
- 109.** Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE, de Oliveira T, Vermeulen M, van der Berg K, Rossouw T, Boswell M, Ueckermann V, Meiring S, von Gottberg A, Cohen C, Morris L, Bhiman JN, Moore PL. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med*. 2021 Apr;27(4):622-625. doi: 10.1038/s41591-021-01285-x. Epub 2021 Mar 2. PMID: 33654292.
- 110.** Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y. V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *BioRxiv*, 2021.01.13.426558. doi: 10.1101/2021.01.13.426558.
- 111.** Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings [Internet]. VirologicalOrg; 2021 [cited 2022 Oct 24] ; Available from: <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>
- 112.** Nascimento VAD, Corado ALG, Nascimento FOD, Costa ÁKAD, Duarte DCG, Luz SLB, Gonçalves LMF, Jesus MS, Costa CFD, Delatorre E, Naveca FG. Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State, Brazil. *Mem Inst Oswaldo Cruz*. 2020 Sep 25;115:e200310. doi: 10.1590/0074-02760200310. PMID: 32997001; PMCID: PMC7523504.
- 113.** Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT, Hulswit RJG, Franco LAM, Ramundo MS, de Jesus JG, Andrade PS, Coletti TM, Ferreira GM, Silva CAM, Manuli ER, Pereira RHM, Peixoto PS, Kraemer MUG, Gaburo N Jr, Camilo CDC, Hoeltgebaum H, Souza WM, Rocha EC, de Souza LM, de Pinho MC, Araujo LJT, Malta FSV, de Lima AB, Silva JDP, Zauli DAG, Ferreira ACS, Schnakenberg RP, Laydon DJ, Walker PGT, Schlüter HM, Dos Santos ALP, Vidal MS, Del Caro VS, Filho RMF, Dos Santos HM, Aguiar RS, Proença-Modena JL, Nelson B, Hay JA, Monod M, Misouridou X, Coupland H, Sonabend R, Vollmer M, Gandy A, Prete CA Jr, Nascimento VH, Suchard MA, Bowden TA, Pond SLK, Wu CH, Ratmann O, Ferguson NM, Dye C, Loman NJ, Lemey P, Rambaut A, Fraiji NA, Carvalho MDPSS, Pybus OG, Flaxman S, Bhatt S, Sabino EC. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021 May 21;372(6544):815-821. doi: 10.1126/science.abh2644. Epub 2021 Apr 14. PMID: 33853970; PMCID: PMC8139423.

- 114.** Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, Feldman J, Pavlovic MN, Gregory DJ, Poznansky MC, Sigal A, Schmidt AG, Iafrate AJ, Naranbhai V, Balazs AB. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. 2021 Apr 29;184(9):2372-2383.e9. doi: 10.1016/j.cell.2021.03.013. Epub 2021 Mar 12. Erratum in: *Cell*. 2021 Apr 29;184(9):2523. PMID: 33743213; PMCID: PMC7953441.
- 115.** Chen LL, Lu L, Choi CY, Cai JP, Tsoi HW, Chu AW, Ip JD, Chan WM, Zhang RR, Zhang X, Tam AR, Lau DP, To WK, Que TL, Yip CC, Chan KH, Cheng VC, Yuen KY, Hung IF, To KK. Impact of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variant-Associated Receptor Binding Domain (RBD) Mutations on the Susceptibility to Serum Antibodies Elicited by Coronavirus Disease 2019 (COVID-19) Infection or Vaccination. *Clin Infect Dis*. 2022 May 3;74(9):1623-1630. doi: 10.1093/cid/ciab656. PMID: 34309648.
- 116.** Wu B, Zhang H, Wang YC, Tang A, Li KF, Li P, Chen JB, Wang HL, Yan JB. Sequencing on an imported case in China of COVID-19 Delta variant emerging from India in a cargo ship in Zhoushan, China. *J Med Virol*. 2021 Dec;93(12):6828-6832. doi: 10.1002/jmv.27239. Epub 2021 Aug 7. PMID: 34314048; PMCID: PMC8426989.
- 117.** Suryadevara N, Shrihari S, Gilchuk P, VanBlargan LA, Binshtein E, Zost SJ, Nargi RS, Sutton RE, Winkler ES, Chen EC, Fouch ME, Davidson E, Doranz BJ, Chen RE, Shi PY, Carnahan RH, Thackray LB, Diamond MS, Crowe JE Jr. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell*. 2021 Apr 29;184(9):2316-2331.e15. doi: 10.1016/j.cell.2021.03.029. Epub 2021 Mar 16. PMID: 33773105; PMCID: PMC7962591.
- 118.** Chen J, Wang R, Wang M, Wei GW. Mutations Strengthened SARS-CoV-2 Infectivity. *J Mol Biol*. 2020 Sep 4;432(19):5212-5226. doi: 10.1016/j.jmb.2020.07.009. Epub 2020 Jul 23. PMID: 32710986; PMCID: PMC7375973.
- 119.** Wang R, Chen J, Gao K, Wei GW. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. *Genomics*. 2021 Jul;113(4):2158-2170. doi: 10.1016/j.ygeno.2021.05.006. Epub 2021 May 15. PMID: 34004284; PMCID: PMC8123493.
- 120.** Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep Med*. 2021 Apr 20;2(4):100255. doi: 10.1016/j.xcrm.2021.100255. Epub 2021 Apr 5. PMID: 33842902; PMCID: PMC8020059.
- 121.** Chen J, Gao K, Wang R, Wei GW. Revealing the Threat of Emerging SARS-CoV-2 Mutations to Antibody Therapies. *J Mol Biol*. 2021 Sep 3;433(18):167155. doi: 10.1016/j.jmb.2021.167155. Epub 2021 Jul 14. PMID: 34273397; PMCID: PMC8277955.
- 122.** Christie A, Brooks JT, Hicks LA, Sauber-Schatz EK, Yoder JS, Honein MA; CDC COVID-19 Response Team. Guidance for Implementing COVID-19 Prevention Strategies in the Context of Varying Community Transmission Levels and Vaccination Coverage. *MMWR Morb Mortal Wkly Rep*. 2021 Jul 27;70(30):1044-1047. doi: 10.15585/mmwr.mm7030e2. PMID: 34324480; PMCID: PMC8323553.

- 123.** SARS-CoV-2 variants of concern and variants under investigation in England Technical briefing 23 [Internet]. Public Health England; 2021 [cited 2022 Oct 24]; Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1018547/Technical_Briefing_23_21_09_16.pdf
- 124.** Kannan SR, Spratt AN, Cohen AR, Naqvi SH, Chand HS, Quinn TP, Lorson CL, Byrareddy SN, Singh K. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *J Autoimmun.* 2021 Nov;124:102715. doi: 10.1016/j.jaut.2021.102715. Epub 2021 Aug 11. PMID: 34399188; PMCID: PMC8354793.
- 125.** SARS-CoV-2 (hCoV-19) Mutation Reports [Internet]. OutbreakInfo; 2021 [cited 2022 Oct 25]; Available from: <https://outbreak.info/situation-reports?pango=AY.1&loc=IND&loc=GBR&loc=USA>
- 126.** Roy B, Roy H. The Delta Plus variant of COVID-19: Will it be the worst nightmare in the SARS-CoV-2 pandemic?. *J. Biomed. Sci.* [Internet]. 2021 Jul. 16 [cited 2022 Oct. 26];8(1):1-2. Available from: <https://www.nepjol.info/index.php/JBS/article/view/38449>
- 127.** Tatsi EB, Filippatos F, Michos A. SARS-CoV-2 variants and effectiveness of vaccines: a review of current evidence. *Epidemiol Infect.* 2021 Nov 4;149:e237. Table 1, Epidemiological and genetic characteristics of the four SARS-CoV-2 variants of concern; p. 3.
- 128.** Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern [Internet]. WHO; 2021 [cited 2022 Oct 25]; Available from: [https://www.who.int/news-room/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news-room/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern)
- 129.** Muralidhar S, Ambi SV, Sekaran S, Krishnan UM. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie.* 2020 Dec;179:85-100. doi: 10.1016/j.biochi.2020.09.018. Epub 2020 Sep 22. PMID: 32971147; PMCID: PMC7505773.
- 130.** Coronavirus Disease 2019 (COVID-19), centers for disease control and prevention [Internet]. CDC; 2020 [cited 2021 Dec 20]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html>
- 131.** Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. *Science.* 2020 Jun 26;368(6498):1422-1424. doi: 10.1126/science.abc6197. Epub 2020 May 27. PMID: 32461212.
- 132.** Stadnytskyi V, Bax CE, Bax A, Anfinrud P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc Natl Acad Sci U S A.* 2020 Jun 2;117(22):11875-11877. doi: 10.1073/pnas.2006874117. Epub 2020 May 13. PMID: 32404416; PMCID: PMC7275719.
- 133.** Transmission of SARS-CoV-2: implications for infection prevention precautions [Internet]. WHO; 2021 [cited 2021 Dec 21]; Available from: <https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions>
- 134.** Novel coronavirus can survive on skin for 9 hours: Study [Internet]. ETHealthworld; 2020 [cited 2021 Dec 21]; Available from:

<https://health.economictimes.indiatimes.com/news/diagnostics/novel-coronavirus-can-survive-on-skin-for-9-hours-study/78587107?redirect=1>

135. WHO acknowledges 'evidence emerging' of airborne spread of COVID-19 [Internet]. ETHEalthworld; 2020 [cited 2021 Dec 21]; Available from: <https://health.economictimes.indiatimes.com/news/industry/who-acknowledges-evidence-emerging-of-airborne-spread-of-covid-19/76847993>
136. Iorio-Morin C, Hodaie M, Sarica C, Dea N, Westwick HJ, Christie SD, McDonald PJ, Labidi M, Farmer JP, Brisebois S, D'Aragon F, Carignan A, Fortin D. Letter: The Risk of COVID-19 Infection During Neurosurgical Procedures: A Review of Severe Acute Respiratory Distress Syndrome Coronavirus 2 (SARS-CoV-2) Modes of Transmission and Proposed Neurosurgery-Specific Measures for Mitigation. *Neurosurgery*. 2020 Aug 1;87(2):E178-E185. doi: 10.1093/neuros/nyaa157. PMID: 32335684; PMCID: PMC7188127.
137. Patri A, Gallo L, Guarino M, Fabbrocini G. Sexual transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A new possible route of infection? *J Am Acad Dermatol*. 2020 Jun;82(6):e227. doi: 10.1016/j.jaad.2020.03.098. Epub 2020 Apr 9. PMID: 32278800; PMCID: PMC7194822.
138. Gu J, Han B, Wang J. COVID-19: Gastrointestinal Manifestations and Potential Fecal-Oral Transmission. *Gastroenterology*. 2020 May;158(6):1518-1519. doi: 10.1053/j.gastro.2020.02.054. Epub 2020 Mar 3. PMID: 32142785; PMCID: PMC7130192.
139. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. *Geroscience*. 2020 Apr;42(2):505-514. doi: 10.1007/s11357-020-00186-0. Epub 2020 Apr 10. Erratum in: *Geroscience*. 2020 May 3;: PMID: 32274617; PMCID: PMC7145538.
140. Nouri-Vaskeh M, Alizadeh L. Fecal transmission in COVID-19: A potential shedding route. *J Med Virol*. 2020 Oct;92(10):1731-1732. doi: 10.1002/jmv.25816. Epub 2020 Apr 6. PMID: 32239515; PMCID: PMC7228240.
141. Alzamora MC, Paredes T, Caceres D, Webb CM, Valdez LM, La Rosa M. Severe COVID-19 during Pregnancy and Possible Vertical Transmission. *Am J Perinatol*. 2020 Jun;37(8):861-865. doi: 10.1055/s-0040-1710050. Epub 2020 Apr 18. PMID: 32305046; PMCID: PMC7356080.
142. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, Li J, Zhao D, Xu D, Gong Q, Liao J, Yang H, Hou W, Zhang Y. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *Lancet*. 2020 Mar 7;395(10226):809-815. doi: 10.1016/S0140-6736(20)30360-3. Epub 2020 Feb 12. Erratum in: *Lancet*. 2020 Mar 28;395(10229):1038. Erratum in: *Lancet*. 2020 Mar 28;395(10229):1038. PMID: 32151335; PMCID: PMC7159281.
143. Lu D, Sang L, Du S, Li T, Chang Y, Yang XA. Asymptomatic COVID-19 infection in late pregnancy indicated no vertical transmission. *J Med Virol*. 2020 Sep;92(9):1660-1664. doi: 10.1002/jmv.25927. Epub 2020 Apr 30. PMID: 32330313; PMCID: PMC7264617.
144. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W, Cheng C, Tang X, Wu X, Wu Y, Sun B, Huang S, Sun Y, Zhang J, Ma T, Lessler J, Feng T. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet*

Infect Dis. 2020 Aug;20(8):911-919. doi: 10.1016/S1473-3099(20)30287-5. Epub 2020 Apr 27. Erratum in: Lancet Infect Dis. 2020 Jul;20(7):e148. PMID: 32353347; PMCID: PMC7185944.

- 145.** Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatr.* 2020 Jun;109(6):1088-1095. doi: 10.1111/apa.15270. Epub 2020 Apr 14. PMID: 32202343; PMCID: PMC7228328.
- 146.** Gu J, Korteweg C. Pathology and pathogenesis of severe acute respiratory syndrome. *Am J Pathol.* 2007 Apr;170(4):1136-47. doi: 10.2353/ajpath.2007.061088. PMID: 17392154; PMCID: PMC1829448.
- 147.** Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med.* 2020 Jun 25;58(7):1021-1028. doi: 10.1515/cclm-2020-0369. PMID: 32286245.
- 148.** Steckler T, Kalin NH, Reul JH. Handbook of stress and the brain part 2: stress: integrative and clinical aspects. Elsevier; 2005 Feb 25. Chapter 2.4, The molecular basis of fever; p. 193-227.
- 149.** Roth J, Blatteis CM. Mechanisms of fever production and lysis: lessons from experimental LPS fever. *Compr Physiol.* 2014 Oct;4(4):1563-604. doi: 10.1002/cphy.c130033. PMID: 25428854.
- 150.** Liu Y, Zhang C, Huang F, Yang Y, Wang F, Yuan J, Zhang Z, Qin Y, Li X, Zhao D, Li S, Tan S, Wang Z, Li J, Shen C, Li J, Peng L, Wu W, Cao M, Xing L, Xu Z, Chen L, Zhou C, Liu WJ, Liu L, Jiang C. Elevated plasma levels of selective cytokines in COVID-19 patients reflect viral load and lung injury. *Natl Sci Rev.* 2020 Jun;7(6):1003-1011. doi: 10.1093/nsr/nwaa037. Epub 2020 Mar 9. PMID: 34676126; PMCID: PMC7107806.
- 151.** Jose RJ, Manuel A. COVID-19 cytokine storm: the interplay between inflammation and coagulation. *Lancet Respir Med.* 2020 Jun;8(6):e46-e47. doi: 10.1016/S2213-2600(20)30216-2. Epub 2020 Apr 27. PMID: 32353251; PMCID: PMC7185942.
- 152.** Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, Huang H, Zhang L, Zhou X, Du C, Zhang Y, Song J, Wang S, Chao Y, Yang Z, Xu J, Zhou X, Chen D, Xiong W, Xu L, Zhou F, Jiang J, Bai C, Zheng J, Song Y. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med.* 2020 Jul 1;180(7):934-943. doi: 10.1001/jamainternmed.2020.0994. Erratum in: *JAMA Intern Med.* 2020 Jul 1;180(7):1031. PMID: 32167524; PMCID: PMC7070509.
- 153.** Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *J Infect.* 2020 Jun;80(6):607-613. doi: 10.1016/j.jinf.2020.03.037. Epub 2020 Apr 10. PMID: 32283152; PMCID: PMC7194613.
- 154.** Smyk W, Janik MK, Portincasa P, Milkiewicz P, Lammert F, Krawczyk M. COVID-19: focus on the lungs but do not forget the gastrointestinal tract. *European Journal of Clinical Investigation.* 2020 Sep;50(9):e13276. doi: <https://doi.org/10.1111/eci.13276>.
- 155.** Hirose R, Ikegaya H, Naito Y, Watanabe N, Yoshida T, Bandou R, Daidoji T, Itoh Y, Nakaya T. Survival of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Influenza Virus on Human Skin: Importance of Hand Hygiene in

Coronavirus Disease 2019 (COVID-19). *Clin Infect Dis.* 2021 Dec 6;73(11):e4329-e4335. doi: 10.1093/cid/ciaa1517. PMID: 33009907; PMCID: PMC7665347.

- 156.** Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin Exp Pediatr.* 2020 Apr;63(4):119-124. doi: 10.3345/cep.2020.00493. Epub 2020 Apr 2. PMID: 32252141; PMCID: PMC7170784.
- 157.** Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003, emergencies preparedness, response [Internet]. WHO; 2021 [cited 2021 Dec 22]; Available from: https://www.who.int/csr/sars/country/table2004_04_21/en/
- 158.** Abd El-Aziz TM, Stockand JD. Recent progress and challenges in drug development against COVID-19 coronavirus (SARS-CoV-2) - an update on the status. *Infect Genet Evol.* 2020 Sep;83:104327. doi: 10.1016/j.meegid.2020.104327. Epub 2020 Apr 19. PMID: 32320825; PMCID: PMC7166307.
- 159.** Liu BM, Yang QQ, Zhao LY, Xie W, Si XY. Epidemiological characteristics of COVID-19 patients in convalescence period. *Epidemiol Infect.* 2020 Jun 3;148:e108. doi: 10.1017/S0950268820001181. PMID: 32487271; PMCID: PMC7287303.
- 160.** Sheervalilou R, Shirvaliloo M, Dadashzadeh N, Shirvalilou S, Shahraki O, Pilehvar-Soltanahmadi Y, Ghaznavi H, Khoei S, Nazarloo Z. COVID-19 under spotlight: A close look at the origin, transmission, diagnosis, and treatment of the 2019-nCoV disease. *J Cell Physiol.* 2020 Dec;235(12):8873-8924. doi: 10.1002/jcp.29735. Epub 2020 May 26. PMID: 32452539; PMCID: PMC7283670.
- 161.** Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C. Virological assessment of hospitalized patients with COVID-2019. *Nature.* 2020 May;581(7809):465-469. doi: 10.1038/s41586-020-2196-x. Epub 2020 Apr 1. Erratum in: *Nature.* 2020 Dec;588(7839):E35. PMID: 32235945.
- 162.** SARS-CoV-2 Variant Classifications and Definitions [Internet]. CDC; 2021 [cited 2021 Dec 22]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>
- 163.** Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, Schaefer-Babajew DJ, DaSilva J, Muecksch F, Gaebler C, Lifton R, Nussenzweig MC, Hatzioannou T, Bieniasz PD, Darnell RB. Vaccine Breakthrough Infections with SARS-CoV-2 Variants. *N Engl J Med.* 2021 Jun 10;384(23):2212-2218. doi: 10.1056/NEJMoa2105000. Epub 2021 Apr 21. PMID: 33882219; PMCID: PMC8117968.
- 164.** Pezzini A, Padovani A. Lifting the mask on neurological manifestations of COVID-19. *Nat Rev Neurol.* 2020 Nov;16(11):636-644. doi: 10.1038/s41582-020-0398-3. Epub 2020 Aug 24. PMID: 32839585; PMCID: PMC7444680.
- 165.** Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, Wade RG. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS One.* 2020 Jun 23;15(6):e0234765. doi: 10.1371/journal.pone.0234765. PMID: 32574165; PMCID: PMC7310678.

- 166.** Coronavirus disease (COVID-19) Symptoms [Internet]. WHO; 2021 [cited 2021 Dec 22]; Available from: https://www.who.int/health-topics/coronavirus#tab=tab_3
- 167.** Coronavirus disease (COVID-19) [Internet]. CDC; 2021 [cited 2021 Dec 22]; Available from: <https://www.cdc.gov/dotw/covid-19/index.html>
- 168.** Coronavirus Disease (COVID-19) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans [Internet]. WHO; 2020 [cited 2021 Dec 22]; Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance>
- 169.** Zhang W, Du RH, Li B, Zheng XS, Yang XL, Hu B, Wang YY, Xiao GF, Yan B, Shi ZL, Zhou P. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect.* 2020 Feb 17;9(1):386-389. doi: 10.1080/22221751.2020.1729071. PMID: 32065057; PMCID: PMC7048229.
- 170.** Chen ZM, Fu JF, Shu Q, Chen YH, Hua CZ, Li FB, Lin R, Tang LF, Wang TL, Wang W, Wang YS, Xu WZ, Yang ZH, Ye S, Yuan TM, Zhang CM, Zhang YY. Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus. *World J Pediatr.* 2020 Jun;16(3):240-246. doi: 10.1007/s12519-020-00345-5. Epub 2020 Feb 5. PMID: 32026148; PMCID: PMC7091166.
- 171.** Shirato K, Nao N, Katano H, Takayama I, Saito S, Kato F, Katoh H, Sakata M, Nakatsu Y, Mori Y, Kageyama T, Matsuyama S, Takeda M. Development of Genetic Diagnostic Methods for Detection for Novel Coronavirus 2019(nCoV-2019) in Japan. *Jpn J Infect Dis.* 2020 Jul 22;73(4):304-307. doi: 10.7883/yoken.JJID.2020.061. Epub 2020 Feb 18. PMID: 32074516.
- 172.** Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 2020 Jan;25(3):2000045. doi: 10.2807/1560-7917.ES.2020.25.3.2000045. Erratum in: *Euro Surveill.* 2020 Apr;25(14): Erratum in: *Euro Surveill.* 2020 Jul;25(30): Erratum in: *Euro Surveill.* 2021 Feb;26(5): PMID: 31992387; PMCID: PMC6988269.
- 173.** Wikipedia. SARS-CoV-2 [report on the Internet]. 2021 Figure 16, Chest X-ray showing COVID-19 pneumonia. Available from: <https://en.wikipedia.org/wiki/COVID-19>
- 174.** Wikipedia. SARS-CoV-2 [report on the Internet]. 2021 Figure 14, A CT scan of a person with COVID-19 shows lesions (bright regions) in the lungs. Available from: <https://en.wikipedia.org/wiki/COVID-19>
- 175.** Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng D, Wang J, Hesketh RL, Yang L, Zheng C. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). *Radiology.* 2020 Jun;295(3):715-721. doi: 10.1148/radiol.2020200370. Epub 2020 Feb 13. PMID: 32053470; PMCID: PMC7233367.
- 176.** Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, Ling Y, Jiang Y, Shi Y. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology.* 2020 Dec;297(3):E346. doi: 10.1148/radiol.2020209021. Erratum for: *Radiology.* 2020 Apr;295(1):210-217. PMID: 33196374; PMCID: PMC8906333.

- 177.** Ding Y, He L, Zhang Q, Huang Z, Che X, Hou J, Wang H, Shen H, Qiu L, Li Z, Geng J, Cai J, Han H, Li X, Kang W, Weng D, Liang P, Jiang S. Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. *J Pathol.* 2004 Jun;203(2):622-30. doi: 10.1002/path.1560. PMID: 15141376; PMCID: PMC7167761.
- 178.** Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012 Nov 8;367(19):1814-20. doi: 10.1056/NEJMoa1211721. Epub 2012 Oct 17. Erratum in: *N Engl J Med.* 2013 Jul 25;369(4):394. PMID: 23075143.
- 179.** Bauer TT, Ewig S, Rodloff AC, Müller EE. Acute respiratory distress syndrome and pneumonia: a comprehensive review of clinical data. *Clin Infect Dis.* 2006 Sep 15;43(6):748-56. doi: 10.1086/506430. Epub 2006 Aug 10. PMID: 16912951; PMCID: PMC7107989.
- 180.** Franks TJ, Chong PY, Chui P, Galvin JR, Lourens RM, Reid AH, Selbs E, McEvoy CP, Hayden CD, Fukuoka J, Taubenberger JK, Travis WD. Lung pathology of severe acute respiratory syndrome (SARS): a study of 8 autopsy cases from Singapore. *Hum Pathol.* 2003 Aug;34(8):743-8. doi: 10.1016/s0046-8177(03)00367-8. Erratum in: *Hum Pathol.* 2004 Jan;35(1):138. PMID: 14506633; PMCID: PMC7119137.
- 181.** Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, Laman JD, de Jong T, van Doornum G, Lim W, Ling AE, Chan PK, Tam JS, Zambon MC, Gopal R, Drosten C, van der Werf S, Escriou N, Manuguerra JC, Stöhr K, Peiris JS, Osterhaus AD. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet.* 2003 Jul 26;362(9380):263-70. doi: 10.1016/S0140-6736(03)13967-0. PMID: 12892955; PMCID: PMC7112434.
- 182.** Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020 Apr 16;181(2):271-280.e8. doi: 10.1016/j.cell.2020.02.052. Epub 2020 Mar 5. PMID: 32142651; PMCID: PMC7102627.
- 183.** Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3. Erratum for: *Nature.* 2020 Dec;588(7836):E6. PMID: 32015507; PMCID: PMC7095418.
- 184.** Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, Cao Y, Yousif AS, Bals J, Hauser BM, Feldman J, Muus C, Wadsworth MH 2nd, Kazer SW, Hughes TK, Doran B, Gatter GJ, Vukovic M, Taliaferro F, Mead BE, Guo Z, Wang JP, Gras D, Plaisant M, Ansari M, Angelidis I, Adler H, Sucre JMS, Taylor CJ, Lin B, Waghray A, Mitsialis V, Dwyer DF, Buchheit KM, Boyce JA, Barrett NA, Laidlaw TM, Carroll SL, Colonna L, Tkachev V, Peterson CW, Yu A, Zheng HB, Gideon HP, Winchell CG, Lin PL, Bingle CD, Snapper SB, Kropski JA, Theis FJ, Schiller HB, Zaragozi LE, Barbry P, Leslie A, Kiem HP, Flynn JL, Fortune SM, Berger B, Finberg RW, Kean LS, Garber M, Schmidt AG, Lingwood D, Shalek AK, Ordovas-Montanes J; HCA Lung Biological Network. Electronic address: lung-

- network@humancellatlas.org; HCA Lung Biological Network. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell*. 2020 May 28;181(5):1016-1035.e19. doi: 10.1016/j.cell.2020.04.035. Epub 2020 Apr 27. PMID: 32413319; PMCID: PMC7252096.
- 185.** Abraham S, Kienzle TE, Lapps W, Brian DA. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology*. 1990 May;176(1):296-301. doi: 10.1016/0042-6822(90)90257-r. PMID: 2184576; PMCID: PMC7157924.
- 186.** Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*. 2020 Apr;176:104742. doi: 10.1016/j.antiviral.2020.104742. Epub 2020 Feb 10. PMID: 32057769; PMCID: PMC7114094.
- 187.** Wang Q, Qiu Y, Li JY, Zhou ZJ, Liao CH, Ge XY. A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility. *Virol Sin*. 2020 Jun;35(3):337-339. doi: 10.1007/s12250-020-00212-7. Epub 2020 Mar 20. PMID: 32198713; PMCID: PMC7091172.
- 188.** Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol*. 2020 Apr 6;30(7):1346-1351.e2. doi: 10.1016/j.cub.2020.03.022. Epub 2020 Mar 19. Erratum in: *Curr Biol*. 2020 Apr 20;30(8):1578. PMID: 32197085; PMCID: PMC7156161.
- 189.** Fink SL, Cookson BT. Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells. *Infect Immun*. 2005 Apr;73(4):1907-16. doi: 10.1128/IAI.73.4.1907-1916.2005. PMID: 15784530; PMCID: PMC1087413.
- 190.** Zhang Y, Gao Y, Qiao L, Wang W, Chen D. Inflammatory Response Cells During Acute Respiratory Distress Syndrome in Patients With Coronavirus Disease 2019 (COVID-19). *Ann Intern Med*. 2020 Sep 1;173(5):402-404. doi: 10.7326/L20-0227. Epub 2020 Apr 13. PMID: 32282871; PMCID: PMC7175423.
- 191.** Huang Z, Jiang Y, Chen J, Zhou Y. Inhibitors of the renin-angiotensin system: The potential role in the pathogenesis of COVID-19. *Cardiol J*. 2020;27(2):171-174. doi: 10.5603/CJ.a2020.0056. Epub 2020 Apr 14. PMID: 32286678; PMCID: PMC8016038.
- 192.** Freeman TL, Swartz TH. Targeting the NLRP3 Inflammasome in Severe COVID-19. *Front Immunol*. 2020 Jun 23;11:1518. doi: 10.3389/fimmu.2020.01518. PMID: 32655582; PMCID: PMC7324760.
- 193.** Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020 Feb 15;395(10223):497-506. doi: 10.1016/S0140-6736(20)30183-5. Epub 2020 Jan 24. Erratum in: *Lancet*. 2020 Jan 30;; PMID: 31986264; PMCID: PMC7159299.
- 194.** Ratajczak MZ, Kucia M. SARS-CoV-2 infection and overactivation of Nlrp3 inflammasome as a trigger of cytokine "storm" and risk factor for damage of hematopoietic stem cells. *Leukemia*. 2020 Jul;34(7):1726-1729. doi: 10.1038/s41375-020-0887-9. Epub 2020 Jun 1. PMID: 32483300; PMCID: PMC7262681.

- 195.** Tian S, Hu W, Niu L, Liu H, Xu H, Xiao SY. Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus (COVID-19) Pneumonia in Two Patients With Lung Cancer. *J Thorac Oncol*. 2020 May;15(5):700-704. doi: 10.1016/j.jtho.2020.02.010. Epub 2020 Feb 28. PMID: 32114094; PMCID: PMC7128866.
- 196.** Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L, Tai Y. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet respiratory medicine*. 2020 Apr 1;8(4):420-2. doi: [https://doi.org/10.1016/S2213-2600\(20\)30076-X](https://doi.org/10.1016/S2213-2600(20)30076-X).
- 197.** Zhang C, Wu Z, Li JW, Zhao H, Wang GQ. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *Int J Antimicrob Agents*. 2020 May;55(5):105954. doi: 10.1016/j.ijantimicag.2020.105954. Epub 2020 Mar 29. PMID: 32234467; PMCID: PMC7118634.
- 198.** Remy KE, Brakenridge SC, Francois B, Daix T, Deutschman CS, Monneret G, Jeannet R, Laterre PF, Hotchkiss RS, Moldawer LL. Immunotherapies for COVID-19: lessons learned from sepsis. *Lancet Respir Med*. 2020 Oct;8(10):946-949. doi: 10.1016/S2213-2600(20)30217-4. Epub 2020 Apr 28. PMID: 32444269; PMCID: PMC7195015.
- 199.** Sun X, Wang T, Cai D, Hu Z, Chen J, Liao H, Zhi L, Wei H, Zhang Z, Qiu Y, Wang J, Wang A. Cytokine storm intervention in the early stages of COVID-19 pneumonia. *Cytokine Growth Factor Rev*. 2020 Jun;53:38-42. doi: 10.1016/j.cytofr.2020.04.002. Epub 2020 Apr 25. PMID: 32360420; PMCID: PMC7182527.
- 200.** Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020 Mar 28;395(10229):1054-1062. doi: 10.1016/S0140-6736(20)30566-3. Epub 2020 Mar 11. Erratum in: *Lancet*. 2020 Mar 28;395(10229):1038. Erratum in: *Lancet*. 2020 Mar 28;395(10229):1038. PMID: 32171076; PMCID: PMC7270627.
- 201.** Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *J Pharm Anal*. 2020 Apr;10(2):102-108. doi: 10.1016/j.jpha.2020.03.001. Epub 2020 Mar 5. PMID: 32282863; PMCID: PMC7104082.
- 202.** Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, Liao P, Qiu JF, Lin Y, Cai XF, Wang DQ, Hu Y, Ren JH, Tang N, Xu YY, Yu LH, Mo Z, Gong F, Zhang XL, Tian WG, Hu L, Zhang XX, Xiang JL, Du HX, Liu HW, Lang CH, Luo XH, Wu SB, Cui XP, Zhou Z, Zhu MM, Wang J, Xue CJ, Li XF, Wang L, Li ZJ, Wang K, Niu CC, Yang QJ, Tang XJ, Zhang Y, Liu XM, Li JJ, Zhang DC, Zhang F, Liu P, Yuan J, Li Q, Hu JL, Chen J, Huang AL. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med*. 2020 Jun;26(6):845-848. doi: 10.1038/s41591-020-0897-1. Epub 2020 Apr 29. PMID: 32350462.
- 203.** Li G, Chen X, Xu A. Profile of specific antibodies to the SARS-associated coronavirus. *N Engl J Med*. 2003 Jul 31;349(5):508-9. doi: 10.1056/NEJM200307313490520. PMID: 12890855.
- 204.** He Z, Zhao C, Dong Q, Zhuang H, Song S, Peng G, Dwyer DE. Effects of severe acute respiratory syndrome (SARS) coronavirus infection on peripheral blood lymphocytes and their subsets. *Int J Infect Dis*. 2005 Nov;9(6):323-30. doi:

10.1016/j.ijid.2004.07.014. Epub 2005 Aug 10. PMID: 16095942; PMCID: PMC7110876.

- 205.** Zheng M, Gao Y, Wang G, Song G, Liu S, Sun D, Xu Y, Tian Z. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol Immunol*. 2020 May;17(5):533-535. doi: 10.1038/s41423-020-0402-2. Epub 2020 Mar 19. PMID: 32203188; PMCID: PMC7091858.
- 206.** Peng H, Yang LT, Li J, Lu ZQ, Wang LY, Koup RA, Bailer RT, Wu CY. Human memory T cell responses to SARS-CoV E protein. *Microbes Infect*. 2006 Aug;8(9-10):2424-31. doi: 10.1016/j.micinf.2006.05.008. Epub 2006 Jun 30. PMID: 16844400; PMCID: PMC7110890.
- 207.** Yang LT, Peng H, Zhu ZL, Li G, Huang ZT, Zhao ZX, Koup RA, Bailer RT, Wu CY. Long-lived effector/central memory T-cell responses to severe acute respiratory syndrome coronavirus (SARS-CoV) S antigen in recovered SARS patients. *Clin Immunol*. 2006 Aug;120(2):171-8. doi: 10.1016/j.clim.2006.05.002. Epub 2006 Jun 16. PMID: 16781892; PMCID: PMC7106132.
- 208.** Torres DA, Ribeiro LDCB, Riello APFL, Horovitz DDG, Pinto LFR, Croda J. Reinfection of COVID-19 after 3 months with a distinct and more aggressive clinical presentation: Case report. *J Med Virol*. 2021 Apr;93(4):1857-1859. doi: 10.1002/jmv.26637. Epub 2020 Nov 10. PMID: 33112002.
- 209.** Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, Laverdure C, Verma SC, Rossetto CC, Jackson D, Farrell MJ, Van Hooser S, Pandori M. Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect Dis*. 2021 Jan;21(1):52-58. doi: 10.1016/S1473-3099(20)30764-7. Epub 2020 Oct 12. PMID: 33058797; PMCID: PMC7550103.
- 210.** Sahu AK, Sreepadmanab M, Rai M, Chande A. SARS-CoV-2: phylogenetic origins, pathogenesis, modes of transmission, and the potential role of nanotechnology. *Virusdisease*. 2021 Mar;32(1):1-12. Figure 1, Transmission and Pathogenesis of SARS-CoV-2; p. 2.
- 211.** Chatterjee SK, Saha S, Munoz MNM. Molecular Pathogenesis, Immunopathogenesis and Novel Therapeutic Strategy Against COVID-19. *Front Mol Biosci*. 2020 Aug 11;7:196. Figure 1, Schematic representation of immunopathogenesis of SARS-CoV-2 infection; p. 6.
- 212.** Medical Conditions, Overview [Internet]. WHO; 2021 [cited 2022 Oct 25]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>
- 213.** Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y, Pan S, Zou X, Yuan S, Shang Y. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020 May;8(5):475-481. doi: 10.1016/S2213-2600(20)30079-5. Epub 2020 Feb 24. Erratum in: *Lancet Respir Med*. 2020 Apr;8(4):e26. PMID: 32105632; PMCID: PMC7102538.
- 214.** Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS; China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 2020 Apr 30;382(18):1708-1720. doi:

10.1056/NEJMoa2002032. Epub 2020 Feb 28. PMID: 32109013; PMCID: PMC7092819.

- 215.** Goumenou M, Sarigiannis D, Tsatsakis A, Anesti O, Docea AO, Petrakis D, Tsoukalas D, Kostoff R, Rakitskii V, Spandidos DA, Aschner M, Calina D. COVID-19 in Northern Italy: An integrative overview of factors possibly influencing the sharp increase of the outbreak (Review). *Mol Med Rep.* 2020 Jul;22(1):20-32. doi: 10.3892/mmr.2020.11079. Epub 2020 Apr 16. PMID: 32319647; PMCID: PMC7248465.
- 216.** Brake SJ, Barnsley K, Lu W, McAlinden KD, Eapen MS, Sohal SS. Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus SARS-CoV-2 (Covid-19). *J Clin Med.* 2020 Mar 20;9(3):841. doi: 10.3390/jcm9030841. PMID: 32244852; PMCID: PMC7141517.
- 217.** Tanya L. Smoking or vaping may increase the risk of a severe coronavirus infection [document on the Internet]: Scientific American; 2020 [cited October 28, 2022]. Available from: <https://www.scientificamerican.com/article/smoking-or-vaping-may-increase-the-risk-of-a-severe-coronavirus-infection1/>
- 218.** Gansevoort RT, Hilbrands LB. CKD is a key risk factor for COVID-19 mortality. *Nat Rev Nephrol.* 2020 Dec;16(12):705-706. doi: 10.1038/s41581-020-00349-4. PMID: 32848205; PMCID: PMC7447963.
- 219.** Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L, Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020 Aug;584(7821):430-436. doi: 10.1038/s41586-020-2521-4. Epub 2020 Jul 8. PMID: 32640463; PMCID: PMC7611074.
- 220.** ERA-EDTA Council; ERACODA Working Group. Chronic kidney disease is a key risk factor for severe COVID-19: a call to action by the ERA-EDTA. *Nephrol Dial Transplant.* 2021 Jan 1;36(1):87-94. doi: 10.1093/ndt/gfaa314. PMID: 33340043; PMCID: PMC7771976.
- 221.** Lew SQ, Wallace EL, Srivatana V, Warady BA, Watnick S, Hood J, White DL, Aggarwal V, Wilkie C, Naljayan MV, Gellens M, Perl J, Schreiber MJ. Telehealth for Home Dialysis in COVID-19 and Beyond: A Perspective From the American Society of Nephrology COVID-19 Home Dialysis Subcommittee. *Am J Kidney Dis.* 2021 Jan;77(1):142-148. doi: 10.1053/j.ajkd.2020.09.005. Epub 2020 Sep 28. PMID: 33002530; PMCID: PMC7521438.
- 222.** Corbett RW, Blakey S, Nitsch D, Loucaidou M, McLean A, Duncan N, Ashby DR; West London Renal and Transplant Centre. Epidemiology of COVID-19 in an Urban Dialysis Center. *J Am Soc Nephrol.* 2020 Aug;31(8):1815-1823. doi: 10.1681/ASN.2020040534. Epub 2020 Jun 19. PMID: 32561681; PMCID: PMC7460899.
- 223.** Hsu CM, Weiner DE, Aweh G, Salenger P, Johnson DS, Lacson E Jr. Epidemiology and Outcomes of COVID-19 in Home Dialysis Patients Compared with In-Center Dialysis Patients. *J Am Soc Nephrol.* 2021 Jun 9;32(7):1569–73. doi: 10.1681/ASN.2020111653. Epub ahead of print. PMID: 34108232; PMCID: PMC8425660.
- 224.** Perl J, Thomas D, Tang Y, Yeung A, Ip J, Oliver MJ, Blake PG. COVID-19 among Adults Receiving Home versus In-Center Dialysis. *Clin J Am Soc Nephrol.*

2021 Sep;16(9):1410-1412. doi: 10.2215/CJN.04170321. Epub 2021 Jun 4. PMID: 34088719; PMCID: PMC8729571.

- 225.** CDC COVID-19 Response Team. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep. 2020 Apr 3;69(13):382-386. doi: 10.15585/mmwr.mm6913e2. PMID: 32240123; PMCID: PMC7119513.
- 226.** Steinberg MH. Management of sickle cell disease. N Engl J Med. 1999 Apr 1;340(13):1021-30. doi: 10.1056/NEJM199904013401307. PMID: 10099145.
- 227.** Pregnant People At Increased Risk for Severe Illness from COVID-19 [Internet]. CDC; 2022 [cited 2022 Oct 25]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/pregnant-people.html>
- 228.** Çakır Edis E. Chronic Pulmonary Diseases and COVID-19. Turk Thorac J. 2020 Sep;21(5):345-349. doi: 10.5152/TurkThoracJ.2020.20091. Epub 2020 Sep 1. PMID: 33031727; PMCID: PMC7544403.
- 229.** Hariyanto TI, Putri C, Arisa J, Situmeang RFV, Kurniawan A. Dementia and outcomes from coronavirus disease 2019 (COVID-19) pneumonia: A systematic review and meta-analysis. Arch Gerontol Geriatr. 2021 Mar-Apr;93:104299. doi: 10.1016/j.archger.2020.104299. Epub 2020 Nov 19. PMID: 33285424; PMCID: PMC7674980.
- 230.** Hüls A, Costa ACS, Dierssen M, Baksh RA, Bargagna S, Baumer NT, Brandão AC, Carfi A, Carmona-Iragui M, Chicoine BA, Ghosh S, Lakhnau M, Manso C, Mayer MA, Ortega MDC, de Asua DR, Rebillat AS, Russell LA, Sgandurra G, Valentini D, Sherman SL, Strydom A; T21RS COVID-19 Initiative. Medical vulnerability of individuals with Down syndrome to severe COVID-19-data from the Trisomy 21 Research Society and the UK ISARIC4C survey. EClinicalMedicine. 2021 Mar;33:100769. doi: 10.1016/j.eclinm.2021.100769. Epub 2021 Feb 22. PMID: 33644721; PMCID: PMC7897934.
- 231.** Safer Drug Use During COVID-19 [Internet]. CDC; 2022 [cited 2022 Oct 25]; Available from: <https://www.cdc.gov/drugoverdose/resources/covid-drugs-QA.html#>
- 232.** Shannon F Singapore: The Model for COVID-19 Response? - Many imported cases, high population density - yet virtually no community spread so far [document on the Internet]: MedPageToday; 2020 [cited 2022 Oct 28]. Available from: <https://www.medpagetoday.com/infectiousdisease/covid19/85254>
- 233.** Protect Yourself & Others [Internet]. CDC; 2022 [cited 2022 Oct 25]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>
- 234.** Kottasova I Italy shuts all schools over coronavirus outbreak [document on the Internet]: CNN; 2020 [cited 2022 Oct 28]. Available from: <https://edition.cnn.com/2020/03/04/europe/italy-schools-closures-coronavirus-intl/index.html>
- 235.** Coronavirus (covid-19): what is social distancing? [Internet]: UK Health Security Agency; 2022 [cited 2022 Oct 28]. Available from: <https://ukhsa.blog.gov.uk/2020/03/04/coronavirus-covid-19-what-is-social-distancing/>
- 236.** Pueyo T Coronavirus : why you must act now [document on the Internet]: Medium.com; 2020 [cited 2022 Oct 28]. Available from:

<https://tomaspueyo.medium.com/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca>

- 237.** What to Do If You Are Sick [Internet]. CDC; 2022 [cited 2022 Oct 25]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/steps-when-sick.html>
- 238.** Qian M, Jiang J. COVID-19 and social distancing. Z Gesundh Wiss. 2022;30(1):259-261. doi: 10.1007/s10389-020-01321-z. Epub 2020 May 25. PMID: 32837835; PMCID: PMC7247774.
- 239.** Beware of Severe Respiratory Disease associated with a Novel Infectious Agent [Internet]. FEHD; 2020 [cited 2022 Oct 25]; Available from: https://www.fehd.gov.hk/english/licensing/guide_general_reference/respiratory_disease_NIA.html
- 240.** Advice on the use of masks in the community, during home care and in health care settings in the context of the novel coronavirus (2019-nCoV) outbreak: interim guidance, 29 January 2020 [Internet]. WHO; 2020 [cited 2022 Oct 25]; Available from: <https://apps.who.int/iris/handle/10665/330987>
- 241.** Interim Infection Prevention and Control Recommendations for Patients with Confirmed Coronavirus Disease 2019 (COVID-19) or Persons Under Investigation for COVID-19 in Healthcare Settings [Internet]. CDC; 2020 [cited 2021 Nov 22]; Available from: <https://www.cdc.gov/coronavirus/2019-ncov/infection-control/control-recommendations.html>
- 242.** Regalis Τύποι Μασκών Προστασίας [report on the Internet]. 2021 Figure 1, Τύποι Μασκών Προστασίας. 2021 Available from: <https://www.regalis.gr/typoi-maskon-prostasias.html>
- 243.** Wikipedia. SARS-CoV-2 [report on the Internet]. 2021 Figure 14, Αναπαράσταση αερολυμάτων μεταξύ μάσκας με βαλβίδα, μάσκας χωρίς βαλβίδα και χωρίς κανένα μέτρο προστασίας. Available from: <https://el.wikipedia.org/wiki/COVID-19>
- 244.** Summary of Guidance for Minimizing the Impact of COVID-19 on Individual Persons, Communities, and Health Care Systems - United States [Internet]. CDC; 2022 [cited 2022 Oct 28]; Available from: www.cdc.gov/mmwr/volumes/71/wr/mm7133e1.htm
- 245.** Πανδημία Κορονοϊού Τεχνικός Οδηγός για Συστήματα Κλιματισμού σε υποστατικά εκτός του τομέα παροχής φροντίδας υγείας. Αναθεωρημένη Έκδοση (Φεβρουάριος 2021) [Internet]. Press and Information Office Cyprus; 2021 [cited 2022 Oct 25]; Available from: https://www.pio.gov.cy/coronavirus/uploads/feb2021_revised_sistimaklimatsimou.pdf
- 246.** Chaabene H, Prieske O, Herz M, Moran J, Höhne J, Kliegl R, Ramirez-Campillo R, Behm DG, Hortobágyi T, Granacher U. Home-based exercise programmes improve physical fitness of healthy older adults: A PRISMA-compliant systematic review and meta-analysis with relevance for COVID-19. Ageing Res Rev. 2021 May;67:101265. doi: 10.1016/j.arr.2021.101265. Epub 2021 Feb 8. PMID: 33571702.
- 247.** Πως ενισχύουμε το ανοσοποιητικό μας μέσω της καθημερινής μας διατροφής [Internet]. Ministry of Health Cyprus; 2021 [cited 2022 Oct 25]; Available from: <https://www.moh.gov.cy/moh/moh.nsf/all/50C324389FC539C5C22585A5003625DC>

- 248.** Brito DTM, Ribeiro LHC, Daltro CHDC, Silva RB. The possible benefits of vitamin D in COVID-19. *Nutrition.* 2021 Nov-Dec;91-92:111356. doi: 10.1016/j.nut.2021.111356. Epub 2021 May 26. PMID: 34352586; PMCID: PMC8149468.
- 249.** Krause PR, Fleming TR, Longini IM, Peto R, Briand S, Heymann DL, Beral V, Snape MD, Rees H, Ropero AM, Balicer RD, Cramer JP, Muñoz-Fontela C, Gruber M, Gaspar R, Singh JA, Subbarao K, Van Kerkhove MD, Swaminathan S, Ryan MJ, Henao-Restrepo AM. SARS-CoV-2 Variants and Vaccines. *N Engl J Med.* 2021 Jul 8;385(2):179-186. doi: 10.1056/NEJMsr2105280. Epub 2021 Jun 23. PMID: 34161052; PMCID: PMC8262623.
- 250.** COVID-19 vaccine tracker [Internet]. LSHTM; 2021 [cited 2021 Nov 22]; Available from: https://vac-lshtm.shinyapps.io/ncov_vaccine_landscape/
- 251.** TheJapanTimes. Japan panel gives Moderna and AstraZeneca vaccines the green light [report on the Internet]. 2021. Available from: <https://www.japantimes.co.jp/news/2021/05/20/national/moderna-astrazeneca-approval/>
- 252.** Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, Frenck RW Jr, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Jansen KU, Gruber WC; C4591001 Clinical Trial Group. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med.* 2020 Dec 31;383(27):2603-2615. doi: 10.1056/NEJMoa2034577. Epub 2020 Dec 10. PMID: 33301246; PMCID: PMC7745181.
- 253.** Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, Hernán MA, Lipsitch M, Reis B, Balicer RD. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N Engl J Med.* 2021 Apr 15;384(15):1412-1423. doi: 10.1056/NEJMoa2101765. Epub 2021 Feb 24. PMID: 33626250; PMCID: PMC7944975.
- 254.** Michos A, Tatsi EB, Filippatos F, Dellis C, Koukou D, Efthymiou V, Kastrinelli E, Mantzou A, Syriopoulou V. Association of total and neutralizing SARS-CoV-2 spike -receptor binding domain antibodies with epidemiological and clinical characteristics after immunization with the 1st and 2nd doses of the BNT162b2 vaccine. *Vaccine.* 2021 Sep 24;39(40):5963-5967. doi: 10.1016/j.vaccine.2021.07.067. Epub 2021 Jul 24. PMID: 34400017; PMCID: PMC8302834.
- 255.** Muik A, Wallisch AK, Sänger B, Swanson KA, Mühl J, Chen W, Cai H, Maurus D, Sarkar R, Türeci Ö, Dormitzer PR, Şahin U. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science.* 2021 Mar 12;371(6534):1152-1153. doi: 10.1126/science.abg6105. Epub 2021 Jan 29. PMID: 33514629; PMCID: PMC7971771.
- 256.** Tada T, Dcosta BM, Samanovic-Golden M, Herati RS, Cornelius A, Mulligan MJ, Landau NR. Neutralization of viruses with European, South African, and United States SARS-CoV-2 variant spike proteins by convalescent sera and BNT162b2 mRNA vaccine-elicited antibodies. *bioRxiv [Preprint].* 2021 Feb 7:2021.02.05.430003. doi: 10.1101/2021.02.05.430003. Update in: *mBio.* 2021 Jun 1;:e0069621. PMID: 33564768; PMCID: PMC7872356.

- 257.** Abu-Raddad LJ, Chemaiteily H, Butt AA; National Study Group for COVID-19 Vaccination. Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants. *N Engl J Med.* 2021 Jul 8;385(2):187-189. doi: 10.1056/NEJMc2104974. Epub 2021 May 5. PMID: 33951357; PMCID: PMC8117967.
- 258.** Hall VJ, Foulkes S, Saei A, Andrews N, Oguti B, Charlett A, Wellington E, Stowe J, Gillson N, Atti A, Islam J, Karagiannis I, Munro K, Khawam J, Chand MA, Brown CS, Ramsay M, Lopez-Bernal J, Hopkins S; SIREN Study Group. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. *Lancet.* 2021 May 8;397(10286):1725-1735. doi: 10.1016/S0140-6736(21)00790-X. Epub 2021 Apr 23. PMID: 33901423; PMCID: PMC8064668.
- 259.** Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Le Guen J, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Courtellemont L, Péré H, Hocqueloux L, Fafi-Kremer S, Prazuck T, Mouquet H, Bruel T, Simon-Lorière E, Rey FA, Schwartz O. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature.* 2021 Aug;596(7871):276-280. doi: 10.1038/s41586-021-03777-9. Epub 2021 Jul 8. PMID: 34237773.
- 260.** Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Dabrera G, Myers R, Campbell CNJ, Amirthalingam G, Edmunds M, Zambon M, Brown KE, Hopkins S, Chand M, Ramsay M. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med.* 2021 Aug 12;385(7):585-594. doi: 10.1056/NEJMoa2108891. Epub 2021 Jul 21. PMID: 34289274; PMCID: PMC8314739.
- 261.** Shen X, Tang H, Pajon R, Smith G, Glenn GM, Shi W, Korber B, Montefiori DC. Neutralization of SARS-CoV-2 Variants B.1.429 and B.1.351. *N Engl J Med.* 2021 Jun 17;384(24):2352-2354. doi: 10.1056/NEJMc2103740. Epub 2021 Apr 7. PMID: 33826819; PMCID: PMC8063884.
- 262.** Wu K, Werner AP, Koch M, Choi A, Narayanan E, Stewart-Jones GBE, Colpitts T, Bennett H, Boyoglu-Barnum S, Shi W, Moliva JI, Sullivan NJ, Graham BS, Carfi A, Corbett KS, Seder RA, Edwards DK. Serum Neutralizing Activity Elicited by mRNA-1273 Vaccine. *N Engl J Med.* 2021 Apr 15;384(15):1468-1470. doi: 10.1056/NEJMc2102179. Epub 2021 Mar 17. PMID: 33730471; PMCID: PMC8008744.
- 263.** Shen X, Tang H, Pajon R, Smith G, Glenn GM, Shi W, Korber B, Montefiori DC. Neutralization of SARS-CoV-2 Variants B.1.429 and B.1.351. *N Engl J Med.* 2021 Jun 17;384(24):2352-2354. doi: 10.1056/NEJMc2103740. Epub 2021 Apr 7. PMID: 33826819; PMCID: PMC8063884.
- 264.** Edara VV, Pinsky BA, Suthar MS, Lai L, Davis-Gardner ME, Floyd K, Flowers MW, Wrammert J, Hussaini L, Cirim CR, Bechnak S, Stephens K, Graham BS, Bayat Mokhtari E, Mudvari P, Boritz E, Creanga A, Pegu A, Derrien-Colemy A, Henry AR, Gagne M, Douek DC, Sahoo MK, Sibai M, Solis D, Webby RJ, Jeevan T, Fabrizio TP. Infection and Vaccine-Induced Neutralizing-Antibody Responses to the SARS-CoV-2 B.1.617 Variants. *N Engl J Med.* 2021 Aug 12;385(7):664-666. doi: 10.1056/NEJMc2107799. Epub 2021 Jul 7. PMID: 34233096; PMCID: PMC8279090.
- 265.** McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, Tortorici MA, Navarro MJ, Silacci-Fregni C, Saliba C, Sprouse KR, Agostini M, Pinto D, Culap K, Bianchi S, Jaconi S, Cameroni E, Bowen JE, Tilles SW, Pizzuto MS, Guastalla SB,

Bona G, Pellanda AF, Garzoni C, Van Voorhis WC, Rosen LE, Snell G, Telenti A, Virgin HW, Piccoli L, Corti D, Veesler D. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science*. 2021 Aug 6;373(6555):648-654. doi: 10.1126/science.abi7994. Epub 2021 Jul 1. PMID: 34210893.

- 266.** Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, Goepfert PA, Truyers C, Fennema H, Spiessens B, Offergeld K, Scheper G, Taylor KL, Robb ML, Treanor J, Barouch DH, Stoddard J, Ryser MF, Marovich MA, Neuzil KM, Corey L, Cauwenberghs N, Tanner T, Hardt K, Ruiz-Guiñazú J, Le Gars M, Schuitemaker H, Van Hoof J, Struyf F, Douoguih M; ENSEMBLE Study Group. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N Engl J Med*. 2021 Jun 10;384(23):2187-2201. doi: 10.1056/NEJMoa2101544. Epub 2021 Apr 21. PMID: 33882225; PMCID: PMC8220996.
- 267.** Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, Baillie VL, Barnabas SL, Bhorat QE, Bibi S, Briner C, Cicconi P, Collins AM, Colin-Jones R, Cutland CL, Darton TC, Dheda K, Duncan CJA, Emary KRW, Ewer KJ, Fairlie L, Faust SN, Feng S, Ferreira DM, Finn A, Goodman AL, Green CM, Green CA, Heath PT, Hill C, Hill H, Hirsch I, Hodgson SHC, Izu A, Jackson S, Jenkin D, Joe CCD, Kerridge S, Koen A, Kwatra G, Lazarus R, Lawrie AM, Lelliott A, Libri V, Lillie PJ, Mallory R, Mendes AVA, Milan EP, Minassian AM, McGregor A, Morrison H, Mujadidi YF, Nana A, O'Reilly PJ, Padayachee SD, Pittella A, Plested E, Pollock KM, Ramasamy MN, Rhead S, Schwarzbold AV, Singh N, Smith A, Song R, Snape MD, Sprinz E, Sutherland RK, Tarrant R, Thomson EC, Török ME, Toshner M, Turner DPJ, Vekemans J, Villafana TL, Watson MEE, Williams CJ, Douglas AD, Hill AVS, Lambe T, Gilbert SC, Pollard AJ; Oxford COVID Vaccine Trial Group. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*. 2021 Jan 9;397(10269):99-111. doi: 10.1016/S0140-6736(20)32661-1. Epub 2020 Dec 8. Erratum in: *Lancet*. 2021 Jan 9;397(10269):98. PMID: 33306989; PMCID: PMC7723445.
- 268.** Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. *N Engl J Med*. 2021 Jun 3;384(22):2092-2101. doi: 10.1056/NEJMoa2104840. Epub 2021 Apr 9. PMID: 33835769; PMCID: PMC8095372.
- 269.** Emary KRW, Golubchik T, Aley PK, Ariani CV, Angus B, Bibi S, Blane B, Bonsall D, Cicconi P, Charlton S, Clutterbuck EA, Collins AM, Cox T, Darton TC, Dold C, Douglas AD, Duncan CJA, Ewer KJ, Flaxman AL, Faust SN, Ferreira DM, Feng S, Finn A, Folegatti PM, Fuskova M, Galiza E, Goodman AL, Green CM, Green CA, Greenland M, Hallis B, Heath PT, Hay J, Hill HC, Jenkin D, Kerridge S, Lazarus R, Libri V, Lillie PJ, Ludden C, Marchevsky NG, Minassian AM, McGregor AC, Mujadidi YF, Phillips DJ, Plested E, Pollock KM, Robinson H, Smith A, Song R, Snape MD, Sutherland RK, Thomson EC, Toshner M, Turner DPJ, Vekemans J, Villafana TL, Williams CJ, Hill AVS, Lambe T, Gilbert SC, Voysey M, Ramasamy MN, Pollard AJ; COVID-19 Genomics UK consortium; AMPHEUS Project; Oxford COVID-19 Vaccine Trial Group. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. *Lancet*. 2021 Apr 10;397(10282):1351-1362. doi: 10.1016/S0140-6736(21)00628-0. Epub 2021 Mar 30. PMID: 33798499; PMCID: PMC8009612.
- 270.** Wall EC, Wu M, Harvey R, Kelly G, Warchal S, Sawyer C, Daniels R, Adams L, Hobson P, Hatipoglu E, Ngai Y, Hussain S, Ambrose K, Hindmarsh S, Beale R, Riddell A, Gamblin S, Howell M, Kassiotis G, Libri V, Williams B, Swanton C, Gandhi

- S, Bauer DL. AZD1222-induced neutralising antibody activity against SARS-CoV-2 Delta VOC. *Lancet*. 2021 Jul 17;398(10296):207-209. doi: 10.1016/S0140-6736(21)01462-8. Epub 2021 Jun 28. PMID: 34197809; PMCID: PMC8238446.
- 271.** Coronavirus (COVID-19) Treatment [Internet]: WebMD.com; 2021 [cited 2022 Oct 28]. Available from: <https://www.webmd.com/lung/covid-treatment-home-hospital>
- 272.** COVID-19 Θεραπευτικός Αλγόριθμος [Internet]: ΕΟΔΥ; 2021 [cited 2021 Nov 28]. Available from: <https://eody.gov.gr/wp-content/uploads/2020/11/covid-19-therapeutikos-algorithmos.pdf>
- 273.** Τι είναι η ενδοτραχειακή διασωλήνωση [Internet]: MedicaHellas.gr; 2021 [cited 2021 Nov 28]. Available from: https://www.medicahellas.gr/blog/ti-einai-i-endotraheiaki-diasolinosi.html?gclid=CjwKCAiAwKyNBhBfEiwA_mrUMhklo9o48pm7S0EAyupRAru3kl288UU2QImXSYqXFbp0wNgVKfjpxoCdY0QAvD_BwE
- 274.** Hsieh CL, Goldsmith JA, Schaub JM, DiVenere AM, Kuo HC, Javanmardi K, Le KC, Wrapp D, Lee AG, Liu Y, Chou CW, Byrne PO, Hjorth CK, Johnson NV, Ludes-Meyers J, Nguyen AW, Park J, Wang N, Amengor D, Lavinder JJ, Ippolito GC, Maynard JA, Finkelstein IJ, McLellan JS. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*. 2020 Sep 18;369(6510):1501-1505. doi: 10.1126/science.abd0826. Epub 2020 Jul 23. PMID: 32703906; PMCID: PMC7402631.
- 275.** Osman EEA, Toogood PL, Neamati N. COVID-19: Living through Another Pandemic. *ACS Infect Dis*. 2020 Jul 10;6(7):1548-1552. doi: 10.1021/acsinfecdis.0c00224. Epub 2020 May 10. PMID: 32388976; PMCID: PMC7216760.
- 276.** Wadman M, Couzin-Frankel J, Kaiser J, Matacic C. A rampage through the body. *Science*. 2020 Apr 24;368(6489):356-360. doi: 10.1126/science.368.6489.356. PMID: 32327580.
- 277.** Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. 2020 Mar 27;367(6485):1444-1448. doi: 10.1126/science.abb2762. Epub 2020 Mar 4. PMID: 32132184; PMCID: PMC7164635.
- 278.** Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nat Rev Microbiol*. 2009 Mar;7(3):226-36. doi: 10.1038/nrmicro2090. Epub 2009 Feb 9. PMID: 19198616; PMCID: PMC2750777.
- 279.** Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020 Apr 16;181(2):281-292.e6. doi: 10.1016/j.cell.2020.02.058. Epub 2020 Mar 9. Erratum in: *Cell*. 2020 Dec 10;183(6):1735. PMID: 32155444; PMCID: PMC7102599.
- 280.** Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020 Mar 13;367(6483):1260-1263. doi: 10.1126/science.abb2507. Epub 2020 Feb 19. PMID: 32075877; PMCID: PMC7164637.
- 281.** Luytjes W, Sturman LS, Bredenbeek PJ, Charite J, van der Zeijst BA, Horzinek MC, Spaan WJ. Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and

- identification of the trypsin cleavage site. *Virology*. 1987 Dec;161(2):479-87. doi: 10.1016/0042-6822(87)90142-5. PMID: 2825419; PMCID: PMC7130946.
- 282.** Abraham S, Kienzle TE, Lapps W, Brian DA. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology*. 1990 May;176(1):296-301. doi: 10.1016/0042-6822(90)90257-r. PMID: 2184576; PMCID: PMC7157924.
- 283.** de Groot RJ, Luytjes W, Horzinek MC, van der Zeijst BA, Spaan WJ, Lenstra JA. Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J Mol Biol*. 1987 Aug 20;196(4):963-6. doi: 10.1016/0022-2836(87)90422-0. PMID: 3681988; PMCID: PMC7131189.
- 284.** Coronavirus Recovery [document on the Internet]: WebMD.com; 2021 [cited 2022 Oct 28]. Available from: <https://www.webmd.com/lung/covid-recovery-overview>
- 285.** Daugherty SE, Guo Y, Heath K, Dasmariñas MC, Jubilo KG, Samranvedhya J, Lipsitch M, Cohen K. Risk of clinical sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ*. 2021 May 19;373:n1098. doi: 10.1136/bmj.n1098. PMID: 34011492; PMCID: PMC8132065.
- 286.** Cohen J Wuhan seafood market may not be source of novel virus spreading globally [document on the Internet]: Science; 2020 [cited 2022 Oct 28]. Available from: <https://www.science.org/content/article/wuhan-seafood-market-may-not-be-source-novel-virus-spreading-globally>
- 287.** Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020 Feb 15;395(10223):470-473. doi: 10.1016/S0140-6736(20)30185-9. Epub 2020 Jan 24. Erratum in: *Lancet*. 2020 Jan 29;; PMID: 31986257; PMCID: PMC7135038.
- 288.** Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020 Feb 15;395(10223):497-506. doi: 10.1016/S0140-6736(20)30183-5. Epub 2020 Jan 24. Erratum in: *Lancet*. 2020 Jan 30;; PMID: 31986264; PMCID: PMC7159299.
- 289.** Joseph A New coronavirus can cause infections with no symptoms and sicken otherwise healthy people, studies show [document on the Internet]: Statnews; 2020 [cited 2022 Oct 28]. Available from: <https://www.statnews.com/2020/01/24/coronavirus-infections-no-symptoms-lancet-studies/>
- 290.** Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 2020 Feb 15;395(10223):514-523. doi: 10.1016/S0140-6736(20)30154-9. Epub 2020 Jan 24. PMID: 31986261; PMCID: PMC7159286.
- 291.** Pekar J, Worobey M, Moshiri N, Scheffler K, Wertheim JO. Timing the SARS-CoV-2 index case in Hubei province. *Science*. 2021 Apr 23;372(6540):412-417. doi: 10.1126/science.abf8003. Epub 2021 Mar 18. PMID: 33737402; PMCID: PMC8139421.
- 292.** To KK, Sridhar S, Chiu KH, Hung DL, Li X, Hung IF, Tam AR, Chung TW, Chan JF, Zhang AJ, Cheng VC, Yuen KY. Lessons learned 1 year after SARS-CoV-

2 emergence leading to COVID-19 pandemic. *Emerg Microbes Infect.* 2021 Dec;10(1):507-535. doi: 10.1080/22221751.2021.1898291. PMID: 33666147; PMCID: PMC8006950.

- 293.** WHO Coronavirus (COVID-19) Dashboard [Internet]. WHO; 2021 [cited 2021 Dec 3]. Available from: <https://covid19.who.int>
- 294.** Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases [Internet]. WHO; 2020 [cited 2021 Dec 3]. Available from: <https://www.who.int/publications/i/item/10665-331501>
- 295.** Total tests for COVID-19 per 1,000 people [Internet]. OurWorldInData; 2021 [cited 2021 Dec 3]. Available from: <https://ourworldindata.org/grapher/full-list-cumulative-total-tests-per-thousand>
- 296.** Sevillano E, P Linde , S Vizoso 640,000 rapid coronavirus tests arrive in spain [document on the Internet]: ELPAIS; 2020 [cited 2022 Oct 28]. Available from: <https://english.elpais.com/society/2020-03-23/640000-rapid-coronavirus-tests-arrive-in-spain.html>
- 297.** Parodi E, S Jewkes , S Cha et al. Special report: italy and south korea virus outbreaks reveal disparity in deaths and tactics [document on the Internet]: Reuters; 2020 [cited 2022 Oct 28]. Available from: <https://www.reuters.com/article/us-health-coronavirus-response-specialre-idUSKBN20Z27P>
- 298.** Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* 2020 May 1;368(6490):489-493. doi: 10.1126/science.abb3221. Epub 2020 Mar 16. PMID: 32179701; PMCID: PMC7164387.
- 299.** Report 13 - Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries [Internet]. Imperial College London; 2020 [cited 2021 Nov 25]; Available from: <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/>
- 300.** Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Ichii H, Schubert J, Bania J, Khosrawipour T. Internationally lost COVID-19 cases. *J Microbiol Immunol Infect.* 2020 Jun;53(3):454-458. doi: 10.1016/j.jmii.2020.03.013. Epub 2020 Mar 14. PMID: 32205091; PMCID: PMC7102572.
- 301.** Streeck H. Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-Cluster-Study (Gemeinde Gangelt) [document on the Internet]: Land NRW; 2020 [cited 2021 Nov 24]. Available from: https://www.land.nrw/sites/default/files/asset/document/zwischenergebnis_covid19_case_study_gangelt_0.pdf
- 302.** Sutton D, Fuchs K, D'Alton M, Goffman D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *N Engl J Med.* 2020 May 28;382(22):2163-2164. doi: 10.1056/NEJMc2009316. Epub 2020 Apr 13. PMID: 32283004; PMCID: PMC7175422.
- 303.** Dutch study suggests 3% of population may have coronavirus antibodies [Internet]. Reuters; 2020 [cited 2022 Oct 28]. Available from: <https://www.reuters.com/article/us-health-coronavirus-netherlands-study-idUSKCN21Y102>

- 304.** Zhang W. China: age distribution of novel coronavirus patients 2020 [document on the Internet]: Statista; 2020 [cited 2021 Nov 24]. Available from: <https://www.statista.com/statistics/1095024/china-age-distribution-of-wuhan-coronavirus-covid-19-patients/>
- 305.** Scott D. The Covid-19 risks for different age groups, explained [document on the Internet]: Vox; 2020 [cited 2021 Nov 25]. Available from: <https://www.vox.com/2020/3/23/21190033/coronavirus-covid-19-deaths-by-age>
- 306.** Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W, Cheng C, Tang X, Wu X, Wu Y, Sun B, Huang S, Sun Y, Zhang J, Ma T, Lessler J, Feng T. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis. 2020 Aug;20(8):911-919. doi: 10.1016/S1473-3099(20)30287-5. Epub 2020 Apr 27. Erratum in: Lancet Infect Dis. 2020 Jul;20(7):e148. PMID: 32353347; PMCID: PMC7185944.
- 307.** Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus 2019 (n-CoV) on 23 January 2020 [Internet]. WHO; 2020 [cited 2021 Nov 25]; Available from: [https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
- 308.** Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020 Jul;26(7):1470-1477. doi: 10.3201/eid2607.200282. Epub 2020 Jun 21. PMID: 32255761; PMCID: PMC7323562.
- 309.** Yu CJ, Wang ZX, Xu Y, Hu MX, Chen K, Qin G. Assessment of basic reproductive number for COVID-19 at global level: A meta-analysis. Medicine (Baltimore). 2021 May 7;100(18):e25837. doi: 10.1097/MD.00000000000025837. PMID: 33950996; PMCID: PMC8104145.
- 310.** Coronavirus Death Toll Climbs in China, and a Lockdown Widens [Internet]. The New York Times; 2020 [cited 2022 Oct 28]; Available from: <https://www.nytimes.com/2020/01/23/world/asia/china-coronavirus.html>
- 311.** Philippines Reports First Coronavirus Death Outside China [Internet]. The New York Times; 2020 [cited 2022 Oct 28]; Available from: <https://www.nytimes.com/2020/02/02/world/asia/phippines-coronavirus-china.html>
- 312.** Coronavirus Death in California Came Weeks Before First Known U.S. Death [Internet]. The New York Times; 2020 [cited 2022 Oct 28]; Available from: <https://www.nytimes.com/2020/04/22/us/coronavirus-first-united-states-death.html>
- 313.** Italy's coronavirus deaths could be underestimated in data: Official [Internet]. Reuters; 2020 [cited 2022 Oct 28]. Available from: <https://www.reuters.com/article/us-health-coronavirus-italy-data-idUSKBN21I250>
- 314.** Coronavirus: Is Covid-19 really the cause of all the fatalities in Italy? [Internet]. Stuff; 2020 [cited 2022 Oct 28]. Available from: <https://www.stuff.co.nz/national/health/coronavirus/120443722/coronavirus-is-covid19-really-the-cause-of-all-the-fatalities-in-italy>
- 315.** 28,000 Missing Deaths: Tracking the True Toll of the Coronavirus Crisis [Internet]. The New York Times; 2020 [cited 2022 Oct 28]; Available from:

<https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html?smid=tw-nytimes&smtyp=cur>

316. What 'Excess Deaths' Do and Don't Tell Us About COVID-19 [Internet]. Reason; 2020 [cited 2022 Oct 28]; Available from: <https://reason.com/2020/04/29/what-excess-deaths-do-and-dont-tell-us-about-covid-19/>
317. What do we know about the risk of dying from COVID-19? [Internet]. Our World in Data; 2020 [cited 2021 Dec 4]; Available from: <https://ourworldindata.org/mortality-risk-covid>
318. Levin AT, Hagine WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol. 2020 Dec;35(12):1123-1138. doi: 10.1007/s10654-020-00698-1. Epub 2020 Dec 8. PMID: 33289900; PMCID: PMC7721859.
319. Tracking covid-19 excess deaths across countries [Internet]. The Economist; 2021 [cited 2022 Oct 28]; Available from: <https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker>
320. Covid-19 Vaccination Latest news, daily rates, information and updates [Internet]. Pharmaceutical-Technology; 2021 [cited 2021 Dec 3]; Available from: <https://www.pharmaceutical-technology.com/covid-19-vaccination-tracker/>
321. COVID-19 Coronavirus Pandemic [Internet]. Worldometer; 2021 [cited 2021 Dec 3]; Available from: <https://www.worldometers.info/coronavirus/>
322. Νοημοσύνη - ορισμός [Internet]. Greek-Language; 2020 [cited 2022 Oct 25]; Available from: https://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/search.html?lq=%CE%BD%CE%BF%CE%B7%CE%BC%CE%BF%CF%83%CF%8D%CE%BD%CE%87
323. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine. 2006 Dec 15;27(4):12-.
324. Moor J. The Dartmouth College artificial intelligence conference: The next fifty years. Ai Magazine. 2006 Dec 15;27(4):87-
325. Turing AM. I.-Computing machinery and intelligence. Mind. 1950 Oct 1;LIX(236):433-60.
326. Turing test success marks milestone in computing history [Internet]. University of Reading; 2014 [cited 2022 Oct 28]. Available from: <https://www.reading.ac.uk/news-archive/press-releases/pr583836.html>
327. Russell SJ. Artificial intelligence a modern approach. Pearson Education, Inc.; 2010.
328. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 1943 Dec;5(4):115-33.
329. Zadeh LA. Fuzzy sets. Information and Control, vol. 8, 1965, pp. 338–353.
330. Zadeh LA. Fuzzy Algorithms, Information and Control, Vol. 12, 1968, pp. 94-102.

- 331.** Colmerauer A, Roussel P. The birth of Prolog. InHistory of programming languages---II 1996 Jan 1 (pp. 331-367).
- 332.** Rechenberg I. Evolution strategy: Optimization of technical systems according to the principles of biological evolutionFrommann-Holzboog. Stuttgart, Germany (in German). 1973.
- 333.** Jh H. Adaptation in natural and artificial systems. Ann Arbor. 1975.
- 334.** Minsky M. A Framework For Representing Knowledge. In: P.H.Winston (ed.) The Psychology of Computer Vision. New York: McGraw-Hill; 1975.
- 335.** Johnson-Laird PN, Wason PC, editors. Thinking: Readings in cognitive science. CUP Archive; 1977 Dec 15.
- 336.** Newell A, Simon HA. Computer science as empirical inquiry: Symbols and search. Communications of the ACM. 1976;19(3):113-26.
- 337.** Sugeno M. Fuzzy theory IV (Lecture note). J. of SICE. 1983;22:554-9.
- 338.** McClelland JL, Rumelhart DE, PDP Research Group. Parallel Distributed Processing, Volume 2: Explorations in the Microstructure of Cognition: Psychological and Biological Models. MIT press; 1987 Jul 29.
- 339.** Koza JR. Genetic programming as a means for programming computers by natural selection. Statistics and computing. 1994 Jun;4(2):87-112.
- 340.** Fogel DB. Toward a new philosophy of machine intelligence. IEEE Evolutionary Computation. 1995;1080.
- 341.** King D. Kasparov versus Deeper Blue: The Ultimate man versus machine challenge. ICGA Journal. 1997 Jan 1;20(3):187-.
- 342.** Serpell LC, Radford SE, Otzen DE. AlphaFold: A Special Issue and A Special Time for Protein Science. J Mol Biol. 2021 Oct 1;433(20):167231. doi: 10.1016/j.jmb.2021.167231. Epub 2021 Sep 8. PMID: 34508726.
- 343.** Lupas AN, Pereira J, Alva V, Merino F, Coles M, Hartmann MD. The breakthrough in protein structure prediction. Biochem J. 2021 May 28;478(10):1885-1890. doi: 10.1042/BCJ20200963. PMID: 34029366; PMCID: PMC8166336.
- 344.** Syam N, Sharma A. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Industrial marketing management. 2018 Feb 1;69:135-46.
- 345.** The figure was designed with the use of the free online diagram software diagrams.net. Available from: <https://www.diagrams.net/>
- 346.** Mitchell TM, Mitchell TM. Machine learning. New York: McGraw-hill; 1997 Mar 1.
- 347.** Machine learning Life cycle [Internet]. Javatpoint; 2021 [cited 2022 Oct 28]. Available from: <https://www.javatpoint.com/machine-learning-life-cycle>
- 348.** Javatpoint. Machine learning Life cycle [report on the Internet]. 2021 Figure 1, Machine learning Life cycle. Available from: <https://www.javatpoint.com/machine-learning-life-cycle>

- 349.** Aggarwa CC. Data Classification: Algorithms and Applications. Data Mining and Knowledge Discovery Series. 2015.
- 350.** Kassambara A. Machine learning essentials: Practical guide in R. Sthda; 2018 Mar 10.
- 351.** Simplilearn. Types of Regression [report on the Internet]. 2021 Figure 1, Regression Types. Available from: https://www.simplilearn.com/ice9/free_resources_article_thumb/types-of-regression.jpg
- 352.** KNN Algorithm - Finding Nearest Neighbors [Internet]. Tutorialspoint; 2021 [cited 2022 Oct 28]. Available from: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm
- 353.** Κωνσταντίνος Τσιώμος. Η Τεχνητή Νοημοσύνη και το Blockchain στην εξέλιξη της Βιομηχανικής Αυτοματοποίησης. 2021 May 23. Figure 1.21, Διαφοροποίηση συσταδοποίησης βάσει επιθυμητού πλήθους συστάδων; p.48.
- 354.** Κωνσταντίνος Τσιώμος. Η Τεχνητή Νοημοσύνη και το Blockchain στην εξέλιξη της Βιομηχανικής Αυτοματοποίησης. 2021 May 23. Figure 1.22, Ομαδοποίηση νέας οντότητας με προσέγγιση εγγύτερου γείτονα (3 clusters); p.49.
- 355.** Κωνσταντίνος Τσιώμος. Η Τεχνητή Νοημοσύνη και το Blockchain στην εξέλιξη της Βιομηχανικής Αυτοματοποίησης. 2021 May 23. Figure 1.23, Ομαδοποίηση νέας οντότητας με προσέγγιση κέντρων (3 clusters); p.50.
- 356.** Κωνσταντίνος Τσιώμος. Η Τεχνητή Νοημοσύνη και το Blockchain στην εξέλιξη της Βιομηχανικής Αυτοματοποίησης. 2021 May 23. Figure 1.24, Ομαδοποίηση νέας οντότητας με προσέγγιση εγγύτερου γείτονα (4 clusters); p.50.
- 357.** Κωνσταντίνος Τσιώμος. Η Τεχνητή Νοημοσύνη και το Blockchain στην εξέλιξη της Βιομηχανικής Αυτοματοποίησης. 2021 May 23. Figure 1.25, Ομαδοποίηση νέας οντότητας με προσέγγιση κέντρων (4 clusters); p.51.
- 358.** Utgoff PE. Incremental induction of decision trees. *Machine learning*. 1989 Nov;4(2):161-86.
- 359.** Kotsiantis SB. Decision trees: a recent overview. *Artificial Intelligence Review*. 2013 Apr;39(4):261-83.
- 360.** Rokach L, Maimon O. Decision trees. InData mining and knowledge discovery handbook 2005 (pp. 165-192). Springer, Boston, MA.
- 361.** Github. Decision Tree Model [report on the Internet]. 2021 Figure 1, A typical Decision Tree. Available from: <https://github.com/topics/decision-tree-model>
- 362.** Colin A. Building decision trees with the ID3 algorithm. *Dr Dobbs Journal*. 1996 Jun 1;21(6):107-109.
- 363.** Quinlan JR. C4. 5: programs for machine learning. Elsevier; 2014 Jun 28.
- 364.** Classification Algorithms - Decision Tree [Internet]. Tutorialspoint; 2021 [cited 2022 Oct 28]. Available from: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_decision_tree.htm

- 365.** What Is a Decision Tree and How Is It Used? [Internet]. Careerfoundry; 2021 [cited 2022 Oct 28]. Available from: <https://careerfoundry.com/en/blog/data-analytics/what-is-a-decision-tree/#what-are-decision-trees-used-for>
- 366.** Keller JM, Liu D, Fogel DB. Fundamentals of computational intelligence: neural networks, fuzzy systems, and evolutionary computation. John Wiley & Sons; 2016 Jul 12.
- 367.** Fuzzy Logic - What is fuzzy logic? [Internet]. Techtarget; 2021 [cited 2021 Dec 21]; Available from: <https://www.techtarget.com/searchenterpriseai/definition/fuzzy-logic>
- 368.** Techtarget. Fuzzy Logic - What is fuzzy logic? [report on the internet]. 2021. Figure 1, Comparing the degrees of truth with Boolean logic vs. fuzzy logic. Available from: <https://www.techtarget.com/searchenterpriseai/definition/fuzzy-logic>
- 369.** Fuzzy Logic Tutorial - What is Fuzzy Logic? [Internet]. Javatpoint; 2021 [cited 2022 Dec 21]; Available from: <https://www.javatpoint.com/fuzzy-logic>
- 370.** Javatpoint. Fuzzy Logic Tutorial - What is Fuzzy Logic? [report on the internet]. 2021. Figure 2, The architecture or process of a Fuzzy Logic system. Available from: <https://www.javatpoint.com/fuzzy-logic>
- 371.** Rechenberg I. Evolution Strategy: Optimization of Technical systems by means of biological evolution. Fromman-Holzboog, Stuttgart. 1973;104:15-6.
- 372.** Koza JR. Genetic programming: On the programming of computers by means of natural selection.,(The MIT Press: Cambridge, MA). Genetic programming: On the programming of computers by means of natural selection. The MIT Press, Cambridge, MA.. 1992.
- 373.** Beale R, Jackson T. Neural Computing-an introduction. CRC Press; 1990.
- 374.** Bezdek JC. On the relationship between neural networks, pattern recognition and intelligence. International journal of approximate reasoning. 1992 Feb 1;6(2):85-107.
- 375.** Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review. 1958 Nov;65(6):386.
- 376.** BecomingHuman. Artificial Neuron Networks(Basics) - Introduction to Neural Networks [report on the internet]. 2017. Figure 3, Figure depicting the Activation function for ANN. Available from: <https://becominghuman.ai/artificial-neuron-networks-basics-introduction-to-neural-networks-3082f1dcca8c>
- 377.** TIBCO. What is a Neural Network? [report on the internet]. 2021. Figure 1, Depiction of a neural network with 3 hidden layers. Available from: <https://www.tibco.com/reference-center/what-is-a-neural-network>
- 378.** Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021 Jan 4;4(1):3. doi: 10.1038/s41746-020-00372-6. PMID: 33398013; PMCID: PMC7782717.
- 379.** COVID-19-Government Data - COVID-19 Reservoir. [Internet]. DataGovIL; 2020 [cited 2022 Jan 24]; Available from: <https://data.gov.il/dataset/covid-19>

- 380.** Yazeed Zoabi, Shira Deri-Rozov, Noam Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptomsMachine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021 Jan 4;4(1):3. Table 1, Characteristics of the dataset and the features used by the model in this study; p.4.
- 381.** Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York: springer; 2009 Aug.
- 382.** Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems?. The journal of machine learning research. 2014 Jan 1;15(1):3133-81.
- 383.** Abou Omar KB. XGBoost and LGBM for Porto Seguro's Kaggle challenge: A comparison. Preprint Semester Project. 2018 Jan 20.
- 384.** Chen T, Guestrin C. Xgboost: A scalable tree boosting system. InProceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794).
- 385.** Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems. 2017;30.
- 386.** Raskutti G, Wainwright MJ, Yu B. Early stopping and non-parametric regression: an optimal data-dependent stopping rule. The Journal of Machine Learning Research. 2014 Jan 1;15(1):335-66.
- 387.** Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30.
- 388.** Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, Liston DE, Low DK, Newman SF, Kim J, Lee SI. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature biomedical engineering. 2018 Oct;2(10):749-60.
- 389.** Yazeed Zoabi, Shira Deri-Rozov, Noam Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptomsMachine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021 Jan 4;4(1):3. Figure 1, Model performance. a ROC curves of the predictive model on the prospective test set; p.2.
- 390.** Yazeed Zoabi, Shira Deri-Rozov, Noam Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptomsMachine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021 Jan 4;4(1):3. Figure 1, Model performance. b A plot of the precision (positive predictive value, PPV) against the recall (sensitivity) of the predictor for different thresholds; p.2.
- 391.** Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020 Mar 13;27(2):taaa021. doi: 10.1093/jtm/taaa021. PMID: 32052846; PMCID: PMC7074654.
- 392.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs.

- 393.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Table 4, Some samples in dataset; p.10.
- 394.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Figure 1, A Vital sign measurement System; p.2.
- 395.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Table 2, Results of 4 algorithms NBC, SVM, DTs, LR; p.9.
- 396.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Figure 14, Accuracy curve of LSTM model; p.10.
- 397.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Figure 13, Loss curve of LSTM model; p.9.
- 398.** Han TT, Pham HY, Nguyen DS, Iwata Y, Do TT, Ishibashi K, Sun G. Machine learning based classification model for screening of infected patients using vital signs. *Informatics in Medicine Unlocked*. 2021 Jan 1;24:100592. Table 3, Results of LSTM algorithm; p.10.
- 399.** Kedia P, Anjum, Katarya R. CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Appl Soft Comput*. 2021 Jun;104:107184. doi: 10.1016/j.asoc.2021.107184. Epub 2021 Feb 15. PMID: 33613140; PMCID: PMC7883765.
- 400.** Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014 Sep 4.
- 401.** Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017 (pp. 4700-4708).
- 402.** Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning?. *arXiv preprint arXiv:1608.08614*. 2016 Aug 30.
- 403.** Kedia P, Anjum, Katarya R. CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Appl Soft Comput*. 2021 Jun;104:107184. Figure 2, CoVNet-19 architecture; p.6.
- 404.** Kedia P, Anjum, Katarya R. CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Appl Soft Comput*. 2021 Jun;104:107184. Figure 17, CoVNet-19 confusion matrix (Test Set); p.12.

- 405.** Kedia P, Anjum, Katarya R. CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Appl Soft Comput*. 2021 Jun;104:107184. Figure 7, CoVNet-19 confusion matrix (Test Set); p.10.
- 406.** Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern Biomed Eng*. 2021 Jul-Sep;41(3):867-879. doi: 10.1016/j.bbe.2021.05.013. Epub 2021 Jun 5. PMID: 34108787; PMCID: PMC8179118.
- 407.** Quinlan JR. Induction of decision trees. *Machine learning*. 1986 Mar;1(1):81-106.
- 408.** Breiman L. Random forests. *Machine learning*. 2001 Oct;45(1):5-32.
- 409.** Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*. 1997 Aug 1;55(1):119-39.
- 410.** Breiman L. Bagging predictors. *Machine learning*. 1996 Aug;24(2):123-40.
- 411.** Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*. 2017;30.
- 412.** Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern Biomed Eng*. 2021 Jul-Sep;41(3):867-879. Figure 6, Performance of different ML classifiers on the COVID- 19 pneumonia classification; p.874.
- 413.** Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern Biomed Eng*. 2021 Jul-Sep;41(3):867-879. Table 3, Comparison of classification performance ($\mu \pm \sigma$) of different machine learning models measured by accuracy; p.874.
- 414.** Kumar M, Shakya D, Kurup V, Suksatan W. COVID-19 prediction through X-ray images using transfer learning-based hybrid deep learning approach. *Mater Today Proc*. 2022;51:2520-2524. doi: 10.1016/j.matpr.2021.12.123. Epub 2021 Dec 13. PMID: 34926174; PMCID: PMC8666290.
- 415.** Nair R, Bhagat A. Genes expression classification through histone modification using temporal neural network. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*. 2021 Jul 1;14(5):1488-96.
- 416.** Kumar M, Shakya D, Kurup V, Suksatan W. COVID-19 prediction through X-ray images using transfer learning-based hybrid deep learning approach. *Mater Today Proc*. 2022;51:2520-2524. Figure 1, Flowchart of the proposed work; p.2522.
- 417.** Kumar M, Shakya D, Kurup V, Suksatan W. COVID-19 prediction through X-ray images using transfer learning-based hybrid deep learning approach. *Mater Today Proc*. 2022;51:2520-2524. Table 1, Comparison of other CNN models with HDCNN; p.2523.

- 418.** Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021 Jul;134:104531. doi: 10.1016/j.compbiomed.2021.104531. Epub 2021 May 29. PMID: 34091385; PMCID: PMC8164361.
- 419.** COVID-19 Open Research Dataset Challenge (CORD-19) [Internet]. Kaagle; 2020 [cited 2021 Dec 22]; Available from: <https://www.kaggle.com/allen-institute-for-ai/COVID-19-research-challenge/discussion/139347>
- 420.** Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021 Jul;134:104531. Table 1, Biochemical, urinalysis, haematological, virological, and bacteriological tests performed on the patients included in the study; p.2.
- 421.** Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics*. 1969 Feb 1;11(1):137-48.
- 422.** Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021 Jul;134:104531. Table 3, Performance comparison of the machine learning models for COVID-19; p.5.
- 423.** Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021 Jul;134:104531. Table 4, Important biomarkers in machine learning models for the diagnosis and classification of COVID-19 severity; p.6.
- 424.** Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, Tonin FS, Pontarolo R. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021 Jul;134:104531. Figure 3, ROC curves of the accuracy of the machine learning models; p.5.
- 425.** Rasmussen L, Christensen S, Lenler-Petersen P, Johnsen SP. Anemia and 90-day mortality in COPD patients requiring invasive mechanical ventilation. *Clin Epidemiol*. 2010 Dec 17;3:1-5. doi: 10.2147/CLEP.S12885. PMID: 21326654; PMCID: PMC3035601.
- 426.** Lomholt FK, Laulund AS, Bjarnason NH, Jørgensen HL, Godtfredsen NS. Meta-analysis of routine blood tests as predictors of mortality in COPD. *Eur Clin Respir J*. 2014 Jun 5;1. doi: 10.3402/ecrj.v1.24110. PMID: 26557244; PMCID: PMC4629760.
- 427.** Liu Z, Sun R, Li J, Cheng W, Li L. Relations of Anemia With the All-Cause Mortality and Cardiovascular Mortality in General Population: A Meta-Analysis. *Am J Med Sci*. 2019 Sep;358(3):191-199. doi: 10.1016/j.amjms.2019.05.016. Epub 2019 Jun 8. PMID: 31331612.
- 428.** Bennett TD, Hayward KN, Farris RW, Ringold S, Wallace CA, Brogan TV. Very high serum ferritin levels are associated with increased mortality and critical care in

pediatric patients. *Pediatr Crit Care Med.* 2011 Nov;12(6):e233-6. doi: 10.1097/PCC.0b013e31820abca8. PMID: 21263363.

- 429.** Fox SE, Akmatbekov A, Harbert JL, Li G, Quincy Brown J, Vander Heide RS. Pulmonary and cardiac pathology in African American patients with COVID-19: an autopsy series from New Orleans. *Lancet Respir Med.* 2020 Jul;8(7):681-686. doi: 10.1016/S2213-2600(20)30243-5. Epub 2020 May 27. PMID: 32473124; PMCID: PMC7255143.
- 430.** Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020 Feb 15;395(10223):507-513. doi: 10.1016/S0140-6736(20)30211-7. Epub 2020 Jan 30. PMID: 32007143; PMCID: PMC7135076.
- 431.** Li J, Wang X, Chen J, Zuo X, Zhang H, Deng A. COVID-19 infection may cause ketosis and ketoacidosis. *Diabetes Obes Metab.* 2020 Oct;22(10):1935-1941. doi: 10.1111/dom.14057. Epub 2020 May 18. PMID: 32314455; PMCID: PMC7264681.
- 432.** Chhetri S, Khamis F, Pandak N, Al Khalili H, Said E, Petersen E. A fatal case of COVID-19 due to metabolic acidosis following dysregulate inflammatory response (cytokine storm). *IDCases.* 2020 May 19;21:e00829. doi: 10.1016/j.idcr.2020.e00829. Erratum in: *IDCases.* 2022;30:e00865. PMID: 32483525; PMCID: PMC7236721.
- 433.** Goldman N, Fink D, Cai J, Lee YN, Davies Z. High prevalence of COVID-19-associated diabetic ketoacidosis in UK secondary care. *Diabetes Res Clin Pract.* 2020 Aug;166:108291. doi: 10.1016/j.diabres.2020.108291. Epub 2020 Jun 29. PMID: 32615280; PMCID: PMC7323680.
- 434.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. *Scientific Programming.* 2021 Apr 19;2021:1-10. DOI: 10.1155/2021/5587188.
- 435.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. *Scientific Programming.* 2021 Apr 19;2021:1-10. Table 2, Description of the dataset; p.5.
- 436.** Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research.* 2002 Jun 1;16:321-57.
- 437.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. *Scientific Programming.* 2021 Apr 19;2021:1-10. Figure 2, Correlation of top 20 features in the dataset; p.No.
- 438.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. *Scientific Programming.* 2021 Apr 19;2021:1-10. Table 7, Performance comparison of classifiers using all features (25) using original and SMOTE data; p.8.
- 439.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients.

Scientific Programming. 2021 Apr 19;2021:1-10. Table 8, Performance comparison of classifiers using top 20 features using original and SMOTE data; p.8.

- 440.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming. 2021 Apr 19;2021:1-10. Table 9, Performance comparison of classifiers using top 15 features using original and SMOTE data; p.8.
- 441.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming. 2021 Apr 19;2021:1-10. Table 10, Performance comparison of classifiers using top 10 features using original and SMOTE data; p.9.
- 442.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming. 2021 Apr 19;2021:1-10. Figure 4, ROC curves of random forest using top 20 features; p.9.
- 443.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming. 2021 Apr 19;2021:1-10. Figure 3, ROC curves of logistic regression using top 20 features; p.9.
- 444.** Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming. 2021 Apr 19;2021:1-10. Figure 5, ROC curves of extreme gradient boosting using top 20 features; p.9.
- 445.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. Comput Biol Med. 2021 Nov;138:104915. doi: 10.1016/j.combiomed.2021.104915. Epub 2021 Oct 5. PMID: 34655896; PMCID: PMC8492016.
- 446.** Crooks GE, Hon G, Chandonia J, Brenner SE. WebLogo: A Sequence Logo Generator: Figure 1. Genome Res. 2004 Jun;14(6):1188-90.
- 447.** Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017 Jan 10;1(1):33-46. doi: 10.1002/gch2.1018. PMID: 31565258; PMCID: PMC6607375.
- 448.** Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. Nature. 2020 Mar;579(7798):265-269. doi: 10.1038/s41586-020-2008-3. Epub 2020 Feb 3. Erratum in: Nature. 2020 Apr;580(7803):E7. PMID: 32015508; PMCID: PMC7094943.
- 449.** Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008 Nov 1;9(11).
- 450.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. Comput Biol Med. 2021 Nov;138:104915. Figure 1, t-SNE plot capturing the distribution of the data collected from the GISAID server; p.3.

- 451.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Figure 2, Plot showing the distribution of the sequences in the data; p.3.
- 452.** Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2021 Apr;592(7852):116-121. doi: 10.1038/s41586-020-2895-3. Epub 2020 Oct 26. Erratum in: *Nature*. 2021 Jul;595(7865):E1. PMID: 33106671; PMCID: PMC8158177.
- 453.** Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA*. 2021 Apr 6;325(13):1324-1326. doi: 10.1001/jama.2021.1612. PMID: 33571356; PMCID: PMC7879386.
- 454.** Wise J. Covid-19: The E484K mutation and the risks it poses. *BMJ*. 2021 Feb 5;372:n359. doi: 10.1136/bmj.n359. PMID: 33547053.
- 455.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Table 1, Hotspots found by analyzing the positional entropy; p.4.
- 456.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Table 2, Location of the mutations in the spike protein of the SARS-CoV-2; p.5.
- 457.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Figure 3, Variation of entropy and the position in the spike protein; p.5.
- 458.** Elnaggar A, Heinzinger M, Dallago C, Rehwald G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D, Rost B. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans Pattern Anal Mach Intell*. 2022 Oct;44(10):7112-7127. doi: 10.1109/TPAMI.2021.3095381. Epub 2022 Sep 14. PMID: 34232869.
- 459.** Mannor S, Jin X, Han J, Jin X, Han J, Jin X, Han J, Zhang X, K-means clustering, in: C. Sammut, G.I. Webb (Eds.), *Encyclopedia of Machine Learning*, Springer US, Boston, MA, 2011, pp. 563–564. doi: 10.1007/978-0-387-30164-8_425.
- 460.** Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R. API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*. 2013 Sep 1.
- 461.** Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011 Nov 1;12:2825-30.

- 462.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Table 4, Clusters where the top 3 dominant mutations in the cluster concur with the cluster center mutation; p.5.
- 463.** Mullick B, Magar R, Jhunjhunwala A, Barati Farimani A. Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. *Comput Biol Med*. 2021 Nov;138:104915. Figure 4, A) Clustering analysis for N439K mutation on the spike protein of SARS-CoV-2, B) Clustering analysis for S477 N mutation on the spike protein of SARS-CoV-2; p.6.
- 464.** Ozger ZB, Cihan P. A novel ensemble fuzzy classification model in SARS-CoV-2 B-cell epitope identification for development of protein-based vaccine. *Appl Soft Comput*. 2022 Feb;116:108280. doi: 10.1016/j.asoc.2021.108280. Epub 2021 Dec 15. PMID: 34931117; PMCID: PMC8673934.
- 465.** Future Corporation. COVID-19/SARS B-cell Epitope Prediction - A simple dataset for epitope prediction used in vaccine development [Internet]. Kaggle; 2021 [cited 2021 Dec 22]; Available from: <https://www.kaggle.com/futurecorporation/epitope-prediction>.
- 466.** Ozger ZB, Cihan P. A novel ensemble fuzzy classification model in SARS-CoV-2 B-cell epitope identification for development of protein-based vaccine. *Appl Soft Comput*. 2022 Feb;116:108280. Table 1, Peptid and protein based features at datasets; p.5.
- 467.** Ishibuchi H, Nakashima T, Murata T. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Trans Syst Man Cybern B Cybern*. 1999;29(5):601-18. doi: 10.1109/3477.790443. PMID: 18252338.
- 468.** Ishibuchi H, Yamamoto T, Nakashima T. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Trans Syst Man Cybern B Cybern*. 2005 Apr;35(2):359-65. doi: 10.1109/tsmc.2004.842257. PMID: 15828663.
- 469.** Chi Z, Yan H, Pham T. Fuzzy algorithms: with applications to image processing and pattern recognition. World Scientific; 1996.
- 470.** Ishibuchi H, Nakashima T. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*. 2001 Aug;9(4):506-15.
- 471.** González A, Pérez R. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2001 Jun;31(3):417-25.
- 472.** Ozger ZB, Cihan P. A novel ensemble fuzzy classification model in SARS-CoV-2 B-cell epitope identification for development of protein-based vaccine. *Appl Soft Comput*. 2022 Feb;116:108280. Figure 5, The proposed ensemble fuzzy classification model; p.5.
- 473.** Ozger ZB, Cihan P. A novel ensemble fuzzy classification model in SARS-CoV-2 B-cell epitope identification for development of protein-based vaccine. *Appl Soft Comput*. 2022 Feb;116:108280. Figure 6, Prediction process for SARS-CoV-2; p.6.

- 474.** Ozger ZB, Cihan P. A novel ensemble fuzzy classification model in SARS-CoV-2 B-cell epitope identification for development of protein-based vaccine. *Appl Soft Comput*. 2022 Feb;116:108280. Table 3, Classification errors for SARS-CoV; p.7.
- 475.** BepiPred - 2.0 Prediction of potential linear B-cell epitopes [Internet]. DTU Health; 2022 [cited 2022 Oct 28]; Available from: <https://services.healthtech.dtu.dk/service.php?BepiPred-2.0>
- 476.** VaxiJen - Prediction of Protective Antigens and Subunit Vaccines [Internet]. DDG-Pharmfac; 2022 [cited 2022 Oct 28]; Available from: <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>
- 477.** Datos Abiertos Dirección General de Epidemiología [Internet]. Gobierno de Mexico; 2022 [cited 2022 Jan 11]; Available from: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
- 478.** Cramer JS. The origins of logistic regression. 2002 [cited 2021 Dec 15]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=360300
- 479.** Logistic Regression in Machine Learning [Internet]. Javatpoint; 2021 [cited 2022 Jun 16]; Available from: <https://www.javatpoint.com/logistic-regression-in-machine-learning>
- 480.** Medium. Activation Functions for Deep Learning [Internet]. 2020. Figure 1, The Sigmoid Function. Available from: <https://medium.com/@toprak.mhmt/activation-functions-for-deep-learning-13d8b9b20e>
- 481.** Random Forest Algorithm [Internet]. Javatpoint; 2021 [cited 2022 Jun 16]; Available from: <https://www.javatpoint.com/machine-learning-random-forest-algorithm>
- 482.** Medium. Random Forests: Explained and implemented [Internet]. 2020. Figure 1, The operating procedure of the Random Forest algorithm. Available from: <https://immohann.medium.com/random-forests-explained-and-implemented-fa4a9446a207>
- 483.** Ramraj S, Uzir N, Sunil R, Banerjee S. Experimenting XGBoost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*. 2016;9(40).
- 484.** Neural network models (supervised) [Internet]. Scikit-Learn; 2021 [cited 2022 Jun 16]; Available from: https://scikit-learn.org/stable/modules/neural_networks_supervised.html
- 485.** Scikit-Learn. Neural network models (supervised) [Internet]. 2021. Figure 1, One hidden layer MLP. Available from: https://scikit-learn.org/stable/modules/neural_networks_supervised.html
- 486.** Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*. 1989 Dec 1;57(3):238-47.
- 487.** Kubat M, Kubat. An introduction to machine learning. Cham, Switzerland: Springer International Publishing; 2017 Sep 2.
- 488.** Alpaydin E. Introduction to machine learning. MIT press; 2020 Mar 24.

- 489.** Analytics Vidhya. Introduction to k-Nearest Neighbors: A powerful Machine Learning Algorithm (with implementation in Python & R) [Internet]. 2018. Figure 1, KNN example. Available from: <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>
- 490.** Everitt BS, Landau S, Leese M, Stahl D. Cluster analysis 5th ed; 2011.
- 491.** Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures How to evaluate the performance of a model in Azure ML and understanding 'Confusion Metrics' [Internet]. Exsilio; 2016 [cited 2022 Jun 20]; Available from: <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>
- 492.** What is a Confusion Matrix in Machine Learning [Internet]. Machine Learning Mastery; 2016 [cited 2022 Jun 21]; Available from: <https://machinelearningmastery.com/confusion-matrix-machine-learning/>
- 493.** Classification: ROC Curve and AUC [Internet]. Google Developers; 2020 [cited 2022 Jul 23]; Available from: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
- 494.** Google Developers. Classification: ROC Curve and AUC [Internet]. 2022. Figure 5, AUC (Area under the ROC Curve. Available from: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
- 495.** Google Developers. Classification: ROC Curve and AUC [Internet]. 2022. Figure 6. Predictions ranked in ascending order of logistic regression score. Available from: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
- 496.** Josephus BO, Nawir AH, Wijaya E, Moniaga JV, Ohyver M. Predict Mortality in Patients Infected with COVID-19 Virus Based on Observed Characteristics of the Patient using Logistic Regression. Procedia Comput Sci. 2021;179:871-877. doi: 10.1016/j.procs.2021.01.076. Epub 2021 Feb 19. PMID: 33643495; PMCID: PMC7894086.
- 497.** Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. medRxiv 2020. Google Scholar. doi: 10.1101/2020.03.30.20047308.
- 498.** Naseem M, Arshad H, Hashmi SA, Irfan F, Ahmed FS. Predicting mortality in SARS-COV-2 (COVID-19) positive patients in the inpatient setting using a novel deep neural network. Int J Med Inform. 2021 Oct;154:104556. doi: 10.1016/j.ijmedinf.2021.104556. Epub 2021 Aug 21. PMID: 34455118; PMCID: PMC8378987.
- 499.** Bárcenas R, Fuentes-García R. Risk assessment in COVID-19 patients: A multiclass classification approach. Inform Med Unlocked. 2022;32:101023. doi: 10.1016/j.imu.2022.101023. Epub 2022 Jul 19. PMID: 35873009; PMCID: PMC9295315.
- 500.** Yan L, Zhang HT, Xiao Y, Wang M, Sun C, Liang J, Li S, Zhang M, Guo Y, Xiao Y, Tang X. Prediction of survival for severe Covid-19 patients with three clinical

features: development of a machine learning-based prognostic model with clinical data in Wuhan. medRxiv. 2020 Mar. doi: 10.1101/2020.02.27.20028027.

Παραρτήματα

Παράρτημα Α: Πίνακας 1 - Κρούσματα και Θάνατοι ανά χώρα ή επικράτεια έως τις 3 Δεκεμβρίου 2021.[319]

#	Country, Other	Total Cases	Total Deaths	Total Recovered	Tot Cases / 1M pop	Deaths/ 1M pop	Population
	World	265,875,654	5,266,834	239,562,126	34,109	675.7	
1	USA	49,934,791	808,608	39,498,111	149,61	2,423	333,765,397
2	India	34,633,255	470,62	34,060,774	24,75	336	1,399,309,544
3	Brazil	22,138,247	615,606	21,359,352	103,108	2,867	214,709,861
4	UK	10,421,104	145,551	9,188,131	152,37	2,128	68,393,618
5	Russia	9,801,613	281,278	8,502,406	67,124	1,926	146,023,563
6	Turkey	8,881,760	77,645	8,412,434	103,722	907	85,630,260
7	France	7,875,012	119,506	7,168,139	120,267	1,825	65,479,150
8	Germany	6,156,061	103,554	5,082,500	73,142	1,23	84,165,340
9	Iran	6,134,465	130,2	5,923,316	71,73	1,522	85,522,195
10	Argentina	5,339,382	116,643	5,200,141	116,616	2,548	45,785,918
11	Spain	5,202,958	88,159	4,927,391	111,221	1,885	46,780,564
12	Italy	5,094,072	134,152	4,736,202	84,429	2,223	60,335,220
13	Colombia	5,078,987	128,733	4,918,619	98,326	2,492	51,654,519
14	Indonesia	4,257,685	143,867	4,106,292	15,335	518	277,638,777
15	Mexico	3,900,293	295,154	3,255,802	29,805	2,255	130,860,095
16	Poland	3,671,421	85,675	3,142,265	97,159	2,267	37,787,584
17	Ukraine	3,497,477	88,28	3,050,659	80,662	2,036	43,359,522
18	South Africa	3,020,569	89,965	2,855,474	50,032	1,49	60,372,382

19	Philippines	2,834,775	49,386	2,771,536	25,389	442	111,655,624
20	Netherlands	2,751,954	19,668	2,158,065	160,101	1,144	17,188,851
21	Malaysia	2,654,474	30,574	2,561,232	80,546	928	32,955,873
22	Peru	2,242,646	201,36	N/A	66,696	5,988	33,625,115
23	Czechia	2,240,721	33,665	1,914,329	208,686	3,135	10,737,268
24	Thailand	2,141,241	20,942	2,048,815	30,568	299	70,049,337
25	Iraq	2,084,346	23,885	2,050,033	50,219	575	41,504,977
26	Belgium	1,827,467	27,167	1,358,257	156,71	2,33	11,661,424
27	Canada	1,805,070	29,766	1,746,170	47,237	779	38,213,055
28	Romania	1,786,036	57,099	1,701,727	93,731	2,997	19,054,895
29	Chile	1,772,547	38,501	1,665,361	91,606	1,99	19,349,711
30	Japan	1,727,828	18,364	1,708,638	13,721	146	125,925,054
31	Bangladesh	1,577,443	28,001	1,542,274	9,444	168	167,023,961
32	Israel	1,345,581	8,204	1,331,900	144,283	880	9,326,000
33	Vietnam	1,309,092	26,26	1,009,277	13,279	266	98,584,854
34	Pakistan	1,286,825	28,767	1,245,606	5,669	127	226,988,925
35	Serbia	1,263,663	11,917	1,210,147	145,462	1,372	8,687,247
36	Sweden	1,212,145	15,136	1,161,927	118,967	1,486	10,188,897
37	Austria	1,198,478	12,796	1,075,482	131,998	1,409	9,079,544
38	Portugal	1,166,787	18,537	1,086,887	114,905	1,826	10,154,367
39	Hungary	1,134,869	35,122	910,745	117,903	3,649	9,625,418
40	Switzerland	1,039,730	11,612	870,684	118,898	1,328	8,744,699
41	Kazakhstan	975,15	12,769	941,285	51,065	669	19,096,213
42	Jordan	971,401	11,715	899,555	93,88	1,132	10,347,257

43	Cuba	963,002	8,307	954,03	85,095	734	11,316,771
44	Greece	962,695	18,516	868,028	93,009	1,789	10,350,540
45	Morocco	950,501	14,787	932,835	25,324	394	37,532,940
46	Georgia	865,293	12,344	806,596	217,505	3,103	3,978,262
47	Nepal	822,592	11,541	804,544	27,528	386	29,881,825
48	UAE	742,328	2,148	737,33	73,787	214	10,060,396
49	Slovakia	729,475	14,826	595,369	133,52	2,714	5,463,407
50	Tunisia	718,278	25,398	691,522	59,884	2,117	11,994,571
51	Bulgaria	702,454	28,805	573,112	102,191	4,19	6,873,938
52	Lebanon	677,147	8,765	634,994	99,846	1,292	6,781,912
53	Belarus	663,813	5,162	652,874	70,283	547	9,444,897
54	Croatia	631,037	11,218	590,341	155,076	2,757	4,069,207
55	Guatemala	619,542	15,98	602,343	33,695	869	18,386,681
56	Azerbaijan	596,388	7,967	564,577	58,075	776	10,269,223
57	Ireland	589,094	5,707	454,362	117,437	1,138	5,016,257
58	Sri Lanka	567,522	14,461	542,326	26,346	671	21,541,060
59	Costa Rica	567,383	7,312	552,158	109,951	1,417	5,160,311
60	Saudi Arabia	549,912	8,844	539,056	15,454	249	35,584,863
61	Bolivia	542,859	19,215	497,115	45,62	1,615	11,899,498
62	Ecuador	526,87	33,25	443,88	29,232	1,845	18,024,044
63	Myanmar	524,407	19,141	500,304	9,547	348	54,928,945
64	Denmark	509,111	2,946	447,152	87,457	506	5,821,240
65	Lithuania	479,839	6,847	444,556	179,838	2,566	2,668,170
66	Panama	478,831	7,373	468,661	108,545	1,671	4,411,353
67	S. Korea	473,034	3,852	407,175	9,215	75	51,332,010
68	Paraguay	463,315	16,477	446,079	63,842	2,27	7,257,171
69	Venezuela	434,133	5,186	421,495	15,329	183	28,321,197
70	Palestine	431,863	4,549	423,981	81,949	863	5,269,899
71	Slovenia	430,064	5,303	395,415	206,827	2,55	2,079,345
72	Kuwait	413,464	2,465	410,701	94,823	565	4,360,391
73	Dominican Republic	408,245	4,212	401,716	37,104	383	11,002,693
74	Uruguay	400,821	6,132	392,388	114,82	1,757	3,490,868
75	Mongolia	383,543	2,016	313,256	114,376	601	3,353,359

76	Honduras	378,251	10,413	122,039	37,345	1,028	10,128,610
77	Libya	374,989	5,493	355,121	53,538	784	7,004,191
78	Ethiopia	372,09	6,794	349,557	3,127	57	119,008,669
79	Moldova	366,162	9,208	350,192	91,071	2,29	4,020,611
80	Egypt	362,26	20,682	300,437	3,448	197	105,078,062
81	Armenia	340,723	7,683	321,106	114,678	2,586	2,971,129
82	Oman	304,603	4,113	300,039	57,569	777	5,291,097
83	Norway	282,918	1,093	88,952	51,613	199	5,481,569
84	Bahrain	277,803	1,394	276,096	155,589	781	1,785,492
85	Bosnia and Herzegovina	277,228	12,687	192,218	85,255	3,902	3,251,752
86	Singapore	268,659	746	258,038	45,414	126	5,915,821
87	Latvia	257,329	4,267	241,683	138,611	2,298	1,856,489
88	Kenya	255,408	5,335	248,404	4,606	96	55,454,327
89	Qatar	244,223	611	241,459	86,98	218	2,807,805
90	Estonia	224,993	1,823	207,85	169,447	1,373	1,327,805
91	Australia	217,843	2,05	196,948	8,404	79	25,920,795
92	North Macedonia	217,301	7,641	203,574	104,309	3,668	2,083,252
93	Nigeria	214,567	2,98	207,427	1,005	14	213,397,144
94	Algeria	211,297	6,096	145,054	4,698	136	44,976,741
95	Zambia	210,312	3,667	206,466	11,002	192	19,115,024
96	Albania	201,402	3,109	191,855	70,094	1,082	2,873,322
97	Botswana	195,302	2,419	192,452	80,733	1	2,419,112
98	Uzbekistan	194,333	1,418	190,939	5,689	42	34,159,056
99	Finland	191,226	1,36	46	34,437	245	5,552,975
100	Kyrgyzstan	183,615	2,758	178,602	27,499	413	6,677,245
101	Montenegro	158,466	2,329	154,027	252,262	3,708	628,179
102	Afghanistan	157,445	7,312	140,721	3,919	182	40,174,039
103	Mozambique	151,868	1,941	149,575	4,673	60	32,498,836
104	Zimbabwe	138,523	4,709	128,858	9,131	310	15,170,633

105	Cyprus	136,525	601	124,37	111,918	493	1,219,868
106	Ghana	130,92	1,209	129,042	4,092	38	31,994,722
107	Namibia	129,638	3,573	125,523	49,74	1,371	2,606,322
108	Uganda	127,655	3,254	97,627	2,671	68	47,791,313
109	Cambodia	120,256	2,96	116,609	7,054	174	17,048,061
110	El Salvador	119,803	3,783	102,982	18,339	579	6,532,792
111	Cameroon	107,148	1,804	102,716	3,898	66	27,484,492
112	Rwanda	100,404	1,343	45,522	7,488	100	13,408,294
113	China	99,142	4,636	93,47	69	3	1,439,323,776
114	Maldives	92,095	254	90,103	166,213	458	554,079
115	Jamaica	91,421	2,407	62,824	30,683	808	2,979,566
116	Luxembourg	90,774	880	84,834	141,733	1,374	640,459
117	Laos	79,833	207	7,339	10,751	28	7,425,709
118	Trinidad and Tobago	74,533	2,236	60,202	53,013	1,59	1,405,940
119	Senegal	74,024	1,886	72,103	4,261	109	17,370,585
120	Angola	65,244	1,735	63,298	1,901	51	34,317,436
121	Malawi	61,966	2,307	58,816	3,125	116	19,832,046
122	Ivory Coast	61,817	705	60,844	2,264	26	27,306,143
123	Réunion	61,188	384	57,781	67,655	425	904,406
124	DRC	58,319	1,107	50,93	624	12	93,426,585
125	Guadeloupe	55,284	748	2,25	138,134	1,869	400,221
126	Fiji	52,543	697	51,116	58,014	770	905,69
127	Suriname	50,993	1,17	29,577	85,831	1,969	594,112
128	Syria	48,538	2,772	29,624	2,681	153	18,104,589
129	Eswatini	47,368	1,248	45,263	40,236	1,06	1,177,255
130	French Polynesia	46,323	636	N/A	163,559	2,246	283,219
131	French Guiana	46,142	328	11,254	149	1,059	309,678
132	Martinique	45,501	718	104	121,384	1,915	374,851

133	Madagascar	44,8	972	43,119	1,561	34	28,702,121
134	Sudan	43,489	3,164	35,491	960	70	45,303,820
135	Malta	39,934	468	37,761	90,1	1,056	443,217
136	Mauritania	39,51	838	37,828	8,192	174	4,823,235
137	Cabo Verde	38,439	351	37,968	68,084	622	564,579
138	Guyana	38,118	1,004	36,17	48,132	1,268	791,955
139	Gabon	37,477	280	33,094	16,292	122	2,300,346
140	Papua New Guinea	35,39	550	34,322	3,852	60	9,187,938
141	Guinea	30,77	387	29,725	2,256	28	13,638,663
142	Belize	30,68	579	28,979	75,216	1,42	407,89
143	Togo	26,291	243	25,921	3,073	28	8,554,386
144	Tanzania	26,27	730	N/A	423	12	62,150,021
145	Barbados	25,964	234	23,253	90,192	813	287,876
146	Haiti	25,51	747	21,644	2,199	64	11,600,267
147	Benin	24,863	161	24,546	1,977	13	12,574,185
148	Seychelles	23,537	127	22,912	237,235	1,28	99,214
149	Somalia	23,051	1,331	12,325	1,395	81	16,522,844
150	Bahamas	22,825	695	21,615	57,264	1,744	398,591
151	Mauritius	22,09	455	20,266	17,329	357	1,274,773
152	Lesotho	21,838	663	13,741	10,08	306	2,166,417
153	Mayotte	21,003	185	2,964	74,435	656	282,166
154	Burundi	20,439	38	773	1,649	3	12,393,539
155	Timor-Leste	19,828	122	19,701	14,643	90	1,354,140
156	Congo	18,97	354	12,421	3,322	62	5,711,024
157	Channel Islands	18,893	104	16,459	107,259	590	176,143
158	Iceland	18,333	35	16,721	53,232	102	344,395
159	Andorra	18,01	132	16,162	232,564	1,705	77,441
160	Mali	17,812	617	15,223	845	29	21,078,985
161	Curaçao	17,47	179	17,19	105,848	1,085	165,048
162	Nicaragua	17,254	210	4,225	2,561	31	6,736,677
163	Tajikistan	17,095	124	16,966	1,737	13	9,842,030

164	Taiwan	16,652	848	15,648	697	36	23,878,150
165	Aruba	16,432	174	16,074	152,982	1,62	107,411
166	Burkina Faso	16	286	15,345	737	13	21,717,221
167	Brunei	15,186	98	14,773	34,246	221	443,44
168	Equatorial Guinea	13,599	175	13,346	9,261	119	1,468,397
169	Djibouti	13,508	187	13,294	13,395	185	1,008,433
170	Saint Lucia	13,027	282	12,636	70,484	1,526	184,822
171	South Sudan	12,758	133	12,463	1,121	12	11,380,424
172	Hong Kong	12,462	213	12,149	1,643	28	7,583,538
173	New Caledonia	12,249	279	11,746	42,325	964	289,403
174	New Zealand	12,195	44	5,834	2,438	9	5,002,100
175	Isle of Man	11,861	66	10,881	138,445	770	85,673
176	CAR	11,742	101	6,859	2,373	20	4,949,155
177	Yemen	10,025	1,954	6,913	326	64	30,759,385
178	Gambia	9,992	342	9,64	3,976	136	2,513,072
179	Cayman Islands	7,529	7	3,57	112,676	105	66,82
180	Eritrea	7,458	61	7,24	2,062	17	3,616,232
181	Gibraltar	7,371	99	6,955	218,873	2,94	33,677
182	Niger	7,061	261	6,693	277	10	25,449,666
183	Guinea-Bissau	6,444	149	6,275	3,168	73	2,034,082
184	Sierra Leone	6,402	121	N/A	780	15	8,207,535
185	Dominica	6,12	39	5,712	84,716	540	72,241
186	San Marino	5,979	93	5,61	175,682	2,733	34,033
187	Grenada	5,907	200	5,632	52,152	1,766	113,265
188	Liberia	5,824	287	5,523	1,114	55	5,226,567
189	Bermuda	5,751	106	5,615	92,833	1,711	61,95
190	Chad	5,701	181	4,874	333	11	17,095,083
191	St. Vincent Grenadines	5,579	74	5,077	50,062	664	111,442

192	Liechtenstein	4,893	62	4,357	127,808	1,619	38,284
193	Sint Maarten	4,598	75	4,504	105,536	1,721	43,568
194	Comoros	4,542	150	4,313	5,07	167	895,925
195	Antigua and Barbuda	4,147	117	4,019	41,854	1,181	99,083
196	Saint Martin	3,973	56	1,399	100,318	1,414	39,604
197	Monaco	3,879	36	3,691	97,861	908	39,638
198	Faeroe Islands	3,796	13	3,263	77,269	265	49,127
199	Sao Tome and Principe	3,732	56	3,675	16,591	249	224,944
200	Turks and Caicos	3,103	25	3,049	78,633	634	39,462
201	Caribbean Netherlands	3,026	22	2,704	113,888	828	26,57
202	British Virgin Islands	2,809	38	N/A	92,047	1,245	30,517
203	Saint Kitts and Nevis	2,79	28	2,749	51,921	521	53,735
204	Bhutan	2,641	3	2,625	3,37	4	783,696
205	St. Barth	1,603	6	N/A	161,593	605	9,92
206	Greenland	1,582		1,261	27,798		56,91
207	Anguilla	1,414	3	1,323	93,063	197	15,194
208	Diamond Princess	712	13	699			
209	Wallis and Futuna	454	7	438	41,431	639	10,958
210	Saint Pierre Miquelon	86		32	14,946		5,754
211	Falkland Islands	83		N/A	22,903		3,624
212	Macao	77		77	116		661,925
213	Montserrat	44	1	41	8,807	200	4,996
214	Vatican City	27		27	33,582		804

215	<u>Solomon Islands</u>	20		20	28		<u>710,812</u>
216	<u>Western Sahara</u>	10	1	8	16	2	<u>618,201</u>
217	<u>MS Zaandam</u>	9	2	7			
218	<u>Palau</u>	8		8	439		<u>18,217</u>
219	<u>Vanuatu</u>	6	1	6	19	3	<u>317,342</u>
220	<u>Marshall Islands</u>	4		4	67		<u>59,758</u>
221	<u>Samoa</u>	3		3	15		<u>200,29</u>
222	<u>Saint Helena</u>	2		2	328		<u>6,103</u>
223	<u>Micronesia</u>	1		1	9		<u>116,738</u>
224	<u>Tonga</u>	1		1	9		<u>107,399</u>
	Total:	265,875,65 4	5,266,83 4	239,562,12 6	34,109.4	675.7	