



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ &
ΠΛΗΡΟΦΟΡΙΚΗΣ**

ΤΟΜΕΑΣ Λογικού των Υπολογιστών

Εργαστήριο Βάσεων Δεδομένων

**Κατηγοριοποίηση Ιστοπαθολογικών Εικόνων με
χρήση διαφορετικών προ-εκπαιδευμένων
Συνελικτικών Δικτύων**

Προπτυχιακή Διπλωματική Εργασία

του

Ορέστη Φακωτάκη

(Α. Μ.: 235673)

Επιβλέπων: Μεγαλοοικονόμου Βασίλειος, Καθηγητής

Μέλη Επιτροπής: Ψαράκης Εμμανουήλ, Αναπληρωτής Καθηγητής

Μακρής Χρήστος, Αναπληρωτής Καθηγητής

Πάτρα, 2020

Ευχαριστίες

Η εκπόνηση της διπλωματικής εργασίας, με τίτλο *«Κατηγοριοποίηση Ιστοπαθολογικών Εικόνων με χρήση διαφορετικών προ-εκπαιδευμένων συνελκτικών δικτύων»* έγινε το διάστημα 5/2019 – 06/2020 στο Εργαστήριο Ανάλυσης Πολυδιάστατων Δεδομένων και Διαχείρισης Γνώσης, του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Πατρών, υπό την επίβλεψη του καθηγητή κ. Βασιλείου Μεγαλοοικονόμου.

Θα ήθελα σε αυτό το σημείο να ευχαριστήσω θερμά τον καθηγητή μου, κ. Βασίλειο Μεγαλοοικονόμου για την ευκαιρία που μου έδωσε με την ανάθεση αυτής της εργασίας. Η συνεργασία μας ήταν άριστη και ιδιαίτερα επικοινωνιακή και μέσα από αυτή κέρδισα πολύτιμες εμπειρίες που θα με συνοδεύσουν στην μελλοντική μου καριέρα ως μηχανικός ηλεκτρονικών υπολογιστών και πληροφορικής.

Ακόμα, ευχαριστώ θερμά το υπ. διδάκτορα κ. Θωμά Παπαστεργίου για την καθοδήγηση που μου προσέφερε σε όλη τη διάρκεια της εργασίας. Τέλος, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής, κυρίους Εμμανουήλ Ψαράκη, Αναπληρωτή Καθηγητή και Χρήστο Μακρή, Αναπληρωτή Καθηγητή, για την προθυμία τους να με βοηθήσουν σε κάθε δυσκολία που συνάντησα στην διαδικασία εκπόνησης της διπλωματικής μου εργασία.

Περίληψη

Η παρούσα διπλωματική εργασία, αναφέρεται στην κατηγοριοποίηση Ιστοπαθολογικών Εικόνων με τη χρήση διαφόρων μοντέλων Συνελκτικών Νευρωνικών Δικτύων. Σκοπός της εργασίας αποτέλεσε η μελέτη της χρήσης των νευρωνικών δικτύων και ιδιαίτερα των τεχνικών βαθιάς μάθησης, στην αυτόματη κατηγοριοποίηση (ταξινόμηση) ιατρικών εικόνων. Οι διαδικασίες ταξινόμησης που δοκιμάστηκαν βασίζονται σε τεχνικές βαθιών νευρωνικών δικτύων (DNN) και συγκεκριμένα των Συνελκτικών Νευρωνικών Δικτύων (ΣΝΔ). Σήμερα, τα δίκτυα αυτά εντάσσονται στο state-of-the-art των τεχνικών αναγνώρισης προτύπων και ιδιαίτερα στις τεχνικές που αφορούν ταξινόμηση ιατρικών δεδομένων (εικόνων).

Τα ΣΝΔ υπάρχουν σε διάφορες αρχιτεκτονικές, με διαφορετικό βάθος, διαφορετικό πλήθος συνελίξεων, διαφορετικές συναρτήσεις ενεργοποίησης και πολλές άλλες παραμέτρους και τεχνολογίες. Στην παρούσα εργασία εξετάσαμε αρχιτεκτονικές που παρουσιάζουν ξεχωριστή συμπεριφορά στην ταξινόμηση ιατρικών δεδομένων. Από αυτές επιλέξαμε οκτώ τις οποίες δοκιμάσαμε κατά την πειραματική διαδικασία: (*Vgg16*, *Vgg19*, *ResNet50*, *ResNet101*, *ResNet152*, *MobileNet*, *MobileNetV2*, *DenseNet121*).

Για την εκπαίδευση και τον έλεγχο των συστημάτων, ως ιατρικά δεδομένα, χρησιμοποιήθηκαν ιστοπαθολογικές εικόνες καρκίνου του μαστού οι οποίες ταξινομήθηκαν (διαχωρίστηκαν) σε θετικές (κακοήγη) και αρνητικές (καλοήγη), ως προς την ύπαρξη όγκου. Συγκεκριμένα, χρησιμοποιήθηκε η Ιστοπαθολογική Βάση Δεδομένων, καρκίνου του μαστού (BreaKHis) που περιέχει συνολικά 7.909 εικόνες, που συλλέχθηκαν από 82 ασθενείς: 58 με καρκίνο (5.429 κακοήγη δείγματα) και 24 με καλοήγη νεοπλασία (2.480 καλοήγη). Η ανάλυση των εικόνων είναι των 700x460 εικονοστοιχείων, RGB 3 καναλιών, βάθους 8-bit (κάθε κανάλι). Οι εικόνες στην βάση, είναι χωρισμένες (περίπου ισομερώς) ανάλογα με την μεγέθυνση του μικροσκοπίου σε 4 κατηγορίες: 40x, 100x, 200x και 400x. Να σημειωθεί ότι όλα τα ΣΝΔ που χρησιμοποιήσαμε ήταν αρχικά εκπαιδευμένα (προεκπαίδευση) στον τεράστιο όγκο των δεδομένων της ImageNet.

Τα οκτώ αυτά δίκτυα, εκπαιδεύτηκαν και ελέγχθηκαν σε τρεις πειραματικές διαδικασίες: μια χωρίς προεπεξεργασία, μία με προεπεξεργασία ZCA, και μια με προεπεξεργασία PCA. Τέλος, όλα τα πειράματα επαναλήφθηκαν τέσσερις φορές, που αντιστοιχούν στις διαφορετικές κατηγορίες δεδομένων εστίασης (40x, 100x, 200x, 400x). Έτσι, το σύνολο των πειραμάτων που διεξήχθησαν ήταν 96 πειράματα, και τα αποτελέσματα αναρτώνται σε 28 συγκεντρωτικούς πίνακες και 8 διπλά σχήματα.

Τα οκτώ δίκτυα που ελέγξαμε συγκρίθηκαν μεταξύ τους και αβίαστα καταλήξαμε στο συμπέρασμα ότι δίκτυο *DenseNet121* έχει την καλλίτερη απόδοση. Ήταν κάτι αναμενόμενο, αφού αυτό αποτελεί εξέλιξη των δικτύων Resnet. Ενδιαφέρον ακόμα είχε το γεγονός ότι εικόνες με χαμηλή εστίαση έδωσαν καλλίτερα αποτελέσματα από αυτές με υψηλή εστίαση.

Στα πειράματα, όπου στα δίκτυα είχε προηγηθεί προεπεξεργασία, με την τεχνική ZCA και την PCA, τα αποτελέσματα, δεν θα μπορούμε να τα χαρακτηρίσουμε αναμενόμενα. Διαπιστώσαμε βελτιώσεις με βάση τις μετρικές, ωστόσο αυτές οι βελτιώσεις δεν ακολούθησαν μια συγκεκριμένη λογική. Η εξήγησή που μπορούμε να δώσουμε, για τα αποτελέσματα αυτά, είναι ο περιορισμένος όγκος των δεδομένων εκπαίδευσης και ελέγχου, δεν επέτρεψαν να αναδειχθεί η σημαντική συνεισφορά του σταδίου της προεπεξεργασίας.

Abstract

This thesis has as a main subject the categorization of Histopathological Images, using various models of Convolutional Neural Networks. The purpose of this study was to examine the use of machine learning and, specifically, deep learning approaches, in the automatic categorization (classification) of medical images. The classification procedures are tested, based on deep neural networks (DNN) and, mainly, Convolutional Neural Networks (CNS).

In the present time, these networks belong to the scientific field of pattern recognition, constituting the state-of-the-art and they are techniques for classifying images with medical data. CNNs exist in different architectures, with different depth, multitude of convolutions, activation functions and many other parameters and technologies. In this paper, we examined architectures that exhibit a distinct behavior in the classification of medical data. From these, we selected the predominant 8, which we tested during the experimental process: (Vgg16, Vgg19, ResNet50, ResNet101, ResNet152, MobileNet, MobileNetV2, DenseNet121).

For the training and control of the systems, histopathological images of breast cancer were used as medical data, which were classified (divided) into positive (malignant) and negative (benign), for tumor recognition. Specifically, the Histopathological Database, Breast Cancer (BreCaHis) was used, containing a total of 7,909 images, collected from 82 patients: 58 with cancer (5,429 malignancies) and 24 with benign neoplasms (2,480 benign). Image resolution is 700x460 pixels, RGB 3 channels, 8-bit depth (each channel). The images in the database are divided (approximately equally), depending on the magnification of the microscope into 4 categories: 40x, 100x, 200x and 400x.

Note that all CNNs we used were initially trained (pre-trained), in the vast amount of ImageNet data. These eight networks were trained and tested in three experimental procedures: one without pre-treatment, one with ZCA pre-treatment, and one with PCA pre-treatment. Finally, all experiments were repeated four times, corresponding to the different categories of focus data (40x, 100x, 200x, 400x). Thus, the total number of experiments carried out was 96 experiments, and the results are posted in 28 confusion matrices and 8 figures.

The eight networks that we tested were compared with each other and we easily came to the conclusion that DenseNet121 network has the best performance. It was expected, as this is a development of Resnet networks. It was also interesting to note that low-focus images performed better results, than high-focus images. In the experiments, where the networks has been pre-processed, with the ZCA technique and the PCA, the results are not the expected.

We found many improvements based on the metrics; however, these improvements did not follow a specific logic. The explanation we can give for these results, is the limited volume of training and test data, which did not allow the significant contribution of the pre-processing stage to emerge.

Περιεχόμενα

ΠΡΟΛΟΓΟΣ.....	1
1 Καρκίνος του Μαστού.....	3
1.1 Εισαγωγή.....	3
1.2 Ανατομία του Μαστού	4
1.3 Καρκίνος του Μαστού.....	5
1.3.1 Αίτια	5
1.3.2 Συμπτώματα	6
1.3.3 Διάγνωση.....	6
1.3.4 Θεραπεία.....	11
2 Τεχνικές Ταξινόμησης.....	12
2.1 Εισαγωγή.....	12
2.2 Βάσεις Δεδομένων Ψηφιακών Εικόνων	13
2.2.1 Ψηφιακή Εικόνα	13
2.2.2 Βάση Δεδομένων Ιστοπαθολογικών Εικόνων	17
2.3 Τεχνικές Εντοπισμού Περιοχών Ενδιαφέροντος.....	18
2.3.1 Εξισορρόπηση Ιστογράμματος (Histogram Equalization)	18
2.3.2 Κατωφλίωση (Binarization)	19
2.3.3 Συσταδοποίηση (Clustering)	19
2.3.4 Μετασχηματισμός Laplace.....	19
2.3.5 Σύγκριση των Σκιάσεων.....	19
2.4 Μέθοδοι Ταξινόμησης.....	20
2.4.1 Ταξινόμηση Υφής (Texture Classification).....	20
2.4.2 Ταξινόμηση Naïve Bayes	21
2.4.3 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines).....	21
2.4.4 Τεχνικές Εξόρυξης Δεδομένων (Data Mining)	21
2.4.5 Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)	22
3 Τεχνητή Νοημοσύνη	23
3.1 Εισαγωγή.....	23
3.2 Εξελικτική Πορεία της Επιστήμης της ΤΝ.....	24
3.3 Μηχανική Μάθηση.....	26
3.4 Νευρωνικά Δίκτυα.....	26
3.4.1 Βιολογικοί Νευρώνες	26
3.4.2 Τεχνητά Νευρωνικά Δίκτυα	28
3.4.3 Διαδικασία Εκπαίδευσης (Μάθησης) Νευρωνικών Δικτύων	34

3.4.4 Σύγχρονες Τάσεις στα Τεχνητά Νευρωνικά Δίκτυα	34
3.4.5 Βαθιά Μάθηση (Deep Learning)	35
3.4.6 Μεγάλα Δεδομένα (Big Data)	36
4 Συνελκτικά Νευρωνικά Δίκτυα	39
4.1 Εισαγωγή	39
4.2 Δομή των Συνελκτικών Νευρωνικών Δικτύων	39
4.2.1 Συνελκτικά Επίπεδα (Convolutional Layers)	40
4.2.2 Συγκεντρωτικά Επίπεδα (Pooling layers)	41
4.2.3 Μη-Γραμμικά Επίπεδα (Non Linearity)	42
4.2.4 Πλήρως Συνδεδεμένα Επίπεδα (Fully Connected Layers)	43
4.3 Τεχνικές Προσαρμογής των Δεδομένων στο Σύστημα	44
4.3.1 Κανονικοποίηση (Regularization)	44
4.3.2 Εγκατάλειψη (Dropout)	44
4.3.3 Επαύξηση των Δεδομένων (Data Augmentation)	45
4.3.4 Batch Normalization	45
4.3.5 Principal Components Analysis (PCA)	46
4.3.6 Zero Components Analysis (ZCA)	46
4.3.7 Overfitting, Underfitting	47
4.3.8 Πρόωρη Διακοπή (Early stopping)	48
4.3.9 Μεταφερόμενη Μάθηση και Συντονισμός	49
4.4 Αρχιτεκτονικές Συνελκτικών Νευρωνικών Δικτύων	49
4.4.1 LeNet (1990)	49
4.4.2 AlexNet (2012)	50
4.4.3 ZFNet (2013)	51
4.4.4 VGGNet (2014)	51
4.4.5 GoogLeNet(InceptionV1) (2014)	53
4.4.6 ResNet (2015)	54
4.4.7 SENet	56
4.4.8 MobileNet	56
4.4.9 DenseNets	57
4.5 Παράμετροι αξιολόγησης νευρωνικού δικτύου	58
5 Πειραματική Διαδικασία	61
5.1 Εισαγωγή	61
5.2 Βάση Δεδομένων	61
5.3 Αξιολόγηση των Συστημάτων	65
5.4 Βιβλιοθήκες	66
5.4.1 Tensorflow	66

5.4.2 Keras.....	66
5.5 Πειραματικά Αποτελέσματα	67
5.5.1 Πειραματικά Αποτελέσματα Δικτύων, Χωρίς Προεπεξεργασία.....	67
5.5.2 Πειραματικά Αποτελέσματα Δικτύων, με χρήση ZCA.....	76
5.5.3 Πειραματικά Αποτελέσματα Δικτύων, με χρήση PCA	81
5.5.4 Συγκεντρωτικοί Πίνακες	85
Συμπεράσματα.....	89
Βιβλιογραφία.....	91

ΠΡΟΛΟΓΟΣ

Σκοπός της παρούσας εργασίας αποτελεί η μελέτη της χρήσης νευρωνικών δικτύων και ιδιαίτερα των τεχνικών βαθιάς μάθησης, στην αυτόματη κατηγοριοποίηση (ταξινόμηση) ιατρικών εικόνων.

Οι τεχνικές ταξινόμησης που δοκιμάστηκαν βασίζονται στην κατηγορία των βαθιών νευρωνικών δικτύων (DNN) και συγκεκριμένα στα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ). Τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Network CNN) είναι μια ξεχωριστή κατηγορία δικτύων μηχανικής μάθησης, τα οποία αποτελούν σήμερα το state-of-the-art στις τεχνικές αναγνώρισης προτύπων και ιδιαίτερα στις τεχνικές που αφορούν ταξινόμηση ιατρικών δεδομένων.

Τα ΣΝΔ παρουσιάζονται με πολλές διαφορετικές αρχιτεκτονικές, με διαφορετικό βάθος, διαφορετικό πλήθος συνελίξεων, διαφορετικές συναρτήσεις ενεργοποίησης και πολλές άλλες παραμέτρους και τεχνολογίες. Στην παρούσα εργασία εξετάσαμε πολλές από αυτές τις αρχιτεκτονικές οι οποίες έχουν παρουσιάσει εξαιρετική συμπεριφορά στην ταξινόμηση ιατρικών δεδομένων. Από τις αρχιτεκτονικές αυτές επιλέξαμε οκτώ με τις οποίες πειραματιστήκαμε. Όλα τα πειράματα έγιναν τρεις φορές: μια χωρίς προεπεξεργασία, μία με προεπεξεργασία ZCA, και μια με προεπεξεργασία PCA.

Κατά τη διαδικασία της εκπαίδευσης και ελέγχου των δικτύων, χρησιμοποιήθηκε η Ιστοπαθολογική Βάση Δεδομένων, καρκίνο του μαστού (BreCaHIS). Η βάση αυτή περιέχει συνολικά 7.909 εικόνες, όγκου μαστού, που συλλέχθηκαν από 82 ασθενείς: 58 με καρκίνο και 24 με κάποια καλοήγη νεοπλασία, των 700x460 εικονοστοιχείων, RGB 3 καναλιών, βάθος (8-bit σε κάθε κανάλι). Οι εικόνες στην βάση, είναι χωρισμένες σε 4 κατηγορίες, ανάλογα με την μικροσκοπική μεγέθυνσή τους: 40x, 100x, 200x και 400x. Να σημειωθεί ότι όλα τα ΣΝΔ που χρησιμοποιήσαμε ήταν αρχικά εκπαιδευμένα (προεκπαίδευση) στον τεράστιο όγκο των δεδομένων της ImageNet.

Η διάρθρωση της παρούσας διπλωματικής αποτελείται από πέντε κεφάλαια:

Στο πρώτο κεφάλαιο, γίνεται μια συνοπτική παρουσίαση του προβλήματος του καρκίνου του μαστού, των αιτιών, της διάγνωσης, και της θεραπείας του. Στόχος αυτής της παρουσίασης είναι η κατανόηση του προβλήματος, η ανάγκη καλύτερης αντίληψης της σπουδαιότητας των τεχνολογιών στον εντοπισμό των καρκινικών κυττάρων του μαστού, από μηχανές υψηλής ταχύτητας και αξιοπιστίας.

Στο δεύτερο κεφάλαιο, γίνεται ανασκόπηση των τεχνικών οι οποίες έχουν προταθεί ή χρησιμοποιηθεί για την αυτοματοποιημένη διάγνωση, κατηγοριοποίηση και επεξεργασία ιατρικών εικόνων. Ακόμα, γίνεται σύντομη παρουσίαση των ψηφιακών εικόνων, για λόγους συμφωνίας με την ορολογία που θα ακολουθηθεί, και παρουσίαση της βάσης δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση και τον έλεγχο των συστημάτων ταξινόμησης.

Στο Τρίτο κεφάλαιο, γίνεται μια σύντομη παρουσίαση της επιστήμης της Τεχνητής Νοημοσύνης, στην οποία εντάσσονται τα συστήματα ταξινόμησης με βαθιά Νευρωνικά Δίκτυα (DNN) που θα χρησιμοποιηθούν στη συνέχεια της εργασίας.

Στο τέταρτο κεφάλαιο, παρουσιάζεται η δομή των Συνελικτικών Νευρωνικών Δικτύων (ΣΝΔ). Γίνεται ανασκόπηση των διαφόρων αρχιτεκτονικών ΣΝΔ που έχουν προταθεί τα τελευταία αρκετά χρόνια, επιλογές των οποίων θα δοκιμαστούν στη συνέχεια της εργασίας.

Το πέμπτο κεφάλαιο, αναφέρεται στην πειραματική διαδικασία. Παρουσιάζονται τα πειραματικά αποτελέσματα από την υλοποίηση και τον έλεγχο των οκτώ διαφόρων αρχιτεκτονικών Συνελκτικών Νευρωνικών Δικτύων που επελέγησαν.

Τέλος, αναφέρονται τα Συμπεράσματα που προέκυψαν από αυτήν τη σπουδή και η Βιβλιογραφία που χρησιμοποιήσαμε για τη μελέτη και υλοποίηση της παρούσας διπλωματικής.

Ιούνιος 2020

Ορέστης Ν. Φακωτάκης

1 Καρκίνος του Μαστού

1.1 Εισαγωγή

Ο Καρκίνος, αποτελεί ένα από τα σοβαρότερα προβλήματα υγείας παγκοσμίως, με μεγαλύτερη συχνότητα εμφάνισης στις αναπτυγμένες χώρες. Τα στατιστικά του παγκόσμιου οργανισμού υγείας το 2008 δηλώνουν 7,6 εκατομμύρια θανάτους από καρκίνο, που αντιστοιχεί στο 13% από όλες τις αιτίες θανάτου. Ακόμη πιο ανησυχητικό είναι το γεγονός ότι την ίδια περίοδο διαγνώστηκαν 12,7 εκατομμύρια άνθρωποι να πάσχουν με καρκίνο. Σχετικές μελέτες του οργανισμού, δείχνουν ότι ο καρκίνος αναμένεται να ξεπεράσει τα 13 εκατομμύρια θανάτους το χρόνο, μέχρι το 2030 [1-4].

Ο καρκίνος αναφέρεται στο σύνολο των σοβαρών και πιθανά θανατηφόρων ασθενειών στις οποίες κύτταρα του οργανισμού πολλαπλασιάζονται με υπερβολικό ρυθμό και εισβάλουν σε κοντινούς ιστούς. Τα κύτταρα μεγαλώνουν και διαιρούνται για να σχηματίσουν νέα, σύμφωνα με τις ανάγκες του οργανισμού. Όταν στα κύτταρα έχει δημιουργηθεί βλάβη, λόγω γήρανσης ή κάποιας ζημιάς, αυτά πεθαίνουν και αναπληρώνονται από νέα κύτταρα. Όταν όμως κάποια κύτταρα, ενώ θα έπρεπε να πεθάνουν και νέα κύτταρα να σχηματιστούν, επιζούν, αυτά ονομάζονται **καρκινικά κύτταρα** και έχουν το χαρακτηριστικό να διαιρούνται χωρίς έλεγχο και να δημιουργούν όγκους (μάζες από με υπερβολικό ιστό).

Τα καρκινικά κύτταρα δεν αναπτύσσονται με κάποιο πρότυπο ή συγκεκριμένη λειτουργία και αγνοούν τις προσταγές του οργανισμού να σταματήσουν να διαιρούνται ή να ξεκινήσουν τη διαδικασία του κυτταρικού θανάτου. Ακόμα, ορισμένα από αυτά τα καρκινικά κύτταρα αποκόπτονται από τη μάζα τους και ταξιδεύουν στο υπόλοιπο σώμα, μέσω του αίματος, για να σχηματίσουν νέους όγκους.

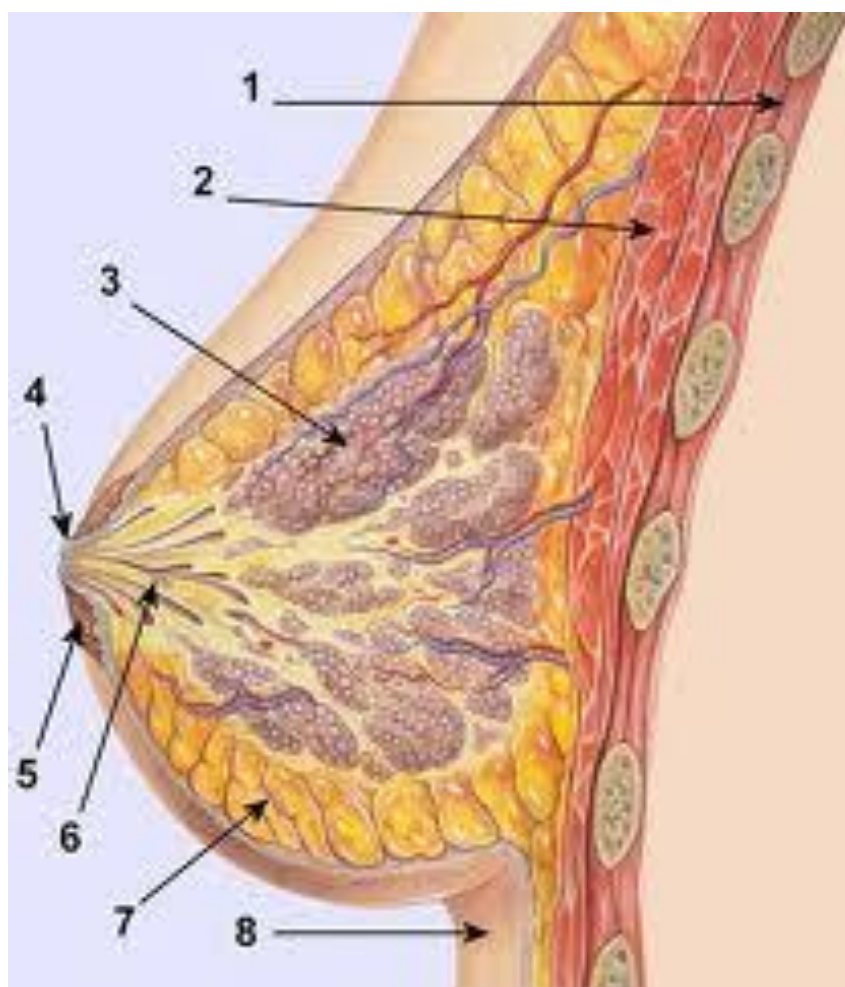
Ο καρκίνος του μαστού τώρα, αποτελεί μία από τις συχνότερα εμφανιζόμενες μορφές καρκίνου παγκοσμίως και είναι η πρώτη σε αριθμό κρουσμάτων στο γυναικείο πληθυσμό. Υπολογίζεται ότι στην Ελλάδα μόνο κάθε χρόνο προσβάλλονται 4.500 γυναίκες και χάνουν τη ζωή τους περισσότερες από το 1/3. Υπολογίζεται δε ότι μία στις 8 γυναίκες παγκοσμίως θα παρουσιάσει καρκίνο μαστού σε κάποια φάση της ζωής της.

Στη συνέχεια του κεφαλαίου γίνεται μια συνοπτική παρουσίαση του καρκίνου του μαστού, τα αίτια, τη διάγνωση, την έκταση και τα είδη αυτού. Στόχος αυτής της παρουσίασης είναι η καλύτερη κατανόηση του προβλήματος, και η ανάδειξη της σπουδαιότητας των σύγχρονων τεχνολογιών στον εντοπισμό των καρκινικών κυττάρων του μαστού, από μηχανές υψηλής αξιοπιστίας.

1.2 Ανατομία του Μαστού

Ο μαστός είναι μία μάζα αδενικού, λιπώδους και ινώδους ιστού, που βρίσκεται πάνω από τους θωρακικούς μύες του θωρακικού τοιχώματος (Εικόνα 1.1). Είναι ένας αδένας που σκοπός του είναι να παράγει γάλα για να θρέψει τους απογόνους μας στην τρυφερή τους ηλικία. Λέγεται και μαστικός ή μαζικός αδένας. Κάθε μαστός είναι στην ουσία μία μάζα λίπους μέσα στην οποία βρίσκονται χιλιάδες **λόβια** (3) που το κάθε ένα περιλαμβάνει μία περιοχή που παράγει το γάλα (αδένας) και μια περιοχή που μεταφέρει το γάλα (**πόρος**). Όλοι οι πόροι από το κάθε λόβιο ενώνονται σε μεγαλύτερους που ονομάζονται γαλακτοφόροι πόροι και εκβάλλουν στη θηλή του μαστού (4). Σε κάθε θηλή, εκβάλλουν περίπου είκοσι πόροι. Λίγο πριν την εκβολή τους διευρύνονται δημιουργώντας τους **γαλακτοφόρους κόλπους** (6). Η σκουρόχρωμη περιοχή γύρω από τη θηλή ονομάζεται θηλαία άλως (5) και οφείλει το χρώμα της στο ότι εκεί το δέρμα έχει κύτταρα με πυκνότερη εναπόθεση της χρωστικής μελανίνης.

Το αν θα είναι ένας μαστός στητός ή πεσμένος εξαρτάται από τον τόνο των ινών του Κούπερ, από το μέγεθος, την ηλικία και τη σωματική διάπλαση. Η πυκνότητά του στη μαστογραφία εξαρτάται από το ποσοστό ινωδών στοιχείων, πόρων και μαζικών λοβίων σε σχέση με το λίπος. Όσο πιο λίγο είναι το λίπος, τόσο πιο πυκνός είναι ένας μαστός στη μαστογραφία και τόσο μεγαλύτερη η πιθανότητα ανάπτυξης καρκίνου. Το μέγεθος του μαστού δε σχετίζεται με την πιθανότητα ανάπτυξης καρκίνου αλλά με τα γονίδια και τον τρόπο ζωής.



Εικόνα 1.1: Ανατομία του γυναικείου μαστού: 1 Μεσοπλεύριοι μύες, 2 Θωρακικοί μύες, 3 Λοβοί του μαστικού αδένος, 4 Θηλή, 5 Θηλαία άλως, 6 Γαλακτοφόροι κόλποι, 7 Περιμαστικό λίπος.

1.3 Καρκίνος του Μαστού

Ο Καρκίνος του Μαστού αναφέρεται στην ανάπτυξη κακοήθους όγκου στην περιοχή του μαστού. Προκαλείται από ανεξέλεγκτο πολλαπλασιασμό παθολογικών κυττάρων με αποτέλεσμα το σχηματισμό κακοήθους όγκου στην περιοχή και ουσιαστικά την κυτταρική νόσο. Τα καρκινικά αυτά κύτταρα έχουν τη δυνατότητα εξάπλωσης σε γειτονικούς ιστούς με προφανείς συνέπειες για ολόκληρο τον οργανισμό.

Ο καρκίνος του μαστού αποτελεί σήμερα μία από τις συχνότερα εμφανιζόμενες μορφές καρκίνου παγκοσμίως και είναι η πρώτη σε αριθμό κρουσμάτων στο γυναικείο πληθυσμό. Παρατηρείται εξαιρετικά σπάνια στους άνδρες, με περίπου 100 φορές μικρότερη συχνότητα συγκριτικά με τις γυναίκες και τα ίδια ποσοστά επιβίωσης. Η πιθανότητα προσβολής από καρκίνο του μαστού αυξάνεται όσο μεγαλώνει μια γυναίκα με κορύφωση τα 75 χρόνια και μετά μειώνεται. Στις νέες γυναίκες, είναι πιο σπάνιος αλλά και πιο επιθετικός από ότι στις ηλικιωμένες.

1.3.1 Αίτια

Έχουν εντοπιστεί αρκετοί παράγοντες που συμβάλουν στη δημιουργία καρκίνου του μαστού, μεταξύ των οποίων είναι:

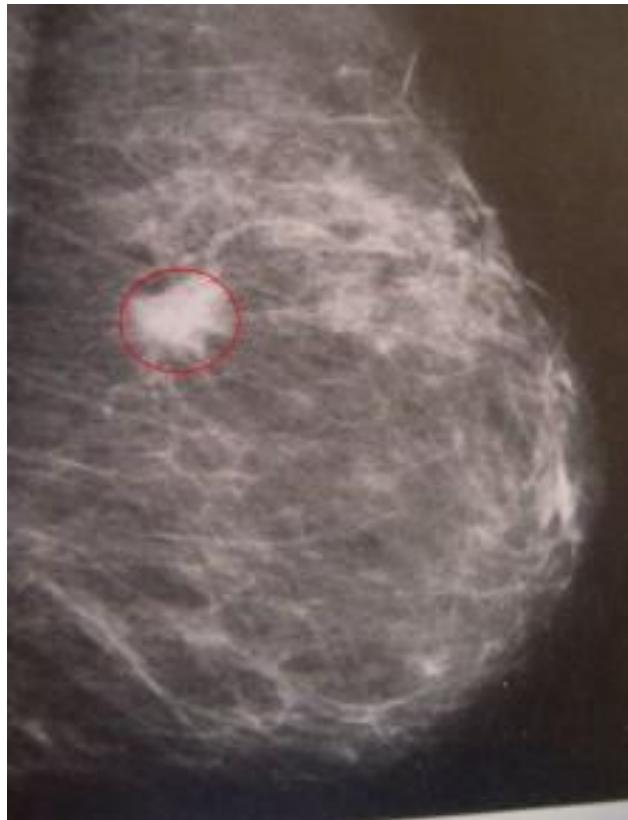
- **Ηλικία:** ο καρκίνος του μαστού μπορεί να προκύψει σε οποιαδήποτε ηλικία μετά την εφηβεία. Όμως, τα ποσοστά αυξάνονται όσο αυξάνεται η ηλικία, με περισσότερες περιπτώσεις να παρουσιάζονται μετά τα 50.
- **Κληρονομικότητα:** Ωστόσο, μόλις το 5-10% των κρουσμάτων σχετίζεται με παράγοντες κληρονομικότητας.
- **Διαταραχές της εμμήνου ρύσης και της κύησης:** γυναίκες με πρόωμη έναρξη της εμμήνου ρύσης (πριν τα 12) ή με καθυστερημένη εμμηνόπαυση (μετά τα 55) αντιμετωπίζουν αυξημένο κίνδυνο εμφάνισης καρκίνου στο μαστό. Ακόμα, η λήψη οιστρογόνων μετά την εμμηνόπαυση έχει συσχετιστεί με αυξημένα ποσοστά εμφάνισης της νόσου. Ανάλογος συσχετισμός έχει προκύψει και για **τις άτεκνες γυναίκες**.
- **Αλκοόλ:** Η χρήση οινοπνευματωδών αυξάνουν τη συγκέντρωση οιστρογόνων στο αίμα.
- **Παχυσαρκία:** αυξάνει τον κίνδυνο καρκίνου του μαστού καθώς αυξάνει τα επίπεδα των οιστρογόνων. Η παραγωγή των οιστρογόνων μετά την εμμηνόπαυση γίνεται κυρίως μέσα σε λιπώδη ιστό.
- **Κάπνισμα:** Η κατανάλωση 20 τσιγάρων ημερησίως για εννέα χρόνια, αυξάνει τον κίνδυνο εμφάνισης καρκίνου του μαστού κατά 60%.
- **Ιστορικό Καρκίνου:** Γυναίκες που έχουν ξαναεμφανίσει καρκίνο έχουν αυξημένες πιθανότητες να εμφανίσουν έναν δεύτερο καρκινικό επεισόδιο στο μαστό.
- **Καθιστική Ζωή:** Η τακτική άσκηση πριν την έναρξη της έμμηνου ρύσης μπορεί να μειώσει τον κίνδυνο καρκίνου του μαστού.
- **Έκθεση σε ακτινοβολία:** Η ακτινοβολία (ιονίζουσα) έχει προσδιοριστεί ως παράγοντας αύξησης του κινδύνου εμφάνισης της νόσου.

1.3.2 Συμπτώματα

Τα συνήθη συμπτώματα που χαρακτηρίζουν τον καρκίνο του μαστού μπορούν να συνοψιστούν στα ακόλουθα:

- Διόγκωση λεμφαδένων,
- Εξόγκωμα ή σκλήρυνση στην περιοχή του μαστού,
- Έκκριση υγρού ή αίματος από τη θηλή,
- Ερυθρότητα, ευαισθησία ή πόνοι στο στήθος,
- Έλξη του δέρματος (εισολκή).

Στο αρχικό στάδιο ο καρκίνος του μαστού δεν παρουσιάζει χαρακτηριστικά συμπτώματα. Αργότερα όμως μπορεί να εμφανιστεί ψηλαφητό μόρφωμα, αλλαγή του χρώματος του δέρματος. Αν δεν δοθεί σημασία στα συμπτώματα τότε μπορεί να εμφανίσει σημάδια προχωρημένης νόσου, (θερμό και ερυθρό μαστό, πόνους στα οστά, διόγκωση), (Εικόνα 1.2).



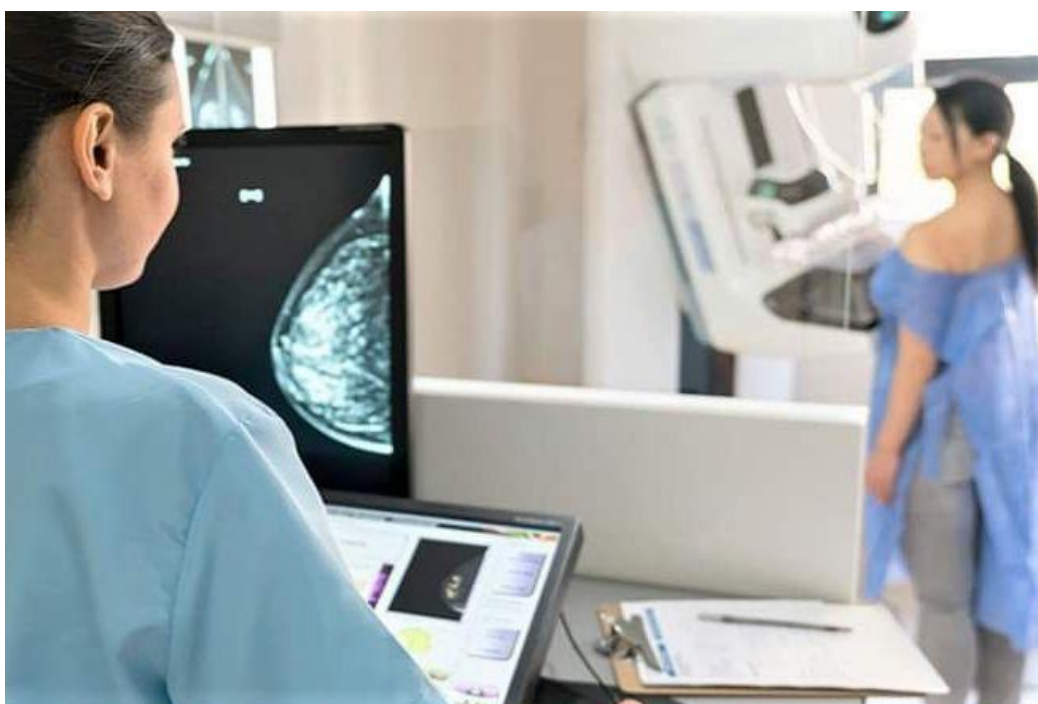
Εικόνα 1.2: Μαστογραφία καρκίνου (Πορογενής διηθητικός καρκίνος). Παρατηρείται το ανώμαλο περίγραμμα.

1.3.3 Διάγνωση

Ο καρκίνος του μαστού, κατά κανόνα, δεν προλαμβάνεται πρωτογενώς, διότι δεν γνωρίζουμε ακόμα τον παράγοντα δημιουργίας του, αλλά δευτερογενώς μετά από έγκαιρη διάγνωση. Να σημειωθεί ότι στην Ευρώπη, το 60% των κρουσμάτων καρκίνου του μαστού διαγιγνώσκεται σε πρώιμο στάδιο.

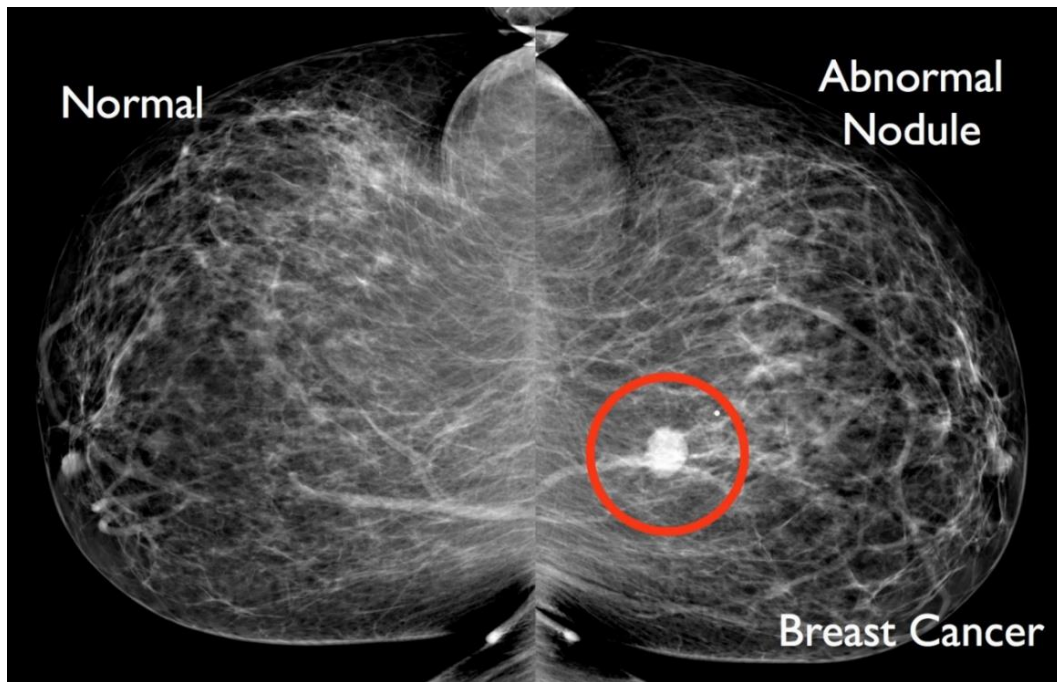
Η πρόληψη περιλαμβάνει την αυτοεξέταση του μαστού από τον πάσχοντα και την κλινική εξέταση του μαστού από το γιατρό. Η διάγνωση του καρκίνου του μαστού αρχίζει όταν μια μάζα γίνεται αντιληπτή στο μαστό με την ψηλάφηση. Στην συνέχεια βέβαια με τη χρήση απεικονιστικών διαγνωστικών μέσων όπως:

- **Μαστογραφία:** Είναι η ακτινογραφία του μαστού από μαστογράφο (Εικόνα 1.3). Η ψηφιακή μαστογραφία αποτελεί τη «φυσική» εξέλιξη της κλασικής αναλογικής μαστογραφίας. Αποτελεί την εξέταση που κατέχει το μεγαλύτερο ποσοστό αξιοπιστίας για τη διερεύνηση ύπαρξης καρκίνου καθώς και για τον εντοπισμό του σε πρώιμο στάδιο. Στη μαστογραφία ο μαστός τοποθετείται σε ειδική θέση του μηχανήματος και λαμβάνεται η ψηφιακή εικόνα (μαστογραφία).



Εικόνα 1.3: Μαστογράφος-Μαστογραφία. Αποτελεί την εξέταση που κατέχει το μεγαλύτερο ποσοστό αξιοπιστίας.

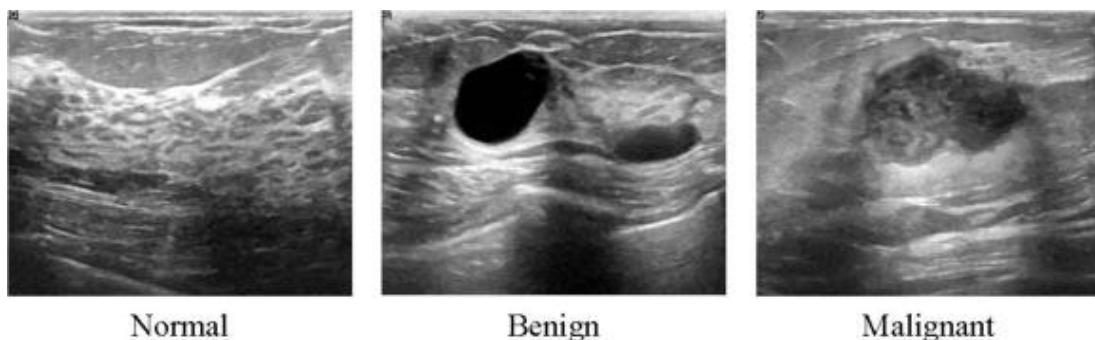
Στην απεικόνιση οι φωτεινές και οι σκοτεινές περιοχές που σχηματίζονται οφείλονται στην απορροφητικότητα του μαστού. Στην εικόνα η οποία προκύπτει με σκούρο χρώμα απεικονίζονται οι περιοχές του λιπώδους ιστού του μαστού και γενικά περιοχές που παρουσιάζουν χαμηλή πυκνότητα και είναι εύκολα διαπερατές από την ακτινοβολία (Εικόνα 1.4). Με ανοιχτό χρώμα πολύ κοντά στο λευκό απεικονίζονται εκείνες οι περιοχές που χαρακτηρίζονται είτε από υψηλή πυκνότητα ιστού είτε είναι οι παθολογικές περιοχές, οπότε αναφερόμαστε σε όγκους.



Εικόνα 1.4: Μαστογραφία. Οι φωτεινές και οι σκοτεινές περιοχές που σχηματίζονται οφείλονται στην απορροφητικότητα του μαστού.

Οι μαστογραφίες λαμβάνονται είτε για προληπτικούς σκοπούς είτε για διαγνωστικούς σκοπούς. Σκοπός της προληπτικής μαστογραφίας είναι η εύρεση καρκινογένεσεων οι οποίες βρίσκονται σε πρώιμο στάδιο και δεν έχουν παρουσιάσει συμπτώματα καρκινογένεσης.

- **Υπερηχογράφημα μαστών** (ultrasound): Ο υπέρηχος επιτρέπει σημαντική ελευθερία στη λήψη των απεικονίσεων του μαστού σχεδόν από οποιοδήποτε προσανατολισμό (Εικόνα 1.5). Η αξία του είναι μεγάλη στο διαχωρισμό μεταξύ συμπαγών και κυστικών μαζών. Ο υπέρηχος μπορεί συχνά να καθορίσει εάν μια ύποπτη περιοχή είναι στην πραγματικότητα μια κύστη ή μια πυκνή μάζα που μπορεί με βιοψία να καθοριστεί η φύση της. Για την καλύτερη απεικόνιση, τοποθετείται ειδικό πυκνόρρευστο υγρό εξωτερικά πάνω στο δέρμα του μαστού που θα γίνει η εξέταση. Στο Σχήμα 1.5 [7], φαίνονται τρία υπερηχογραφήματα, που απεικονίζουν τρεις ακραίες περιπτώσεις.



Εικόνα 1.5: Δείγματα συνόλου δεδομένων εικόνων μαστού. Υπερηχογραφήματα (a) Normal, (b) Benign, (c) Malignant. [7]).

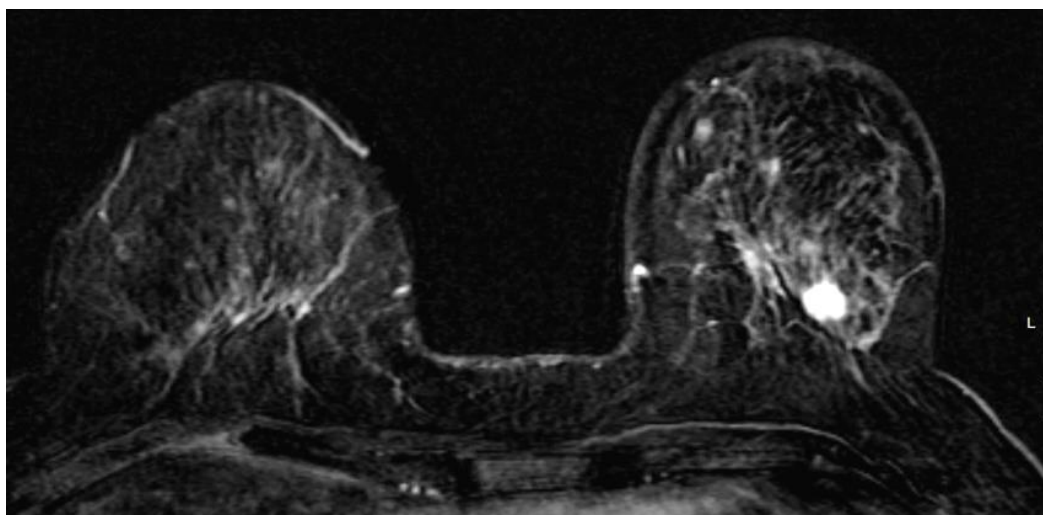
- **Αξονική τομογραφία** (Εικόνα 1.6): Είναι μια ακτινολογική μέθοδος στην οποία χρησιμοποιούνται ακτίνες X για την απεικόνιση εγκάρσιων τομών του ανθρώπινου σώματος. Για την παραγωγή μιας τέτοιας εξέτασης, ο ασθενής τοποθετείται κάτω από έναν

κυλινδρικού σχήματος σαρωτή. Οι ακτίνες X με τις οποίες ακτινοβολείται ο ασθενής ψηφιοποιούνται από ανιχνευτές που βρίσκονται τοποθετημένοι στην απέναντι πλευρά του ασθενή. Με τη χρήση πολλαπλών εικόνων μπορεί να υπολογιστεί ένα τρισδιάστατο μοντέλο και να γίνει η περαιτέρω επεξεργασία με τη χρήση υπολογιστή. Είναι εξέταση κυρίως σε επείγουσες και οξείες καταστάσεις σε καρκινοπαθείς.



Εικόνα 1.6: Αξονική τομογραφία. χρησιμοποιούνται ακτίνες X για την απεικόνιση εγκάρσιων τομών του ανθρώπινου σώματος.

- **Μαγνητική τομογραφία (MRI):** Με αυτή τη μέθοδο γίνεται η διάγνωση καρκίνου στο μαστό, χρησιμοποιώντας ισχυρά μαγνητικά πεδία για να δημιουργήσουν την απεικόνιση του μαστού (Εικόνα 1.7). Το σύστημα μαγνητικής μαστογραφίας (Magnetic resonance, MRI/MR) μετατρέπει τα μαγνητικά πεδία για να πετύχει προβολές σε οποιοδήποτε επίπεδο και προσανατολισμό.



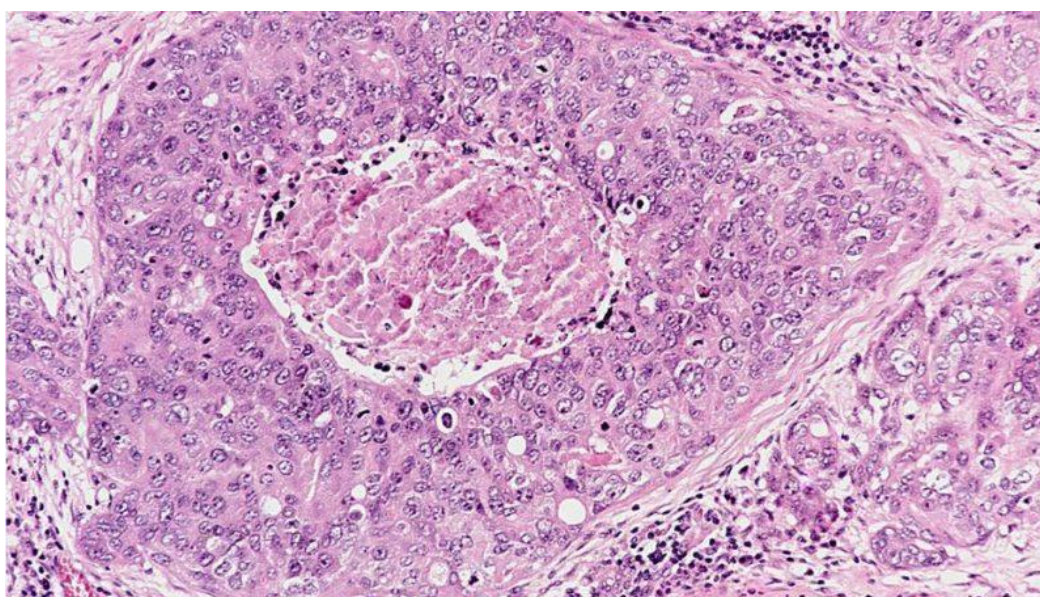
Εικόνα 1.7: Μαγνητική Τομογραφία. *Ταχεία Μαγνητική Μαστογραφία, καρκίνος 1.4 εκ. αριστερού μαστού.*

- **Σπινθηρογράφημα:** είναι μια συμπληρωματική εξέταση μαστών που μπορεί να χρησιμοποιηθεί σε μερικούς ασθενείς για να ερευνήσει μια ανωμαλία μαστών. Αναφέρεται ως εξέταση **πυρηνικής ιατρικής** και δεν είναι ένα αρχικό εξεταστικό εργαλείο για τον καρκίνο του μαστού αλλά μπορεί να είναι χρήσιμη σε επιλεγμένες περιπτώσεις μετά από διαγνωστική μαστογραφία. Η πυρηνική ιατρική περιλαμβάνει την έγχυση ενός ραδιενεργού ανιχνευτή (χρωστική ουσία) στον ασθενή. Δεδομένου ότι η χρωστική ουσία συσσωρεύεται διαφορετικά στους καρκινώδεις και μη-καρκινώδεις ιστούς, το σπινθηρογράφημα μπορεί να βοηθήσει τους παθολόγους να καθορίσουν εάν υπάρχει καρκίνος.

Όταν εντοπιστεί ένας ύποπτος όγκος, με τη βοήθεια των πιο πάνω διαγνωστικών διαδικασιών, ακολουθεί η διάγνωση με λήψη υλικού από τον όγκο για μικροσκοπική εξέταση ώστε να διευκρινιστεί η διαφορά ανάμεσα σε ένα συμπαγή όγκο και μια κύστη γεμάτη υγρό. Υπάρχουν 4 τέτοιες διαγνωστικές μέθοδοι:

- **Κυτταρολογική εξέταση:** Με λεπτή βελόνα (σύριγγα) γίνεται αναρρόφηση κυττάρων από την ύποπτη περιοχή του μαστού.
- **Ιστολογική βιοψία:** Με μεγαλύτερη βελόνα (απαιτείται τοπική αναισθησία) και τομή, γίνεται αφαίρεση ιστολογικού υλικού, από το οποίο προκύπτουν ιστοπαθολογικές εικόνες μικροσκοπίας. Με τη μελέτη και παρατήρηση των ιστών μέσω του μικροσκοπίου, (Εικόνα 1.8) καθίσταται δυνατή η τελική διάγνωση της νόσου. Ακολουθεί η γνωμάτευση του ιατρού, που κατηγοριοποιεί τα ευρήματα σε ένα από τα στάδια εξέλιξης του καρκίνου. Στα πλαίσια τις παρούσας διπλωματικής γίνεται επεξεργασία **Ιστοπαθολογικών Εικόνων καρκίνου του μαστού** και με τη χρήση Συνελκτικών Νευρωνικών Δικτύων, επιδιώκεται η αυτόματη κατηγοριοποίηση (διαχωρισμός) σε θετικές και αρνητικές εικόνες, στην ύπαρξη όγκου.

Η Ιστοπαθολογία είναι ένας από τους τομείς που ειδικεύεται στη μικροσκοπική μελέτη ιστών. Είναι ένας από τους πιο σημαντικούς τομείς της παθολογίας και χρησιμοποιείται για την διάγνωση καρκίνου, όγκων και άλλων ασθενειών.



Εικόνα 1.8: Ιστοπαθολογική εικόνα μικροσκοπίας. Τύπος καρκίνου του μαστού.

- **Στερεοτακτική βιοψία:** Αποτελεί τη λήψη ιστολογικού υλικού, υπό στερεοτακτική (απεικονιστική) καθοδήγηση.
- **Ανοιχτή χειρουργική βιοψία:** η ταυτοποίηση της βλάβης γίνεται με ταχεία βιοψία κατά τη διάρκεια του χειρουργείου πριν την αφαίρεση του όγκου.

Ωστόσο, δεν είναι όλες οι αλλοιώσεις (διογκώσεις ή σκληρύνσεις) καρκινικές και κακοήθεις. Υπολογίζεται πως το 90% των όγκων των μαστών (σε ηλικίας 20 – 50 ετών) προκαλούνται από κάποια πάθηση καλοήθους φύσης.

- **Καλοήθεις** παθήσεις μπορούν να είναι: *Μασταλγία, Μαστίτιδα, Ινοαδένωμα, Ινοκυστική μαστοπάθεια, Γαλακτοφορεκτασία, Κύστη, Λίπωμα, Φυλλοειδής όγκος, Θήλωμα, Φλεγμονή-Απόστημα, Γαλακτοκήλη, Αδενολίπωμα,*
- **Κακοήθεις** παθήσεις του μαστού (καρκινώματα), μπορούν να είναι: *Επί τόπου Πορογενές, Πορογενές διηθητικό, Σωληνώδες, Μυελώδες, Βλεννοπαράγωγο Μαστικό, Θηλώδες, Διάτρητο, Διηθητικό Λοβιακό, Φλεγμονώδες.*

Μέχρι πρότινος, οι αξιολογήσεις εικόνων γινόταν αποκλειστικά από ειδικευμένους. Βέβαια, η αξιοπιστία της αξιολόγησης από άνθρωπο περιορίζεται από τα μη συστηματικά πρότυπα αναζήτησης του ανθρώπου, την παρουσία δομικού θορύβου στην εικόνα και την ύπαρξη πολύπλοκων ασθενειών που απαιτούν το συνδυασμό τεράστιας ποσότητας απεικονιστικών και κλινικών δεδομένων.

Τα τελευταία χρόνια η διάγνωση έχει εξελιχθεί εξαιτίας της υποβοηθητικής διάγνωσης μέσω υπολογιστή (Computer Aided Diagnosis - CAD). Μετά τη λήψη των ψηφιοποιημένων ιατρικών εικόνων ακολουθεί η ανάλυσή τους στον υπολογιστή. Στο στάδιο αυτό αξιοποιούνται τεχνικές που εντάσσονται στον ευρύτερο χώρο της Ψηφιακής Επεξεργασίας Σήματος και της Τεχνητής Νοημοσύνης (επεξεργασία εικόνα, ανίχνευση πιθανών ύποπτων περιοχών με μεθόδους κατάτμησης και μοντέλα αυτοματοποιημένης ταξινόμησης όπως τα νευρωνικά δίκτυα **που αποτελεί και το αντικείμενο της παρούσας διπλωματικής**). Τα εργαλεία αυτά συντελούν στην καλύτερη αξιολόγηση του περιστατικού από το γιατρό, την πιο έγκυρη διάγνωση της ασθένειας και την αισθητά πιο εύκολη ανίχνευση περισσότερων περιπτώσεων καρκίνου λόγω του αυξημένου επιπέδου εξειδίκευσης και ευαισθησίας.

1.3.4 Θεραπεία

Η υποβοηθητική διάγνωση με CAD συστήματα ή συστήματα Βαθιάς Μάθησης που θα δούμε στη συνέχεια, και η επιπρόσθετη πληροφορία που αυτά προσφέρουν έχουν σαν αποτέλεσμα την μεγαλύτερη σταθερότητα στην περίθαλψη.

Τις περισσότερες φορές, το επόμενο στάδιο, μετά την αποτελεσματική διάγνωση, της θεραπείας της νόσου είναι η χειρουργική επέμβαση για την αφαίρεση του όγκου. Ακολουθεί το τελευταίο στάδιο που είναι η μετεγχειρητική θεραπεία που συνήθως περιλαμβάνει ακτινοθεραπεία, χημειοθεραπεία ή ορμονοθεραπεία. Τα στοιχεία που λαμβάνονται υπόψη στην επιλογή της μετεγχειρητικής θεραπείας είναι το μέγεθος και η διαφοροποίηση του όγκου, η ύπαρξη ορμονικών υποδοχέων και η γενική κατάσταση της υγείας.

2 Τεχνικές Ταξινόμησης

2.1 Εισαγωγή

Οι ταξινομητές αποτελούν την καρδιά της αυτοματοποιημένης κατηγοριοποίησης ιατρικών εικόνων. Σήμερα, οι μηχανές αυτές συμβάλουν στο να ανιχνεύουν διάφορες ασθένειες μέσω της μελέτης ιατρικών εικόνων, εμφανίζοντας παρόμοια επίπεδα ακρίβειας (ή ξεπερνώντας τα) με επαγγελματίες και εξειδικευμένους γιατρούς [44]. Στην περίπτωση της ανίχνευσης και διάγνωσης καρκίνου του μαστού οι ταξινομητές, σε συνδυασμό με αλγόριθμους λήψης αποφάσεων, βρίσκουν πολύ μεγάλη εφαρμογή στην παροχή υποβοηθητικής διάγνωσης [9]. Ως γνωστό, οι γιατροί παρέχουν διαγνώσεις συνδυάζοντας κατάρτιση, εμπειρία και ατομική κρίση, πράγμα που μπορεί να δημιουργήσει έναν αναπόφευκτο βαθμό υποκειμενικότητας στην ερμηνεία της εικόνας, αφού βασίζεται στον ανθρώπινο παράγοντα.

Με τον όρο ιατρικές εικόνες ή απεικονιστικές εξετάσεις αναφερόμαστε σε αναπαραστάσεις οργάνων, ιστών και κυττάρων, διαφορετικών σημείων του ανθρώπινου σώματος. Η ταξινόμηση των ιατρικών εικόνων αναφέρεται στον εντοπισμό (αναγνώριση) περιστατικών ή φυσικών αντικειμένων που υπάρχουν στις εικόνες αυτές. Οι ταξινομητές μαστογραφιών, σε συνδυασμό με αλγόριθμους κατάτμησης και στατιστικά μοντέλα, μπορούν με υψηλή αξιοπιστία να κατηγοριοποιήσουν τις εικόνες σε προκαθορισμένες κλάσεις, σύμφωνα με τα ευρήματα.

Οι σύγχρονες διαδικασίες ταξινόμησης βασίζονται σε τεχνικές ανάλυσης και βελτιστοποίησης βαθιών νευρωνικών δικτύων (DNN), που σήμερα αποτελούν τη σύγχρονη προσέγγιση της μηχανικής μάθησης. Τη βαθιά μάθηση (Deep Learning). Θεωρούνται ως οι πλέον αξιόπιστες τεχνικές αναγνώρισης προτύπων, και βρίσκουν μεγάλη εφαρμογή στη σύγχρονη ιατρική για την ταξινόμηση ιατρικών δεδομένων.

Ωστόσο, η καλή αξιοπιστία των μηχανών βαθιάς μάθησης, σχετίζεται με την καλή εκπαίδευση των συστημάτων. Αυτό απαιτεί σωστές βάσεις δεδομένων, δηλαδή κατάλληλα δεδομένα και κυρίως δεδομένα σε μεγάλο όγκο. Στα πλαίσια αυτής της διπλωματικής χρησιμοποιήθηκε μια βάση δεδομένων που περιελάμβανε Ιστοπαθολογικές εικόνες μικροσκοπίας, καρκίνο του μαστού (Breast Cancer Histopathological Database).

Στη συνέχεια του κεφαλαίου, γίνεται ανασκόπηση των τεχνικών οι οποίες έχουν προταθεί για την αυτοματοποιημένη διάγνωση, κατηγοριοποίηση και επεξεργασία ιατρικών εικόνων. Ακόμα, γίνεται παρουσίαση της βάσης δεδομένων που χρησιμοποιήθηκε στο πειραματικό μέρος της διπλωματικής, για εκπαίδευση και έλεγχο των συστημάτων ταξινόμησης. Τέλος,

γίνεται μια σύντομη παρουσίαση των ψηφιακών εικόνων, για λόγους συμφωνίας με την ορολογία που θα ακολουθηθεί.

2.2 Βάσεις Δεδομένων Ψηφιακών Εικόνων

Οι σωστές και επαρκείς βάσεις δεδομένων ιατρικών εικόνων αποτελούν προϋπόθεση για τη σχεδίαση συστημάτων ταξινόμησης εικόνων (αναγνώρισης προτύπων). Στην αξιόπιστη λειτουργία των συστημάτων ταξινόμησης βασικό ρόλο παίζουν τα δεδομένα (πλήθος και είδος) που θα χρησιμοποιηθούν για να εκπαιδευθεί το σύστημα (**training data**) και τα δεδομένα με τα οποία αυτό θα ελεγχθεί (**testing data**).

Πριν προχωρήσουμε στην παρουσίαση των τεχνικών ταξινόμησης θεωρούμε σκόπιμο να αναφερθούμε σε αυτά τα δεδομένα και ακόμα να κάνουμε μια σύντομη παρουσίαση των διαφόρων κατηγοριών ψηφιακών εικόνων, για λόγους συνέπειας, σε ότι αφορά την ορολογία που θα ακολουθηθεί.

2.2.1 Ψηφιακή Εικόνα

Μια ψηφιακή εικόνα $i[m, n]$ αναπτύσσεται σε ένα διακριτό χώρο δύο διαστάσεων και παράγεται από την ψηφιοποίηση μιας αναλογικής εικόνας $i[x, y]$ που αναπτύσσεται σε ένα συνεχή χώρο με διαστάσεις x και y . Έτσι, μια ψηφιακή εικόνα θεωρείται ως η κατανομή της πληροφορίας στο επίπεδο $\{x, y\}$ [10-11].

Η συνεχής εικόνα διαιρείται σε N σειρές και M στήλες. Τα σημεία τομής των σειρών με τις στήλες είναι τα εικονοστοιχεία (pixels), και αποτελεί το δομικό στοιχείο κάθε εικόνας. Έτσι, μία ψηφιακή εικόνα διαστάσεων $M \times N$ γενικά παριστάνεται από ένα δυσδιάστατο πίνακα, $\text{Array}(i, j)$, όπου σε κάθε κελί (εικονοστοιχείο) αποθηκεύονται διάφορες τιμές. Η φωτεινότητα της κάθε ψηφιακής εικόνας είναι ανάλογη της τιμής $I(i, j)$ του κάθε εικονοστοιχείου (i, j) . Συνεπώς, ο πίνακας $I(i, j)$ είναι μια διακεκριμένη συνάρτηση που εκφράζει την ένταση της φωτεινότητας σε κάθε εικονοστοιχείο.

Οι ψηφιακές εικόνες κυριάρχησαν με την ραγδαία ανάπτυξη της τεχνολογίας. Μέσα από την πληθώρα των τεχνολογιών που προέκυψε, προτάθηκαν και διάφοροι τύποι ψηφιακών εικόνων. Κυριάρχησαν όμως δύο μεγάλες κατηγορίες: τα αρχεία τύπου **ψηφιοκουκίδων (raster)** και τα **διανυσματικά (vector)** αρχεία εικόνων. Η διαφορά μεταξύ των δύο κατηγοριών είναι στο ότι τα αρχεία τύπου ψηφιοκουκίδων σχηματίζονται από το άθροισμα κουκίδων, ενώ στα διανυσματικά η εικόνα σχηματίζεται από το άθροισμα γεωμετρικών σχημάτων. Από τις δύο αυτές κατηγορίες, μεγάλο προβάδισμα έχουν επιτύχει οι ψηφιακές εικόνες τύπου **raster**.

Για την παραγωγή μίας ψηφιακής φωτογραφίας απαιτούνται μέσα δειγματοληψίας, όπως ψηφιακές φωτογραφικές μηχανές, κάμερες και σαρωτές. Ανάλογα με το μέσο παραγωγής της μπορεί να είναι δυαδική εικόνα (Binary Image), μονοχρωματική αποχρώσεων του γκρι (Gray-level images) ή έγχρωμη εικόνα (color image):

- **Δυαδική Εικόνα (Binary Image):** Η δυαδική εικόνα, αποτελεί την απλούστερη μορφή μιας ψηφιακής εικόνας. Οι στάθμες φωτεινότητας της είναι δύο (π.χ. άσπρο και μαύρο) (Εικόνα 2.1). Το χρώμα άσπρο παρίσταται με την τιμή ένα (1) και το χρώμα μαύρο με την τιμή μηδέν (0). Ο κυριότερος λόγος διατήρησης αυτής της μορφής ψηφιακών εικόνων είναι γιατί καταλαμβάνουν πολύ μικρή μνήμη και η επεξεργασία τους έχει μικρότερο υπολογιστικό κόστος. Πολλές φορές χρησιμοποιούνται σαν δείκτες μονοχρωματικών και έγχρωμων εικόνων για πληροφορίες όπως θέση, εμβადών και μορφή αντικειμένων.



Εικόνα 2.1: Δυαδική (Binary) εικόνα, αναπαράσταση με μαύρο και άσπρο χρώμα.

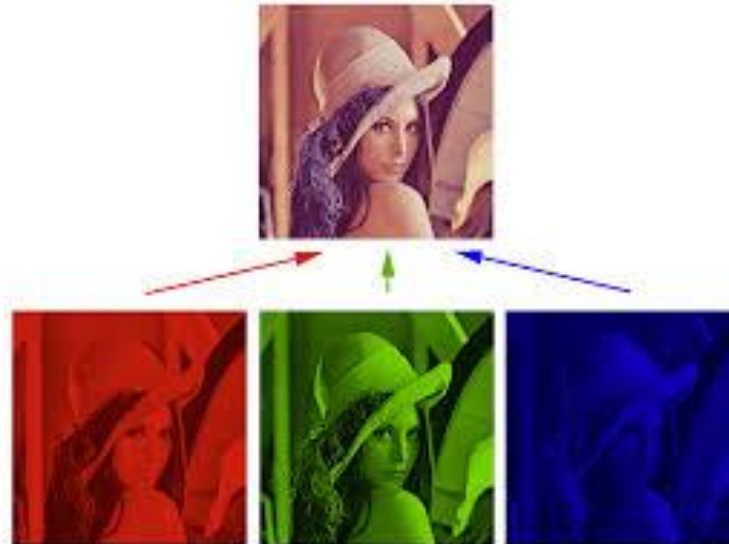
- **Μονοχρωματική Εικόνα με αποχρώσεις του Γκρι (Gray-level images):** Η μονοχρωματική ψηφιακή εικόνα θα μπορούσε να χαρακτηριστεί και ως η εξέλιξη της δυαδικής εικόνας με τη μόνη διαφορά τους την ένταση χρώματος του κάθε εικονοστοιχείου. Στη μονοχρωματική ψηφιακή εικόνα ή εικόνα αποχρώσεων του γκρι, το κάθε εικονοστοιχείο μπορεί να λάβει τιμές 2^n (για $n=8$, 256 τιμές), από 0 (μαύρο) έως 255 (λευκό), που στη δυαδική αρίθμηση σημαίνει 8bit το κάθε εικονοστοιχείο.

Λόγω του μεγάλου αριθμού αποχρώσεων της έντασης κάθε εικονοστοιχείου στις μονοχρωματικές εικόνες μπορούμε να δούμε με μεγάλη ευκρίνεια στις μορφές και τα σχήματα σε μια φυσική τους αναπαράσταση. Μία ψηφιακή εικόνα αποχρώσεων του γκρι διαστάσεων $N \times M$ (256x256) παριστάνεται στην Εικόνα 2.2.



Εικόνα 2.2: Εικόνα αποχρώσεων του γκρι (Gray-level), διαστάσεων $N \times M$ (256x256), 8bit κάθε εικονοστοιχείο.

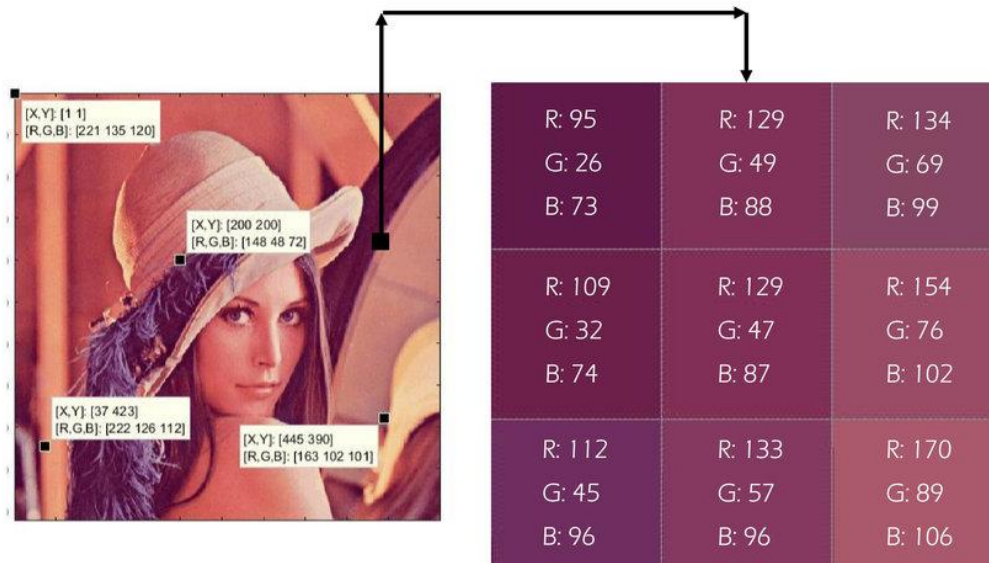
- **Έγχρωμη εικόνα (color image):** Στην έγχρωμη ψηφιακή εικόνα με το χρωματικό σύστημα RGB (είναι το επικρατέστερο), το κάθε χρώμα συντίθεται από τα χρώματα: κόκκινο (**Red**), πράσινο (**Green**) και μπλε (**Blue**), (Εικόνα 2.3). Το χρώμα του κάθε εικονοστοιχείου μίας έγχρωμης εικόνας αναπαρίσται μέσα από τις τρεις αυτές συνιστώσες: το επίπεδο του R, το επίπεδο του G και το επίπεδο του B. Μια έγχρωμη ψηφιακή εικόνα διαστάσεων $N \times M$ αντιπροσωπεύεται από ένα τρισδιάστατο πίνακα $N \times M \times 3$ (3 επίπεδα): M (αριθμός γραμμών) \times N (αριθμός στηλών) \times Z (τα 3 επίπεδα R-G-B).



Εικόνα 2.3: RGB εικόνα και η διάσπασή της στα τρία επιμέρους χρώματα.

Το κάθε επίπεδο αντιπροσωπεύεται από ένα δυσδιάστατο πίνακα όπου μέσα είναι καταχωρημένες οι εντάσεις του κάθε εικονοστοιχείου για κάθε επίπεδο. Η ένταση κάθε εικονοστοιχείου «**ERGB**» παράγεται από το γινόμενο των τριών εντάσεων: E_R (επίπεδο R), επί E_G (επίπεδο G), επί E_B (επίπεδο B), και μας διατυπώνει συγκεκριμένο χρώμα από το ορατό φάσμα του φωτός σε κάθε εικονοστοιχείο. Δηλαδή: **Ένταση $ERGB = E_R \times E_G \times E_B$.**

Η ένταση κάθε χρώματος κάθε εικονοστοιχείου συνήθως αποθηκεύεται χρησιμοποιώντας 8 bits, (επίπεδα κβαντοποίησης 256). Συνολικά, ένα εικονοστοιχείο στην έγχρωμη εικόνα απαιτεί 24 bits ($3 \times 8 \text{bits}$) για την αποθήκευση (Εικόνα 2.4). Έτσι για παράδειγμα, σε μία έγχρωμη ψηφιακή εικόνα χρωματικού βάθους 8bits είναι δυνατό να διατυπωθούν 16,77 εκατομμύρια ($2^8 \times 2^8 \times 2^8$) πιθανά χρώματα.



Εικόνα 2.4: Έγχρωμη απόδοση γνωστής εικόνα (Lenna's). Δεξιά και πάνω στην εικόνα δίνονται πληροφορίες των αντίστοιχων εικονοστοιχείων, (τιμές των R, B, G).

Άλλα χαρακτηριστικά των ψηφιακών εικόνων μπορούν να συνοψιστούν στα ακόλουθα:

- Ανάλυση Εικόνας (image resolution):** Ένας τρόπος προσδιορισμού της ανάλυσης μιας εικόνας είναι μέσω της αναφοράς στο σύνολο των εικονοστοιχείων. Μπορεί απλά να γίνει με αναφορά στον αριθμό των εικονοστοιχείων ανά γραμμή και ανά στήλη. Παράλληλα υπάρχει και η σχετική ανάλυση, όπου υπολογίζεται ο αριθμός των εικονοστοιχείων σε μία μονάδα μήκους. Ευρέως διαδεδομένο είναι το **ppi** (pixel per inch = εικονοστοιχείο ανά ίντσα) ή και **dpi** (dots per inch = κουκίδες ανά ίντσα) (για εκτύπωση). Το φυσικό μέγεθος dpi ή ppi στην ουσία καθορίζει την πυκνότητα πληροφορίας της εικόνας αλλά δεν μπορεί να καθορίσει το ποσό πληροφορίας της εικόνας.
- Χρωματικό βάθος (deep color):** Στην ψηφιακή εικόνα αποχρώσεων γκρι, ορίζουμε χρωματικό βάθος τις αποχρώσεις της κλίμακας του γκρι. Δηλαδή είναι η ποσότητα πληροφορίας που περιέχουν οι ψηφιακές εικόνες στην απόδοση χρώματος. Το χρωματικό βάθος σε μία εικόνα εκφράζεται από συντελεστή m , τα bits που μπορούμε να αποθηκεύσουμε μία πληροφορία, υψωμένα στην δύναμη του δύο (2^m). Για παράδειγμα, για μονοχρωματική 8bits (2^8) 256 χρώματα, για έγχρωμη 24bits ($2^8 \times 2^8 \times 2^8$) 16 εκατομμύρια χρώματα.
- Ευκρίνεια:** Ευκρίνεια μίας εικόνας είναι ένας δείκτης που καθορίζει πόσο ξεκάθαρα μπορούμε να βλέπουμε τις λεπτομέρειες σε μία εικόνα. Η υψηλή ανάλυση αυξάνει γενικά την ευκρίνεια των εικόνων σε διάφορες διαστάσεις. Εξαρτάτε τόσο από της διαστάσεις της εικόνας όσο και από το χρωματικό βάθος της εικόνας. Η ευκρίνεια καθορίζεται από τα όρια μεταξύ των ζωνών από τους διαφορετικούς τόνους ή χρώματα.
- Μέγεθος Εικόνας (Image size):** Το μέγεθος του αρχείου μίας ψηφιακής εικόνας μετριέται σε KB, MB και GB όπως και όλα τα αρχεία σε ένα ηλεκτρονικό υπολογιστή. Το μέγεθος αρχείου σε μία ψηφιακή εικόνα είναι ανάλογο των διαστάσεών της σε εικονοστοιχεία αλλά και το χρωματικό της βάθος. Συγκεκριμένα, το μέγεθος μίας ψηφιακής εικόνας υπολογίζεται: M (αριθμός γραμμών) \times N (αριθμός στηλών) \times m (χρωματικό βάθος bits). Εικόνες μεγάλων διαστάσεων και μεγάλου χρωματικού βάθους έχουν πολύ περισσότερη ευκρίνεια και λεπτομέρεια από εικόνες με μικρότερες διαστάσεις ή μικρότερο χρωματικό βάθος.

- **Συμπίεση Εικόνας (Image Compression):** Ο μεγάλος όγκος των δεδομένων στις ψηφιακές εικόνες δημιουργεί προβλήματα σε ανάγκες χώρου αποθήκευσης. Για το λόγο αυτόν, έχουν προταθεί πολλοί αλγόριθμοι συμπίεσης της εικόνας. Οι τρόποι συμπίεσης γενικότερα χωρίζονται σε συμπίεση με απώλεια και συμπίεση χωρίς απώλεια. Από τους επικρατέστερους αλγόριθμους συμπίεσης χωρίς απώλειες είναι οι τύπου «**jpeg**» ή «**jrg**» και οι τύπου «**tiff**» ή «**tif**».
- **Τύπος JPG:** είναι ο πιο γνωστός τύπος αρχείου ψηφιακής εικόνας. Επιτυγχάνει εξαιρετικά υψηλή συμπίεση με μηδενικές απώλειες στην πληροφορία της εικόνας. Ένας από τους λόγους της μεγάλης διάδοσης του είναι γιατί ο χρήστης έχει την ικανότητα να επιλέξει το βαθμό συμπίεσης της εικόνας. Σε αυτό το τύπο μπορούμε να αποθηκεύσουμε εικόνες με χρωματικό μοντέλο RGB ή χρωματικό μοντέλο CMYK.
- **Τύπος TIFF:** είναι ο δεύτερος πιο γνωστός τύπος αρχείου ψηφιακής εικόνας. Υποστηρίζεται από τα περισσότερα λειτουργικά συστήματα. Τα αρχεία τύπου TIFF χρησιμοποιούνται από προσωπικούς υπολογιστές έως και ψηφιακές φωτογραφικές μηχανές. Δέχονται οποιαδήποτε ανάλυση και επιδέχονται συμπίεση με διάφορες μεθόδους χωρίς να έχουμε απώλεια στην πληροφορία της εικόνας.

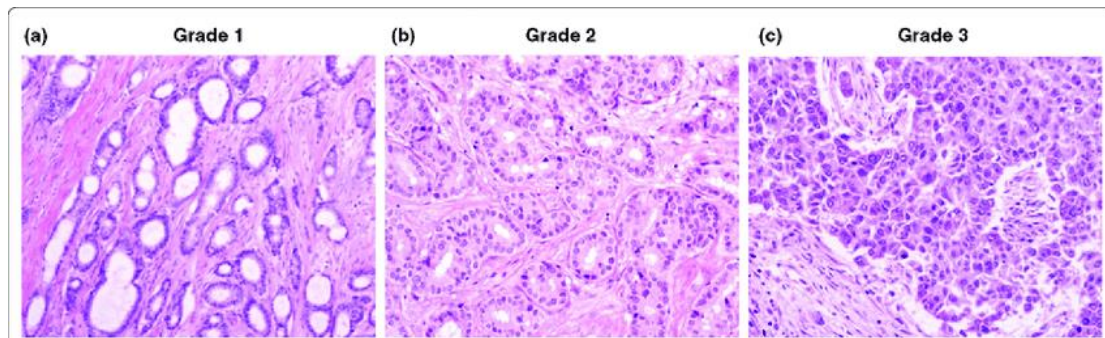
2.2.2 Βάση Δεδομένων Ιστοπαθολογικών Εικόνων

Οι μεγάλες βάσεις δεδομένων, το πολλά (μεγάλα) δεδομένα σε ψηφιακές ιατρικές εικόνες, αποτελούν τη βάση στη σχεδίαση σύγχρονων συστημάτων αναγνώρισης προτύπων. Οι μηχανές αυτές αποτελούν τη σύγχρονη προσέγγιση της μηχανικής μάθησης, τη βαθιά μάθηση και έχουν ως βάση τα βαθιά νευρωνικά δίκτυα (DNN), τα οποία βρίσκουν τεράστιες εφαρμογή στην ιατρική, σήμερα.

Το αντικείμενο της παρούσας διπλωματικής αφορά στην κατηγοριοποίηση Ιστοπαθολογικών Εικόνων, που σημαίνει ότι η βάση δεδομένων που χρησιμοποιήθηκε στην παρούσα διπλωματική περιλαμβάνει Ιστοπαθολογικές εικόνες μικροσκοπίας, καρκίνο του μαστού (Breast Cancer Histopathological Database).

Οι Ιστοπαθολογικές εικόνες μικροσκοπίας συνήθως παράγονται από ιστούς που αφαιρούνται από ανθρώπους, που ονομάζουμε χειρουργικά παρασκευάσματα ή βιοψίες. Με τη μελέτη και παρατήρηση των ιστών μέσω του μικροσκοπίου, καθίσταται δυνατή η τελική διάγνωση της νόσου. Με τη γνωμάτευση του ιατρού, κατηγοριοποιούνται σε ένα από τα τρία στάδια εξέλιξης του καρκίνου (Εικόνα 2.6):

- **Βαθμός 1:** καλά διαφοροποιημένος όγκος. (τα κύτταρα φαίνονται πιο όμοια με τα φυσιολογικά και δεν αυξάνεται ο όγκος με ταχείς ρυθμούς.
- **Βαθμός 2:** μέτρια διαφοροποιημένος όγκος. Τα κύτταρα μοιάζουν κάπως διαφορετικά από τα φυσιολογικά.
- **Βαθμός 3:** ελάχιστα διαφοροποιημένος όγκος. Κύτταρα ακανόνιστα παθολογικά που μπορούν να μεγαλώνουν και να εξαπλώνονται πιο επιθετικά από ότι στους άλλους βαθμούς.



Εικόνα 2.6: Ιστολογικός βαθμός καρκίνου του μαστού: (α) καλά διασκορπισμένος, (β) μέτρια διαφοροποιημένος και (γ) κακώς διαφοροποιημένος όγκος [45].

Στην παρούσα διπλωματική εργασία, όπως προαναφέραμε, ερευνούμε την δυνατότητα της αυτοματοποιημένης διάγνωσης με τη χρήση μηχανών αναγνώρισης ιστοπαθολογικών εικόνων μικροσκοπίας που αφορούν βιοψίες καρκίνου του μαστού. Εδώ χρησιμοποιήθηκε η Ιστοπαθολογική Βάση Δεδομένων, καρκίνου του μαστού BreaKHis. Η βάση περιέχει συνολικά 7.909 εικόνες, ιστού όγκου μαστού, που συλλέχθηκαν από 82 ασθενείς: 58 που πάσχουν από καρκίνο και 24 που έχουν κάποια καλοήγη νεοπλασία. Μέχρι σήμερα, περιέχει 2.480 καλοήγη και 5.429 κακοήγη δείγματα, των 700x460 εικονοστοιχείων, RGB, 3 καναλιών, βάθος 8-bit σε κάθε κανάλι, μορφή PNG.

2.3 Τεχνικές Εντοπισμού Περιοχών Ενδιαφέροντος

Στις μέρες μας, έρευνες γίνονται γύρω από τον αυτόματο εντοπισμό περιοχών ενδιαφέροντος σε μαστογραφίες χωρίς την ανάγκη επισημάνσης των περιοχών αυτών, από χρήστη. Οι έρευνες αυτές περιλαμβάνουν τεχνικές κατάτμησης εικόνων (image segmentation), σε συνδυασμό με διάφορα στάδια προ-επεξεργασίας των εικόνων. Τα στάδια της προ-επεξεργασίας, μέσω διαφόρων μετασχηματισμών, είναι αναγκαία για τον περιορισμό θορύβου ή των όποιων άλλων προβλημάτων υπάρχουν στην μορφολογία της εικόνας. Το στάδιο της προ-επεξεργασίας και κατάτμησης των εικόνων, θεωρείται επιβεβλημένο για την προσαρμογή των δεδομένων στον εκάστοτε ταξινομητή, που ακολουθεί μετά αυτής της διαδικασίας.

2.3.1 Εξισορρόπηση Ιστογράμματος (Histogram Equalization)

Η τεχνική εξισορρόπησης ιστογράμματος είναι μια ευρέως διαδεδομένη μέθοδος βελτιστοποίησης εικόνων που έχουν ληφθεί σε συνθήκες κακού φωτισμού (πολύ χαμηλού ή πολύ υψηλού). Βασίζεται στην τροποποίηση του ιστογράμματος, αυξάνοντας τοπικά την αντίθεση της φωτεινότητας. Μετασχηματίζει τις γκρι φωτεινότητες μιας εικόνας έτσι ώστε αυτές να κατανέμονται ομοιόμορφα σ' όλη την κλίμακα φωτεινότητας. Η εικόνα που προκύπτει με τον τρόπο αυτό είναι αυξημένης αντίθεσης σε σχέση με την αρχική.

Για εικόνες προερχόμενες από ακτίνες X ή εικόνες με πολύ χαμηλό ή υψηλό φωτισμό γενικότερα, οι τεχνικές εξισορρόπησης ιστογράμματος αποτελούν τις πιο διαδεδομένες μεθόδους. Η λειτουργία της συγκεκριμένης τεχνικής βασίζεται στη μεγιστοποίηση της αντίθεσης της φωτεινότητας, σε συγκεκριμένες περιοχές της εικόνας. Επομένως, η μέθοδος εφαρμόζεται καλύτερα όταν στο ιστόγραμμα μιας εικόνας υπάρχει συγκεντρωτισμός εικονοστοιχείων σε συγκεκριμένη φωτεινότητα. Στην ουσία, μέσω της εφαρμογής της μεθόδου εξισορρόπησης οι περιοχές της εικόνας με χαμηλή ένταση γίνονται σκοτεινότερες και οι περιοχές με υψηλότερη ένταση γίνονται φωτεινότερες.

2.3.2 Κατωφλίωση (Binarization)

Η τεχνική της κατωφλίωσης είναι μια μέθοδος κατάτμησης της εικόνας η οποία χωρίζει το ιστόγραμμα σε περιοχές, με βάση την φωτεινότητα τους. Επίσης, ορίζει τιμές κατωφλίου για τα οριακά σημεία και για την κάθε περιοχή. Συνήθως, το ιστόγραμμα χωρίζεται σε δυο περιοχές (ζώνες) με βάσει ενός κατωφλίου T : τη $B1$ και τη $B2$. Η πρώτη ζώνη ($B1$) καθορίζεται βάσει του φόντου της εικόνας (background), ενώ η δεύτερη ($B2$) βάσει του αντικειμένου. Στις μαστογραφικές εικόνες οι φωτεινότερες περιοχές αντιστοιχούν στους όγκους και τον ιστό του στήθους, ενώ οι σκοτεινότερες περιοχές στο φόντο.

Η επιλογή του κατωφλίου, γίνεται από το συγκεντρωτικό ιστόγραμμα φωτεινότητας της εικόνας. Για παράδειγμα, επίπεδο φωτεινότητας L δείχνει τον αριθμό των εικονοστοιχείων της εικόνας που έχουν φωτεινότητα μικρότερη ή ίση της τιμής L . Κατώφλι T επιλέγεται το επίπεδο φωτεινότητας για το οποίο 90% των pixels έχουν φωτεινότητα μικρότερη ή ίση της τιμής L . Δηλαδή, ορίζουμε μια σταθερά φωτεινότητας, σύμφωνα με την οποία αν σε κάποια περιοχή τα εικονοστοιχεία έχουν τιμή φωτεινότητας μικρότερη της σταθεράς, εκεί ορίζεται η τιμή του κατωφλίου.

2.3.3 Συσταδοποίηση (Clustering)

Η Συσταδοποίηση (ομαδοποίηση) είναι μια εναλλακτική μέθοδος κατάτμησης της εικόνας. Η λειτουργία της βασίζεται στο να δημιουργεί ομάδες σε περιοχές από εικονοστοιχεία με κοινά χαρακτηριστικά. Η τεχνική της συσταδοποίησης ψάχνει για περιοχές εικονοστοιχείων, όπου οι τιμές της φωτεινότητας να είναι μεγαλύτερες από κάποιο, αρχικά ορισμένο, κατώφλι.

Ο αλγόριθμος της συσταδοποίησης αρχίζει από μια ομάδα εικονοστοιχείων και προσθέτει γειτονικά εικονοστοιχεία, με κοινές ιδιότητες, βάση προκαθορισμένων κριτηρίων. Κατά καιρούς έχουν προταθεί πολλοί αλγόριθμοι συσταδοποίησης, μεταξύ των οποίων, Ανταγωνιστικό νευρωνικό δίκτυο (Competitive Neural Network), ασαφής C-means, K-means, Gaussian mixture model).

2.3.4 Μετασχηματισμός Laplace

Με το μετασχηματισμό Laplace επιτυγχάνεται ο εντοπισμός οποιασδήποτε πιθανής σκίασης γύρω από τα όρια κάποιας προκαθορισμένης περιοχής. Τα όρια της κάθε περιοχής εστίασης ονομάζετε περίγραμμα και ισούται με μια οζώδη σκίαση. Πρακτικά ο μετασχηματισμός μεταβαίνει από την οζώδη σκίαση στο περίγραμμα. Αυτό επιφέρει αλλαγή στην κυρτότητα της επιφάνειας και επομένως χρειάζεται την αλλαγή του προσήμου της Laplacian επιφανείας ώστε να καθοριστούν οι γραμμές της οζώδους σκίασης.

Ο κυριότερος λόγος χρήσης του μετασχηματισμού Laplace είναι για την ανίχνευση όγκων σε μαστογραφίες. Στην τεχνική αυτή φαίνεται ότι η συμπίεση έχει μεγάλη επίδραση στις καλοήθεις μάζες και έτσι κάποιο μέτρο συμπίεστότητας που βασίστηκε στη μαζική μορφή επιτρέπει τη διαφοροποίηση των καλοηθών και κακοηθών μαζών.

2.3.5 Σύγκριση των Σκιάσεων

Η σύγκριση των σκιάσεων μας επιτρέπει την διάκριση των δύο μαστών. [8]. Πρακτικά, η τεχνική αυτή βασίζεται την κατάτμηση της εικόνας, ανάλογα με την ένταση που υπάρχει σε

κάθε περιοχή εικονοστοιχείων. Αυτό βοηθά στο να συγκεντρωθούν στατιστικά στοιχεία που να επιτρέψουν τη σύγκριση μεταξύ των δύο μαστών (αριστερού και δεξιού) και να εντοπιστούν πιθανές ανωμαλίες.

Η μέθοδος αυτή συνδυάζει την κατάτμηση και την κατωφλίωση της φωτεινότητας της εικόνας. Ως πρώτο βήμα εφαρμόζει κατωφλίωση της φωτεινότητας της εικόνας σε διάφορες μαστογραφίες. Έπειτα, εφαρμόζει κατάτμηση σε συγκεκριμένες περιοχές, ανάλογα με τις τιμές της κατωφλίωσης της φωτεινότητας. Κατόπιν, εξετάζει τις παραλειπόμενες εικόνες, που προέκυψαν κατά τη διαδικασία της κατάτμησης, ώστε μέσω της σύγκρισης των εικόνων να εντοπιστούν οι σκιάσεις.

2.4 Μέθοδοι Ταξινόμησης

Οι ταξινομητές αποτελούν τη βάση στην διαδικασία της αυτοματοποιημένης κατηγοριοποίησης ιατρικών εικόνων. Στην περίπτωση της ανίχνευσης και διάγνωσης καρκίνου του μαστού οι ταξινομητές, σε συνδυασμό με αλγόριθμους λήψης αποφάσεων, βρίσκουν πολύ μεγάλη εφαρμογή παρέχοντας υποβοηθητική διάγνωση σε γιατρούς [9]. Οι γιατροί παράγουν διαγνώσεις συνδυάζοντας κατάρτιση, εμπειρία και ατομική κρίση, πράγμα που μπορεί να δημιουργήσει έναν αναπόφευκτο βαθμό υποκειμενικότητας στην ερμηνεία της εικόνας, αφού βασίζεται κατά πολύ στην ανθρώπινη οπτική αντίληψη, με τις όποιες αδυναμίες της. Οι ταξινομητές, συνδυάζουν αλγορίθμους κατάτμησης (όπως αυτοί που προαναφέρθηκαν) αλλά και στατιστικά μοντέλα ώστε να μπορούν επιτυχώς να κατηγοριοποιούν τις εικόνες σε προκαθορισμένες κλάσεις. Σαν είσοδο τα συστήματα αυτά μπορεί να έχουν είτε μονόχρωμες είτε έγχρωμες (3 χρώματα) ιατρικές εικόνες. Και με τις δύο περιπτώσεις τα συστήματα δεν διαφέρουν ως προς τη διαδικασία ταξινόμησης αλλά σε τροποποιήσεις στον τρόπο προεπεξεργασίας της εκάστοτε εικόνας.

2.4.1 Ταξινόμηση Υφής (Texture Classification)

Η υφή μια εικόνας αποτελεί το σύνολο των παραμέτρων που μπορούν να εξαχθούν από την εικόνα. Αυτό περιλαμβάνει πληροφορίες για τις τιμές που παίρνει το χρώμα σε όλες τις περιοχές των εικονοστοιχείων αλλά και την ένταση όπου θα έχουν τα εικονοστοιχεία σε αυτές τις περιοχές.

Στη διαδικασία της κατηγοριοποίησης της υφής ο στόχος είναι η ταξινόμηση μιας εικόνας σε κάποια από τις γνωστές κατηγορίες υφής. Η αποτελεσματικότητα των διαδικασιών ταξινόμησης και κατάτμησης εξαρτάται πάρα πολύ από την περιγραφή της υφής της εικόνας. Για το λόγο αυτό έχουν προταθεί πολλές τεχνικές ανάλυσης υφής που θα μπορούσαν να ομαδοποιηθούν στις ακόλουθες τέσσερις γενικές κατηγορίες Μεθόδων: Στατιστικές, Γεωμετρικές, Βασισμένες στο Μοντέλο, και Επεξεργασίας Σήματος.

Με βάση την ποσοτικοποίηση και το χαρακτηρισμό της υφής της εικόνας, τα χαρακτηριστικά υφής ομαδοποιούνται σε τρεις γενικές κατηγορίες: **Συντακτικά**, **Στατιστικά** και **Φασματικά** χαρακτηριστικά. Η συντακτική ανάλυση υφής αναγνωρίζει θεμελιώδη στοιχεία των εικόνων. Τα συντακτικά στοιχεία πρώτης τάξης (μέσος όρος, τυπική απόκλιση, ενέργεια και διασπορά) χρησιμοποιούνται με την ίδια έννοια που έχουν και στη στατιστική. Τα στατιστικά στοιχεία δεύτερης τάξης εξάγονται από πιθανολογικές μήτρες (συσχέτιση, αντίθεση, εντροπία κλπ). Με τον ίδιο τρόπο υπολογίζεται και η τραχύτητα μιας υφής σε μια προκαθορισμένη περιοχή της εικόνας. Τα φασματικά χαρακτηριστικά (run-length, cooccurrence κλπ) υπάρχουν επίσης αλλά συχνά υστερούν σε ευαισθησία στις μεγάλους μεγέθους μεταβολές της υφής μιας εικόνας.

2.4.2 Ταξινόμηση Naïve Bayes

Στη βάση της θεωρίας του Bayes, σχεδιάστηκαν αυτοματοποιημένα συστήματα στατιστικής ταξινόμησης, για διάγνωση καρκίνου του μαστού υψηλής αξιοπιστίας [9]. Τα συστήματα Ταξινόμησης Naïve Bayes, έχουν ως προϋπόθεση τη σωστή προ-επεξεργασία των εικόνων, συμπεριλαμβανομένης και της αποθορυβοποίησης των δεδομένων εισόδου. Ακόμα, απαιτούν σωστή κατάτμηση των εικόνων, ώστε να εντοπιστούν οι πυρήνες του κυττάρου, σύμφωνα με τους οποίους θα γίνει η ταξινόμηση. Για την διαδικασία αυτή χρησιμοποιούνται αλγόριθμοι συσταδοποίησης, (π.χ. Competitive Neural Network, K-means, Gaussian mixture model).

Στη συνέχεια της διαδικασίας της προεπεξεργασίας, τα στοιχεία εισόδου τροφοδοτούν και εκπαιδεύουν τον ταξινομητή Naïve Bayes.

2.4.3 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

Οι Μηχανές Διανυσμάτων Υποστήριξης (SVM) ανήκουν στην κατηγορία των γραμμικών, μη-πιθανοτικών δυαδικών ταξινομητών. Ο χαρακτηρισμός δυαδικός προκύπτει επειδή συνήθως η ταξινόμηση γίνεται σε δυο κλάσεις. Ωστόσο, με κατάλληλες τροποποιήσεις (χρησιμοποιώντας μια SVM για κάθε κλάση) μπορούν να εξελίσσουν τη λειτουργία τους σε ταξινομητές πολλαπλών κατηγοριών, έχοντας δομή νευρωνικού δικτύου.

Η λειτουργία των SVM είναι μη-πιθανοτικοί, αφού η επιλογή της κλάσης, που ανήκουν τα δεδομένα προς κατηγοριοποίηση, γίνεται βάση του σφάλματος ταξινόμησης. Δηλαδή, η σωστή κλάση για μια εικόνα να αποδίδει μια τιμή μεγαλύτερη της τιμής που αποδίδει η λάθος κλάση, κατά ένα περιθώριο (ένα κατώφλι D).

Οι SVM χρησιμοποιούνται σε προβλήματα ιατρικής διάγνωσης που περιορίζονται σε δυαδικές απαντήσεις, όπως για παράδειγμα: υπάρχει ή δεν υπάρχει καρκίνος στην εικόνα, είναι ο καρκίνος τύπου Α ή Β κλπ.

2.4.4 Τεχνικές Εξόρυξης Δεδομένων (Data Mining)

Οι τεχνολογίες Εξόρυξης Δεδομένων εντάσσονται στα πλέον δυναμικά εξελισσόμενα επιστημονικά πεδία. Αφορά την ανάλυση μεγάλων συνόλων δεδομένων με σκοπό τον εντοπισμό μη προφανών σχέσεων και την σύναψη των δεδομένων, με τρόπους που είναι κατανοητοί στον κάτοχο των δεδομένων. Δηλαδή οι τεχνικές Data Mining είναι αλγόριθμοι εύρεσης συσχετίσεων και ιδιοτήτων μεταξύ πολυάριθμων δεδομένων. Οι τεχνολογίες αυτές, συνδυάζουν τεχνικές μηχανικής μάθησης, τεχνικές αναγνώρισης προτύπων, αλλά και μαθηματικά και στατιστικά μοντέλα. Οι σχέσεις και οι συνοψίσεις που παράγονται από μία διαδικασία εξόρυξης δεδομένων αναφέρονται ως μοντέλα ή πρότυπα.

Το πρώτο βήμα, στη διαδικασία εξόρυξης των δεδομένων, είναι η προεπεξεργασία των δεδομένων και πιθανά η μείωση των διαστάσεων της βάσης δεδομένων, εφαρμόζοντας κατάλληλα επιλεγμένες τεχνικές. Αυτό μπορεί να γίνει, είτε χρησιμοποιώντας τεχνικές εντοπισμού περιοχών ενδιαφέροντος, είτε σαρώνοντας την εικόνα με φίλτρα, επιδιώκοντας την εξομάλυνσή της. Στις τεχνικές εξόρυξης δεδομένων μπορεί να προστεθεί και ο κανόνας συσχέτισης (Association rule), ο οποίος βοηθά στη δημιουργία προτύπων, ώστε να μπορέσουν να βγουν οι κανόνες κατηγοριοποίησης.

Τα πρότυπα που εμφανίζονται στα δεδομένα, με τα αντίστοιχα χαρακτηριστικά τους, προκύπτουν από έναν αλγόριθμο με μορφή δέντρου αποφάσεων. Γνωστός τέτοιος αλγόριθμος δέντρου αποφάσεων είναι ο FP-Tree (Frequent Pattern Tree). Έτσι, η τελική διάγνωση βασίζεται στην σύγκριση αποτελεσμάτων μεταξύ των ιατρικών δεδομένων από το δέντρο αποφάσεων με τα πιο κοινά χαρακτηριστικά που εμφανίζονται.

Συνοψίζοντας, τα αποτελέσματα της ταξινόμησης με βάση τις τεχνολογίες Εξόρυξης Δεδομένων περιλαμβάνουν την προεπεξεργασία των εικόνων, χρησιμοποιώντας φίλτρα εξομάλυνσης, κατάτμηση κλπ. για τον εντοπισμό των περιοχών ενδιαφέροντος, και στη συνέχεια τους κανόνες συσχέτισης, για να προκύψει η ακριβής διάγνωση. Η οριστική διάγνωση στηρίζεται στη σύγκριση του αποτελέσματος του δέντρου αποφάσεων (FP-Tree) με εισόδους (εικόνες), με τα πιο συχνά εμφανιζόμενα στοιχεία στην εικόνα που αξιολογείται.

2.4.5 Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)

Τα Νευρωνικά Δίκτυα ή Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), αποτελούν σήμερα την κορωνίδα των τεχνικών αναγνώρισης προτύπων. Εντάσσονται στις πιο αποτελεσματικές μεθόδους και χρησιμοποιούνται ευρύτατα από ερευνητές στο χώρο της ταξινόμησης των μαστογραφιών.

Τα ΤΝΔ παρουσιάζουν ξεχωριστά πλεονεκτήματα σε σχέση με τις κλασικές μεθόδους ταξινόμησης προτύπων. Είναι αυτοπροσαρμοστικές διατάξεις οδηγούμενες από τα δεδομένα, δηλαδή μπορούν να προσαρμοστούν τα ίδια στα δεδομένα με τα οποία εκπαιδεύονται, χωρίς την απαίτηση λειτουργικής διευκρίνισης για τη μορφή των υποκείμενων μοντέλων. Ακόμα, τα ΤΝΔ είναι μη-γραμμικά μοντέλα, πράγμα που τα καθιστά ευέλικτα στη μοντελοποίηση σύνθετων προβλημάτων του πραγματικού κόσμου. Ακόμα και αν υπάρχουν προβλήματα στα δεδομένα το σύστημα καταφέρνει να τα ταξινομήσει με σχετικά καλή ακρίβεια, δεδομένου ότι κάθε διαδικασία ταξινόμησης αναζητά στην ουσία μια σχέση μεταξύ των χαρακτηριστικών ενός αντικειμένου και των ιδιοτήτων του, ειδικά στις περιπτώσεις της αυτοματοποιημένης ιατρικής διάγνωσης (των μαστογραφιών στην παρούσα εργασία). Τέλος, τα ΤΝΔ πραγματοποιούν τη λειτουργία της ταξινόμησης χρησιμοποιώντας τεχνικές εκπαίδευσης, είτε με επίβλεψη είτε χωρίς επίβλεψη. Η δυνατότητα των νευρωνικών δικτύων να λειτουργούν και χωρίς επίβλεψη μας επιτρέπει να εκπαιδεύσουμε ένα σύστημα ενώ έχουμε έλλειψη κάποιων πληροφοριών.

Τα Τεχνητά Νευρωνικά Δίκτυα, αποτελούν την βάση την παρούσας διπλωματικής εργασίας. Εντάσσονται στο ευρύτερο επιστημονικό πεδίο της Τεχνητής Νοημοσύνης, που παρουσιάζεται στο κεφάλαιο που ακολουθεί.

3 Τεχνητή Νοημοσύνη

3.1 Εισαγωγή

Κατά την δεκαετία του 1950 [13] έκανε την εμφάνισή του ένας νέος τομέας της επιστήμης των υπολογιστών που καθιερώθηκε διεθνώς με την ονομασία Τεχνητή Νοημοσύνη - Artificial Intelligence (AI). Στόχος της έρευνας στην Τεχνητή Νοημοσύνη (TN) είναι η ανάπτυξη υπολογιστικών συστημάτων τα οποία θα είναι ικανά να επιλύουν σύνθετα πραγματικά προβλήματα επιδεικνύοντας μία νοήμονα συμπεριφορά ή αλλιώς, η επιστήμη που προσπαθεί να καταστήσει τους υπολογιστές ικανούς να κάνουν πράγματα στα οποία οι άνθρωποι, προς το παρόν, τα καταφέρνουν καλύτερα [14].

	Ανθρώπινο μοντέλο	Ιδανικό πρότυπο
Σκέψη	Συστήματα που σκέφτονται όπως ο άνθρωπος	Συστήματα που σκέφτονται ορθολογικά
	<i>Η αυτοματοποίηση λειτουργιών που σχετίζονται με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση...</i>	<i>Η μελέτη νοητικών ικανοτήτων με την χρήση υπολογιστικών μοντέλων</i>
Συμπεριφορά	Συστήματα που δρουν όπως ο άνθρωπος	Συστήματα που δρουν ορθολογικά
	<i>Η τέχνη της δημιουργίας μηχανών που κάνουν λειτουργίες, οι οποίες, όταν πραγματοποιούνται από ανθρώπους, απαιτούν νοημοσύνη</i>	<i>Η μελέτη της σχεδίασης ευφυών πρακτόρων</i>

Πίνακας 3.1: Ορισμοί της Τεχνητής Νοημοσύνης, σε συνάρτηση του τρόπου σκέψης και συμπεριφοράς του ανθρώπου.

Διαφορετικά, σκοπός της TN είναι η δημιουργία μηχανών ικανών να χειρίζονται στοιχεία και δεδομένα, ώστε να λαμβάνουν αποφάσεις για διάφορα θέματα με την ίδια ή και μεγαλύτερη ευχέρεια και με ανάλογη επιτυχία με ανθρώπους ειδικούς στα θέματα αυτά [15]. Πολλοί διαφορετικοί ορισμοί έχουν δοθεί για την TN από διαφορετικούς επιστήμονες. Στον Πίνακα

3.1, κωδικοποιούνται γνωστοί ορισμοί, σε συνάρτηση με τον τρόπο σκέψης και την συμπεριφορά του ανθρώπου. Οι υποστηρικτές τις μοντελοποίησης του ανθρώπινου νου, της ανθρώπινης σκέψης, δεν προχώρησαν όσο θα επιθυμούσαν. Επικράτησε η μοντελοποίηση της συμπεριφοράς. Ωστόσο, μέσα από αυτήν την προσπάθεια προέκυψε ένα νέο επιστημονικό πεδίο, η **γνωσιακή επιστήμη (cognitive science)**, που ασχολείται με τη μελέτη του νου.

Η επιστήμη της Τεχνητής Νοημοσύνης αξιοποιεί πολλούς επιστημονικούς κλάδους όπως, αυτόν της πληροφορικής, της ψυχολογίας, της φιλοσοφίας, της νευρολογίας, της γλωσσολογίας καθώς και της επιστήμης των μηχανικών. Χαρακτηριστικά επιτεύγματα και στόχοι της TN μπορούν να θεωρηθούν, η δυνατότητα των μηχανών να μαθαίνουν να εξελίσσονται (ικανότητα μάθησης), η δυνατότητα εξαγωγής συμπερασμάτων, η προσαρμογή στο περιβάλλον, η πρόβλεψη και η επίλυση πολύπλοκων προβλημάτων, η αντίληψη μέσω της όρασης, η κατανόηση φυσικής γλώσσας, κλπ.

Στην εποχή μας, η TN αποτελεί ένα από τα ταχύτερα εξελισσόμενα επιστημονικά πεδία, ενώ με τη χρήση εργαλείων των εφαρμοσμένων μαθηματικών και επιστήμες μηχανικών έχει ξεφύγει από τα πλαίσια της θεωρητικής πληροφορικής.

Στη συνέχεια του κεφαλαίου, γίνεται μια σύντομη παρουσίαση της επιστήμης της Τεχνητής Νοημοσύνης, στην οποία εντάσσονται τα συστήματα ταξινόμησης με Νευρωνικά Δίκτυα, τα οποία θα παρουσιάσουμε με μεγαλύτερη προσοχή στο επόμενο κεφάλαιο, αφού αποτελούν την βάση την παρούσας διπλωματικής εργασίας.

3.2 Εξελικτική Πορεία της Επιστήμης της TN

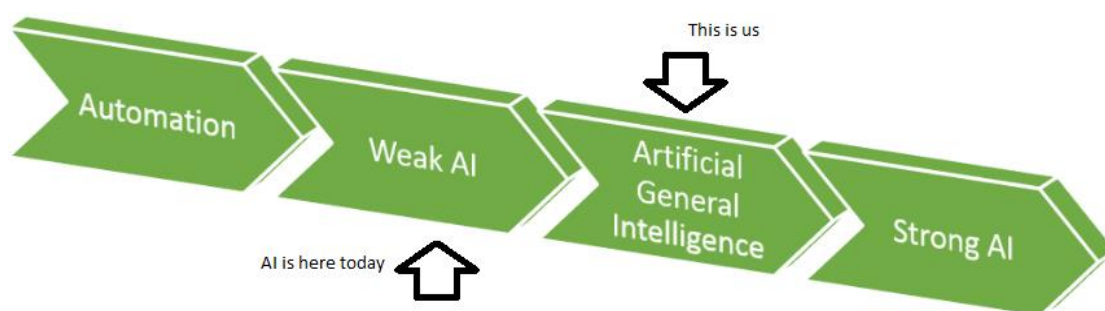
Αν και ο όρος Τεχνητή Νοημοσύνη μας παραπέμπει σε μία μηχανή που δημιουργήθηκε κατά τέτοιο τρόπο ώστε να έχει η ίδια τη δική της νοημοσύνη και να ακολουθεί τις αρχές λειτουργίας ενός ανθρώπου, οι μέχρι τώρα εξελίξεις στον τομέα δεν έχουν φτάσει στο επίπεδο αυτό. Ωστόσο, αν θέλουμε να δούμε την εξελικτική πορεία της Τεχνητής Νοημοσύνης, αναφερόμενοι και στο κοντινό μέλλον, μπορούμε να τη διακρίνουμε σε τρεις χαρακτηριστικούς σταθμούς: στην **Αδύναμη** Τεχνητή Νοημοσύνη, στην **Γενική** Τεχνητή Νοημοσύνη και στην **Ισχυρή** Τεχνητή Νοημοσύνη.

- **Αδύναμη Τεχνητή Νοημοσύνη (Weak Artificial Intelligence):** Έχουμε μηχανές που **δείχνουν να έχουν ευφυΐα**, και μπορούν να εκτελούν εργασίες περιορισμένου εύρους (Narrow AI). Στην Weak AI, η μηχανή είναι απλώς ένα καλά σχεδιασμένο εργαλείο το οποίο λειτουργεί υπό τις εντολές του προγραμματιστή δίχως τη δυνατότητα αντίληψης. Είναι σε κάθε περίπτωση ένας ισχυρός βοηθός για εργασίες περιορισμένου εύρους. Η λειτουργία της μπορεί να θεωρηθεί ότι ίσως να προσομοιώνει την ανθρώπινη νοημοσύνη όμως η μηχανή δεν διαθέτει νοημοσύνη παρόμοια αυτής του ανθρώπου. Βασιζόμενη στα δεδομένα, μπορεί να κάνει πολύ καλά μία μόνο εργασία. Οι δεξιότητες που μπορεί να αποκτήσει, για να εκτελέσει μια εργασία, δεν είναι **μεταβιβάσιμες** σε μια άλλη εργασία, κάτι που συνιστά μια σημαντική πτυχή της ανθρώπινης νοημοσύνης. Οι έως τώρα εφαρμογές της Τεχνητής Νοημοσύνης ανήκουν στην κατηγορία της Αδύναμης Τεχνητής Νοημοσύνης, αλλά φαίνεται να προχωράμε γρήγορα για το επόμενο στάδιο που είναι η Γενική τεχνητή Νοημοσύνη.
- **Γενική Τεχνητή Νοημοσύνη (General Artificial Intelligence):** αφορά μηχανές **με γενική ευφυΐα**, που θα μπορούν να μεταφέρουν εμπειρία από μια νοητική λειτουργία σε άλλη και είναι το επίπεδο που εντάσσεται η ανθρώπινη ευφυΐα. Η αρχική αναζήτηση της AI ήταν να δημιουργηθούν μηχανές που θα μπορούσαν να εμφανίσουν το ίδιο επίπεδο νοημοσύνης με τον άνθρωπο. Αυτό σήμερα αναφέρεται ως

Γενική Τεχνητή Νοημοσύνη. Αυτή αφορά την ικανότητα ενός τεχνολογικού συστήματος να μεταφέρει εμπειρία από μια νοητική λειτουργία που εκτελεί σε μία άλλη, ώστε να μην αρχίζει η ανάπτυξη του νέου συστήματος (εφαρμογής) από μηδενική βάση. Για παράδειγμα, αν η μηχανή γνωρίζει ελληνικά και θέλει να μάθει γερμανικά, να μπορεί να μεταφέρει εμπειρία από τη μία γλώσσα στην άλλη, να μπορεί να εκμεταλλεύεται το γεγονός ότι υπάρχει σύνταξη και κλίση σε όλες τις γλώσσες. Αν παίζει σκάκι, να μπορεί να αξιοποιήσει την εμπειρία (στη χάραξη στρατηγικής) για να παίζει πόκερ.

Σήμερα βρισκόμαστε ακόμα στο επίπεδο της Αδύναμης TN, (Σχήμα 3.1), όπως προαναφέρθηκε, γεγονός που σημαίνει ότι απαιτείται ένα τεράστιο τεχνολογικό άλμα, για να φτάσουμε στη Γενική TN. Βέβαια η εξέλιξη τα τελευταία χρόνια είναι, θα λέγαμε, εκθετική και αυτό επιτρέπει σε πολλούς επιστήμονες να προβλέπουν το άλμα να προκύψει εντός του 21^{ου} αιώνα. Τότε, οι μηχανές θα έχουν τη δυνατότητα να αυτοβελτιώνεται, η ευφυΐα τους θα αυξάνεται εκθετικά με αποτέλεσμα να φτάσουν και να ξεπεράσουν την ανθρώπινη νοημοσύνη. Το οριακό αυτό σημείο ονομάζεται «τεχνολογική ιδιαιτερότητα» - **singularity**.

Η Γενική Τεχνητή Νοημοσύνη θα είναι αυτή που θα χαράξει το δρόμο για το τελευταίο στάδιο εξέλιξης της Τεχνητής Νοημοσύνης που είναι η Ισχυρή Τεχνητή Νοημοσύνη.



Σχήμα 3.1: Εξελικτική Πορεία της Τεχνολογίας της TN.

- **Ισχυρή Τεχνητή Νοημοσύνη (Strong AI):** Θα έχουμε **μηχανές με ευφυΐα**. Θα εκτελούν με επιτυχία οποιοδήποτε νοητικό έργο που (κατ' ελάχιστον) μπορεί να κάνει ένας άνθρωπος. Η μηχανή θα μπορεί κυριολεκτικά να καταλαβαίνει και να λαμβάνει και άλλα γνωστικά επίπεδα και άρα δεν θα αποτελεί απλά προσομοίωση της ανθρώπινης αλλά αναπαραγωγή νοημοσύνης. Ισχυρή TN σημαίνει ότι τα μηχανήματα θα εκτελούν διαφορετικά καθήκοντα και όχι μόνο τα προκαθορισμένα, θα έχουν κοινή λογική και δημιουργικότητα. Πολύ πιθανό να αρχίσουν να θέτουν και τους δικούς τους στόχους.

Ακόμα, την εξελικτική πορεία της Τεχνολογίας της TN μέχρι σήμερα, συνοπτικά μπορεί να θεωρηθεί ως μια πορεία από την Κλασική Τεχνητή Νοημοσύνη στη Συνδεδετική TN:

- **Κλασική ή συμβολική (Symbolic AI):** Βασίζεται στην κατανόηση των νοητικών διεργασιών και ασχολείται με τη προσομοίωση της ανθρώπινης νοημοσύνης προσεγγίζοντάς την με αλγορίθμους και συστήματα που βασίζονται στη γνώση χρησιμοποιώντας ως δομικές μονάδες τα σύμβολα. Επιχειρεί να προσομοίωση την ανθρώπινη συμπεριφορά αλγοριθμικά, (π.χ. σύμβολα και λογικούς κανόνες).
- **Συνδεδετική (connectionist approach) ή μη-Συμβολική:** Βασίζεται στη μίμηση της βιολογικής λειτουργίας του εγκεφάλου, τη λειτουργία των δομικών του κύτταρων, προσεγγίζοντας τα με τα Τεχνητά Νευρωνικά Δίκτυα, τα οποία αποτελούν την μίμηση της βιολογικής λειτουργίας του εγκεφάλου.

Κύριο χαρακτηριστικό των συστημάτων που εφαρμόζουν τις αρχές της Τεχνητής Νοημοσύνης είναι η δυνατότητα τους να λαμβάνουν κάποια σήματα ως εισόδους, να τα επεξεργάζονται με τρόπο παρόμοιο με εκείνο της ανθρώπινης σκέψης και να εξάγουν κάποια συμπεράσματα ως έξοδο. Αντίστοιχα στους ανθρώπους, το άτομο επιλέγει και προσλαμβάνει τα ερεθίσματα (πληροφορίες) με τους αισθητηριακούς υποδοχείς, τα μετασχηματίζει σε μία μορφή μηνυμάτων που μπορούν να χρησιμοποιηθούν από το νοητικό (γνωστικό) σύστημα, τα επεξεργάζεται, τα συγκρατεί και τα χρησιμοποιεί όταν τα χρειάζεται. Αυτή η διαδικασία της απόκτησης, συγκράτησης και χρησιμοποίησης των πληροφοριών συγκροτεί το φαινόμενο της μάθησης.

Η Μηχανική Μάθηση αποτελεί κλάδο του ευρύτερου τομέα της Τεχνητής Νοημοσύνης, είναι ο τομέας της Επιστήμης που ασχολείται με την ανάπτυξη συστημάτων ικανά να μαθαίνουν από εμπειρία χωρίς να έχουν προγραμματιστεί ρητά από πριν.

3.3 Μηχανική Μάθηση

Η Μηχανική Μάθηση (machine learning), ένας από τους σημαντικότερους κλάδους της ΤΝ, αναφέρεται στη σχεδίαση και ανάπτυξη αλγορίθμων που να μαθαίνουν να αναγνωρίζουν μοτίβα σε δεδομένα, να λαμβάνουν αποφάσεις βασισμένες είτε στην στατιστική είτε στις πιθανότητες και να μπορούν να βελτιώνουν τη συμπεριφορά τους χρησιμοποιώντας την εμπειρία τους. Με άλλα λόγια, η δυνατότητα μάθησης να προσδιορίζεται ως η ικανότητα απόκτησης επιπλέον γνώσης, που επιφέρει μεταβολές στην υπάρχουσα καταχωρημένη γνώση είτε αλλάζοντας χαρακτηριστικά της είτε με αυξομειώσή της.

Στην περίπτωση των συστημάτων ΤΝ που ανήκουν στη Μη-Συμβολική (Συνδεδετική) ΤΝ, όπως η περίπτωση των **Τεχνητών Νευρωνικών Δικτύων** (ΤΝΔ), ως μάθηση προσδιορίζεται η δυνατότητα που διαθέτουν τα συστήματα στο να μετασχηματίζουν την εσωτερική τους δομή, παρά στο να μεταβάλλουν κατάλληλα τη γνώση που έχει καταχωρηθεί μέσα σε αυτά κατά το σχεδιασμό τους. Ακόμα, στον τομέα της μηχανικής μάθησης αναπτύσσεται, και η Εξελικτική Μάθηση (**Evolutionary Learning**), η οποία βασίζεται στην μίμηση των διαδικασιών φυσικής αναπαραγωγικής σε έμβια όντα (γενετικοί αλγόριθμοι) [42-43].

3.4 Νευρωνικά Δίκτυα

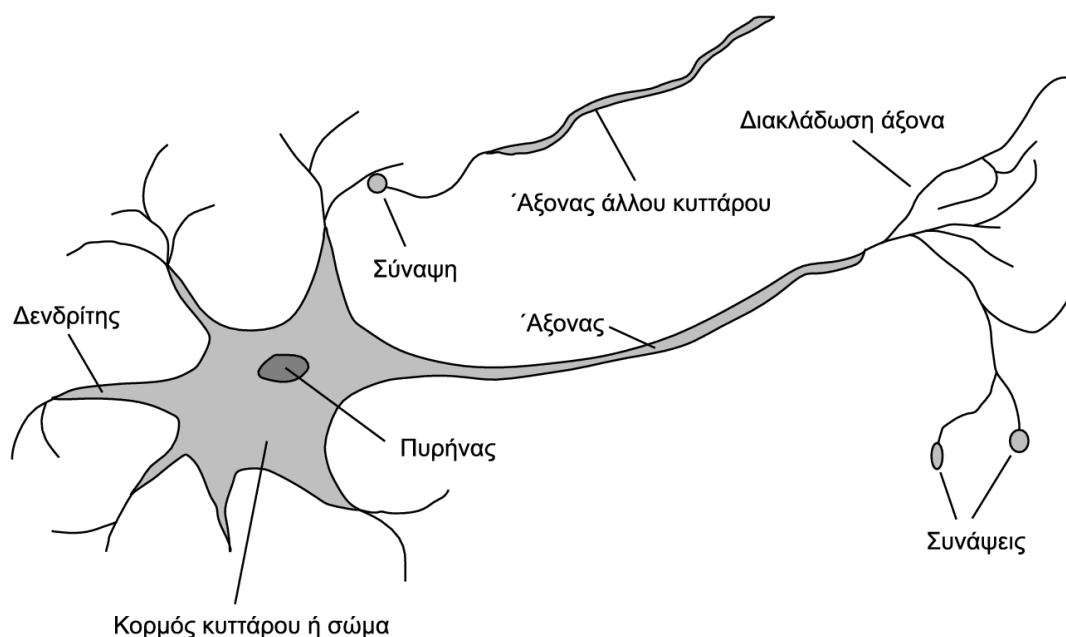
Τα **Τεχνητά Νευρωνικά Δίκτυα** πρωτοπαρουσιάστηκαν ως μαθηματικά μοντέλα τα οποία προσομοιώνουν την πολύπλοκη λειτουργία των νευρώνων του ανθρώπινου εγκεφάλου. Μιμούνται, κατά το δυνατό, εκείνα του **Βιολογικού Νευρωνικού Δικτύου** (ΒΝΔ) του ανθρώπινου εγκεφάλου. Με τον όρο λοιπόν Τεχνητό Νευρωνικό Δίκτυο εννοούμε ένα μαθηματικό μοντέλο που προσπαθεί να μιμηθεί τη δομή και τη λειτουργία ενός Βιολογικού Νευρωνικού δικτύου.

3.4.1 Βιολογικοί Νευρώνες

Το **Βιολογικό Νευρωνικό Δίκτυο** (ΒΝΔ) αποτέλεσε το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί [44]. Ο νευρώνας είναι ένας εξειδικευμένος τύπος κυττάρου, που αποτελεί τη βασική μονάδα των συστημάτων επεξεργασίας πληροφοριών που απαρτίζουν το νευρικό σύστημα. Έτσι, σε

αναλογία με τους Ηλεκτρονικούς Υπολογιστές, οι νευρώνες αποτελούν το σύνολο των στοιχείων μνήμης, λογικών κυκλωμάτων και εντολών λειτουργίας του εγκεφάλου.

Τα ΒΝΔ αποτελούνται από το βιολογικό ιστό, χημικές ουσίες και ηλεκτρικά σήματα και αποτελούν μέρος του κεντρικού νευρικού συστήματος του εγκεφάλου. Η δομή και οι συνδέσεις του εγκεφάλου είναι τέτοιες ώστε να καθιστούν δυνατή την παράλληλη επεξεργασία δεδομένων και τη δυνατότητα συνεχούς μάθησης μέσω αλληλεπίδρασης με το περιβάλλον. Αυτά τα δύο θεμελιώδη χαρακτηριστικά συμβάλλουν στην ικανότητα εκτέλεσης δύσκολων καθηκόντων (γρήγορη αναγνώριση πολύπλοκων μορφών, ταξινόμηση αντικειμένων κ.ά.), και στη συνεχή εξέλιξη και προσαρμογή, δηλαδή μαθαίνοντας από το περιβάλλον κατά την αλληλεπίδρασή του με αυτό.



Σχήμα 3.2: Παράσταση ενός Βιολογικού Νευρωνικού Δικτύου.

Το ανθρώπινο σώμα αποτελείται από ένα τεράστιο αριθμό κυττάρων. Το νευρικό κύτταρο, η δομική και λειτουργική μονάδα του νευρικού συστήματος, αποτελεί μια στοιχειώδη μονάδα μεταφοράς και επεξεργασίας ηλεκτρικών σημάτων και χημικών διαδικασιών. Ένας νευρώνας αποτελείται από ένα κυτταρικό σώμα (περικάρυο) και τις αποφύσεις του, έναν άξονα και έναν ή περισσότερους δενδρίτες. (Σχήμα 3.2) Οι δενδρίτες επιτρέπουν την επικοινωνία μεταξύ των νευρώνων, λαμβάνοντας ηλεκτρικούς παλμούς από τον άξονα άλλων νευρώνων μέσω χημικών διεργασιών. Τα σημεία ανταλλαγής του ηλεκτρικού δυναμικού μεταξύ των νευρώνων ονομάζεται **συνάψεις**. Δηλαδή, οι συνάψεις ενώνουν τους βιολογικούς νευρώνες και δίνουν την δυνατότητα ανταλλαγής του ηλεκτρικού δυναμικού μεταξύ των δενδριτών. Οι ανταλλαγές αυτές λαμβάνουν χώρα σε μεγάλη πυκνότητα με πολύ αργό ρυθμό περίπου 100Hz. Ο μέσος άνθρωπος έχει περίπου 10^{14} συνάψεις και επομένως συμβαίνουν 10^{16} αλληλεπιδράσεις ανά δευτερόλεπτο (interactions/sec).

Στο σώμα του νευρώνα γίνεται η επεξεργασία όλων των σημάτων που έχουν ληφθεί και στη συνέχεια αποστέλλεται το επεξεργασμένο σήμα μέσω του άξονα σε όλους τους γειτονικούς νευρώνες. Υπολογίζεται πως ο φλοιός του ανθρώπινου εγκεφάλου αποτελείται από 10δισ νευρώνες και 60τρις συνδέσεις μεταξύ τους (κάθε νευρώνας συνδέεται με χιλιάδες γειτονικούς).

Συνεπώς, παρ' ότι η επεξεργαστική ισχύς του νευρώνα είναι πολύ μικρότερη από αυτή που διαθέτουν οι σύγχρονες ψηφιακές λογικές πύλες, οι δυνατότητες του εγκεφάλου ξεπερνούν κάθε υπολογιστικό σύστημα που έχει σήμερα κατασκευαστεί [16]. Οι σημερινοί υπολογιστές προσεγγίζουν τις 10^9 interactions/sec.

3.4.2 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), (Artificial Neural Network -ANN) αποτελούν κλάδο της Τεχνητής Νοημοσύνης. Το όνομά τους προέρχεται από τα δομικά κύτταρα του εγκεφάλου, τους νευρώνες (ΒΝΔ), λόγω του ότι προσπαθούν να προσομοιώσουν τον τρόπο λειτουργίας τους. Το Νευρωνικό Δίκτυο είναι ένας προσαρμοστικός μηχανισμός με την έννοια ότι αλλάζει τη δομή του ανάλογα με τις πληροφορίες που έχει να επεξεργαστεί. Τα βασικά σημεία που προσομοιάζει συγκριτικά με τη λειτουργία του ανθρώπινου εγκεφάλου είναι:

- Η γνώση του ΤΝΔ αποκτάται μέσω μιας επαναληπτικής διαδικασίας εκπαίδευσης, γνωστή ως μηχανική μάθηση (Machine Learning).
- Οι σύνδεσμοι μεταξύ των νευρώνων που ονομάζονται βάρη χρησιμοποιούνται για να αποθηκευτεί αυτή η γνώση.

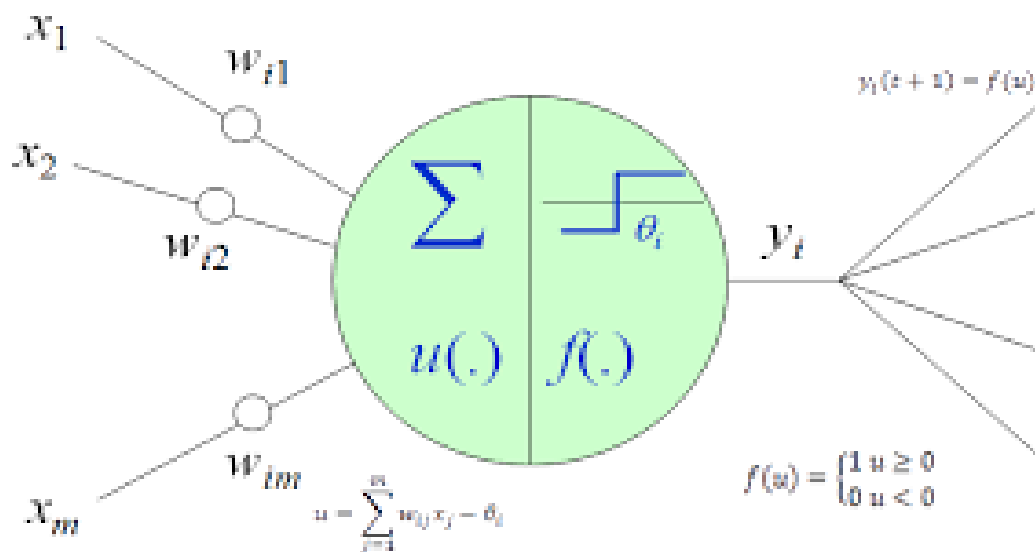
Έτσι, κατ' αντιστοιχία με ένα βιολογικό δίκτυο νευρώνων, ένα τεχνητό Νευρωνικό δίκτυο δομείται από ένα σύνολο τεχνητών νευρώνων που επικοινωνούν και αλληλοεπιδρούν, συνδεδεμένοι μεταξύ τους με τις λεγόμενες συνάψεις (synapses). Κάθε σύναψη έχει διαφορετικό βαθμό αλληλεπίδρασης, ο οποίος καθορίζεται από τα συναπτικά βάρη (synaptic weights), τα οποία μεταβάλλονται (επαναπροσδιορίζονται) κατά τη διαδικασία της εκπαίδευσης. Πιο συγκεκριμένα, καθώς το νευρωνικό δίκτυο επικοινωνεί και προσαρμόζεται με το περιβάλλον ώστε να εκπαιδευτεί από αυτό, τα συναπτικά βάρη μεταβάλλονται συνεχώς, ενδυναμώνοντας ή αποδυναμώνοντας την ισχύ του κάθε δεσμού. Επομένως, αποτελούν κωδικοποιημένη την εμπειρική γνώση που αποκτά το νευρωνικό δίκτυο από την αλληλεπίδρασή του με το περιβάλλον. Αυτό είναι και το χαρακτηριστικό εκείνο που προσδίδει στο δίκτυο την ικανότητα για εξέλιξη και προσαρμογή στο περιβάλλον.

Το κύριο πλεονέκτημα των νευρωνικών δικτύων είναι ότι έχουν τη δυνατότητα να αποθηκεύσουν γνώση και εμπειρία από το περιβάλλον, την οποία στη συνέχεια μπορούν να ανακαλέσουν. Επιπλέον, ένα τεχνητό νευρωνικό δίκτυο προσφέρει την δυνατότητα γενίκευσης μίας εισόδου, δηλαδή την εξαγωγή των βασικών χαρακτηριστικών ενός συστήματος, ακόμα και από θορυβώδη δεδομένα.

Ο τεχνητός νευρώνας αποτελεί θεμελιώδες στοιχείο κάθε νευρωνικού δικτύου. Ο κάθε νευρώνας επεξεργάζεται ουσιαστικά την είσοδο, καθώς δέχεται δεδομένα, τα οποία αναλύει και δίνει μία τιμή εξόδου. Οι εισοδοί του ενδέχεται να είναι είτε οι εξοδοί άλλων νευρώνων, είτε το αρχικό σήμα εισόδου του δικτύου.

Στο νευρώνα του Σχήματος 3.3, η πληροφορία διαδίδεται πάντα από αριστερά προς τα δεξιά. Κάθε τιμή εισόδου $x = (x_1, x_2, \dots, x_n)$ (διάνυσμα εισόδου εκπαίδευσης) πολλαπλασιάζεται με τους αντίστοιχους **συντελεστές βαρών** $W=(w_1, w_2, \dots, w_n)$ και στη συνέχεια αθροίζονται μαζί με μια σταθερή μονάδα την **προκατάληψη** b (bias unit). Αυτό το άθροισμα θα δοθεί στη συνέχεια ως όρισμα στη **συνάρτηση ενεργοποίησης (activation function)** $f(x)$ η οποία και θα καθορίσει αν θα πυροδοτείτε ο νευρώνας.

$$f(x) = \sum_{i=1}^n x_i w_i + b$$



Σχήμα 3.3: Παράσταση ενός Τεχνητού Νευρωνικού Δικτύου.

3.4.2.1 Συναρτήσεις Ενεργοποίησης

Η Συνάρτηση Ενεργοποίησης (**activation function**), ανάλογα με το όρισμα που δέχεται αποφασίζει για το αν ένας νευρώνας ενεργοποιείται ή όχι. Παρακάτω δίνονται ορισμένες συναρτήσεις ενεργοποίησης για νευρωνικά δίκτυα.

- **Διαδική συνάρτηση (binary):** Η έξοδος του νευρώνα παίρνει την τιμή 0, όταν το όρισμα της συνάρτησης ενεργοποίησης έχει αρνητικές τιμές, ενώ όταν πάρει μηδενικές ή θετικές τιμές τότε η συνάρτηση ενεργοποίησης γίνεται 1:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- **Γραμμική συνάρτηση (linear):** Η έξοδος του νευρώνα ισούται με την τιμή του ορίσματος:

$$f(x) = C * x$$

- **Συνάρτηση εφαπτομένης (tanh):** Η σιγμοειδής συνάρτηση υπερβολικής εφαπτομένης (tanh) είναι παρόμοιας λογικής με την απλή σιγμοειδή συνάρτηση και η έξοδος του νευρώνα παίρνει τιμές από -1 έως 1:

$$f(x) = \tanh(x) = 2\sigma(2x) - 1$$

- **Συνάρτηση ReLU (rectified linear unit):** Η Συνάρτηση ReLU ή Γραμμική Μονάδα Επιδιόρθωσης, είναι από τις πιο γνωστές συναρτήσεις ενεργοποίησης. Η έξοδος του νευρώνα παίρνει την τιμή από 0, όταν το όρισμα είναι αρνητικό, ενώ όταν το όρισμα είναι θετικό τότε η συνάρτηση ενεργοποίησης παραμένει ίδια:

$$f(x) = \max(0, x)$$

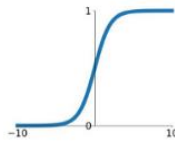
- **Λογιστική / Σιγμοειδής συνάρτηση (Sigmoid):** Η λογαριθμική σιγμοειδής συνάρτηση είναι μη γραμμική και η έξοδος του νευρώνα παίρνει την τιμή από 0 έως 1. Επίσης, όταν η συνάρτηση τείνει στο άπειρο τότε οι μεγάλες αρνητικές τιμές γίνονται μηδέν και οι μεγάλες θετικές γίνονται 1:

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

Στον Πίνακα 3.1, δίνονται συγκεντρωτικά χαρακτηριστικές συναρτήσεις ενεργοποίησης Νευρωνικών δικτύων.

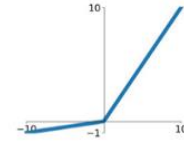
Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



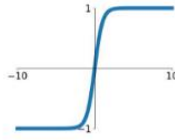
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

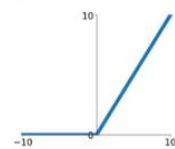


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

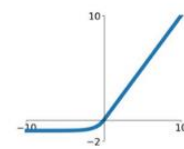
ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

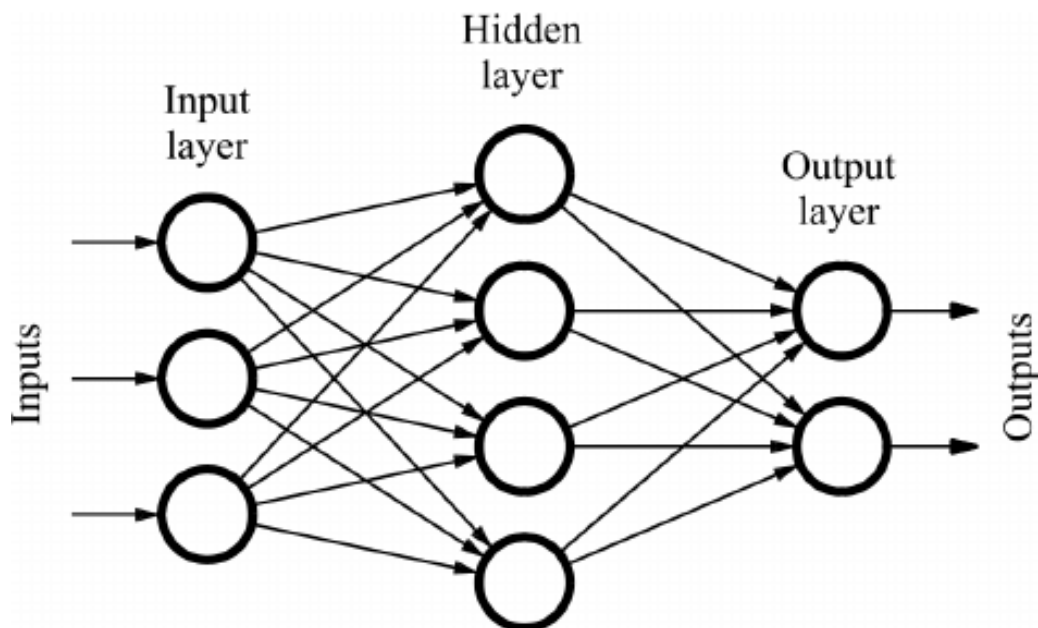


Πίνακας 3.1: Χαρακτηριστικές συναρτήσεις ενεργοποίησης Νευρωνικών δικτύων.

3.4.2.2 Πολυεπίπεδα ΤΝΔ

Κοινό χαρακτηριστικό της δομής των πολυεπίπεδων ΤΝΔ είναι ότι διαθέτουν τουλάχιστον ένα κρυφό επίπεδο. Η στοιχειώδης δομή ενός ΤΝΔ έχει τα ακόλουθα διαφορετικά επίπεδα (Σχήμα 3.4):

- Το **επίπεδο της εισόδου (input layer)** αποτελεί το 1^ο επίπεδο ενός ΤΝΔ, σε αυτό εμφανίζονται τα δεδομένα εισόδου και ο αριθμός των νευρώνων του είναι ίσος με τις μεταβλητές των δεδομένων.
- Το **κρυφό επίπεδο (hidden layer)**, το ενδιάμεσο επίπεδο μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου. Ωστόσο, ο αριθμός των κρυφών επιπέδων που περιέχονται σε ένα ΤΝΔ, ορίζονται από τον σχεδιαστή και επηρεάζει το «βάθος» του δικτύου.
- Το **επίπεδο εξόδου (output layer)** αποτελεί το τελευταίο επίπεδο ενός ΤΝΔ. Σε αυτά εμφανίζονται τα αποτελέσματα μετά το τέλος της εκπαίδευσης και ο αριθμός των νευρώνων του είναι ίσος με τις μεταβλητές εξόδου των αποτελεσμάτων.



Σχήμα 3.4: Παράσταση Τεχνητού Νευρωνικού Δικτύου 3^{ων} επιπέδων.

Στο επίπεδο εισόδου δεν υπάρχει συνάρτηση ενεργοποίησης και απλά χρησιμοποιείτε για να λάβει την είσοδο (x_i) και να την μεταδώσει στο επόμενο επίπεδο. Αν κάποιο ΤΝΔ περιέχει και ενδιάμεσα στρώματα τότε αυτό ονομάζεται πολύ-επίπεδο νευρωνικό δίκτυο και τα ενδιάμεσα επίπεδα κρυφά (**hidden layers**). Τα επίπεδα δεν είναι αναγκαίο να έχουν το ίδιο πλήθος νευρώνων μεταξύ τους αλλά οι νευρώνες του ίδιου επιπέδου πρέπει να έχουν τις ίδιες συναρτήσεις ενεργοποίησης.

Τα ΤΝΔ ανάλογα με το είδος της σύνδεσης που υπάρχει μεταξύ των νευρώνων στο κάθε επίπεδο, χωρίζονται σε **μερικώς συνδεδεμένα (partially connected)**, αν δεν υπάρχουν συνδέσεις μεταξύ όλων των νευρώνων του επιπέδου, και **πλήρως συνδεδεμένα (fully connected)**, αν κάθε νευρώνας ενός επιπέδου είναι συνδεδεμένος με το επόμενο επίπεδο.

Τα ΤΝΔ κατηγοριοποιούνται ανάλογα με την αρχιτεκτονική τους και τον τρόπο με τον οποίο συνδέονται οι νευρώνες μεταξύ τους. Τα πιο συνηθισμένα δίκτυα ονομάζονται **πρόσθιας τροφοδότησης (feedforward)** νευρωνικά δίκτυα, το οποίο σημαίνει ότι η πληροφορία εντός του δικτύου είναι πάντα προς τα εμπρός και δεν επιτρέπεται η προς τα πίσω τροφοδότηση του δικτύου. Μεταφέρεται η πληροφορία από την είσοδο του δικτύου μέχρι την έξοδο χωρίς δυνατότητα οπισθοδιάδοσης. Οι νευρώνες κάθε επιπέδου έχουν σαν εισόδους τα σήματα εξόδου του προηγούμενου επιπέδου.

Ωστόσο, υπάρχουν άλλα είδη νευρωνικών δικτύων, στα οποία επιτρέπεται η προς τα πίσω τροφοδότηση, τα οποία ονομάζονται **οπίσθιας τροφοδότησης, Recurrent Neural Networks (RNN)**. Στα οπίσθιας τροφοδότησης δίκτυα υπάρχουν συνδέσεις και μεταξύ νευρώνων του ίδιου επιπέδου ή και προηγούμενων. Αν η ανατροφοδότηση αφορά κόμβους στο ίδιο επίπεδο, τότε τα δίκτυα καλούνται αυτοσυσχετιζόμενες μνήμες (**autoassociated memories**) διαφορετικά, καλούνται ετεροσυσχετιζόμενες μνήμες (**heteroassociated memories**).

3.4.2.3 Αλγόριθμος Οπισθοδιάδοσης (Backpropagation Algorithm)

Σημαντικό ρόλο στην απόδοση των ΣΝΔ διαδραματίζει η δυναμική σχέση των σφαλμάτων στα αποτελέσματα με τις τιμές των συναπτικών βαρών (synaptic weights).. Για το λόγο αυτό, έχουν

προταθεί αρκετές μέθοδοι ανάδρασης σφάλματος που βελτιστοποιηθούν την απόδοση του δικτύου ελαχιστοποιώντας το σφάλμα.

Χαρακτηριστικός είναι ο **αλγόριθμος οπισθοδιάδοσης σφάλματος** (Backpropagation Algorithm), ο οποίος έχει αντικείμενο ακριβώς αυτή τη βελτιστοποίηση μέσω ελαχιστοποίησης του σφάλματος. Χρησιμοποιείται κυρίως στην επιβλεπόμενη μάθηση, δηλαδή στην περίπτωση που έχουμε ένα επισημασμένο σύνολο δεδομένων για εκπαίδευση. Η λογική αυτού του τρόπου εκμάθησης είναι πως χρησιμοποιούμε την έξοδο του δικτύου μετά τη διαδικασία της πρόσθιας-τροφοδότησης για να τη συγκρίνουμε με την επιθυμητή (γνωστή) έξοδο, ώστε να μετρήσουμε το λάθος του δικτύου.

Το λάθος υπολογίζεται από τη **συνάρτηση κόστους J** και στη συνέχεια διαδίδεται στο δίκτυο από το τελευταίο επίπεδο διαδοχικά προς τα πίσω με στόχο την κατάλληλη ρύθμιση των παραμέτρων έτσι ώστε να ελαχιστοποιηθεί η J. Ουσιαστικά, η διαδικασία εκπαίδευσης προσπαθεί να ρυθμίσει τους **συντελεστές βαρών W** και τις **σταθερές προκατάληψης b**, έτσι ώστε να ελαχιστοποιηθεί η διαφορά μεταξύ των προβλέψεων του δικτύου και των πραγματικών τιμών.

Συγκεκριμένα, μέσω τις εφαρμογής του αλγορίθμου, αναπροσαρμόζονται τα βάρη στους νευρώνες των κρυφών επιπέδων για να ελαχιστοποιηθεί το λάθος. Το λάθος υπολογίζεται από μια **συνάρτηση κόστους J(x)**, το οποίο διαδίδεται από το τελευταίο επίπεδο προς το πρώτο, με στόχο την ελαχιστοποίησή της.

Η συνάρτηση κόστους **J(W, b; x, y)**, υπολογίζεται συναρτήσει των **βαρών (w)**, **συντελεστών προκατάληψης (b)**, των **εισόδων (x₁, x₂, x₃, ..., x_n)**, και των **εξόδων (y₁, y₂, y₃, ..., y_n)**. Η εκτίμηση των συναρτήσεων κόστους μπορεί να γίνει από διαφορετικές προσεγγίσεις. Χαρακτηριστικές εκτιμήσεις των συναρτήσεων κόστους είναι οι ακόλουθες:

- **Μέσο τετραγωνικό σφάλμα** (mean squared error):

$$j(W, b; x, y) = \frac{1}{2} \sum_{i=1}^n (a_i - y_i)^2$$

- **Σφάλμα διασταυρούμενης εντροπίας** (cross-entropy cost):

$$j(W, b; x, y) = - \sum_{j=0}^n (y_j \ln(a_j) + (1 - y_j) \ln(1 - a_j))$$

- **Εκθετικό σφάλμα** (exponential cost):

$$j(W, b; x, y) = \exp(\sum_{i=0}^n (a_i - y_i)^2).$$

Τα δε βήματα εφαρμογής του αλγορίθμου, μπορούν να συνοψιστούν στα ακόλουθα:

- Αρχικοποιούμε όλες τις τιμές, με τυχαίες τιμές κανονικής κατανομής. Τα βάρη (W) τα αρχικοποιούμε με μικρές μη-μηδενικές τιμές, (μηδενικές τιμές παρουσιάζουν προβλήματα στη μάθηση). Στο πρώτο (πρόσθιο) πέρασμα υπολογίζουμε τις τιμές των συναρτήσεων ενεργοποίησης σε όλα τα επίπεδα του νευρωνικού δικτύου.
- Για το κάθε επίπεδο k και για τα επόμενα k+1, θεωρούμε:

$$\delta^{(k)} = ((W^{(k)})^T * \delta^{(k+1)}) * f'(z^{(k)})$$

και για το κάθε επίπεδο εξόδου kn, θεωρούμε:

$$\delta^{(n_i)} = -(y - a^{(n_i)}) * f'(z^{(n_i)})$$

- Υπολογίζουμε τις μερικές παραγώγους της $j(W, b; x, y)$, ως προς W και b .

$$\nabla_W j(W, b; x, y) = \delta^{(k+1)}(a^{(k)})^T$$

$$\nabla_b j(W, b; x, y) = \delta^{(k+1)}$$

- Ανανεώνουμε τις παραμέτρους σύμφωνα με τη μέθοδο gradient descend:

$$W^{(k)} := W^{(k)} - \alpha \frac{\partial j(W, b; x, y)}{\partial W^{(k)}}$$

$$b^{(k)} := b^{(k)} - \alpha \frac{\partial j(W, b; x, y)}{\partial b^{(k)}}$$

3.4.2.4 Βελτιστοποίηση: Stochastic Gradient Descent

Ο αλγόριθμος backpropagation, θεωρεί ότι στο δίκτυο περνά το σύνολο των δειγμάτων της εκπαίδευσης και στο τέλος υπολογίζεται η ανανέωση των βαρών. Όταν συμβαίνει αυτό τότε ουσιαστικά εφαρμόζουμε μεθόδους υπολογισμού με πακέτα (batch methods). Στην πραγματικότητα, με δεδομένα εκπαίδευσης για πολλές κλάσεις, η τεχνική γίνεται όλο και περισσότερο μη πρακτική γιατί αυτό απαιτεί πάρα πολλούς υπολογισμούς για ένα και μόνο βήμα στον υπολογισμό της νέας κατεύθυνσης των βαρών. Για το λόγο αυτόν προτείνεται η χρήση της τεχνικής **Stochastic Gradient Descent (SGD)** που εξαλείφει τα παραπάνω προβλήματα. Η SGD υπολογίζει τη νέα κλίση όχι υποβάλλοντας ολόκληρο το σετ αλλά ένα τυχαίο υποσύνολο του.

Ο Stochastic Gradient Descent είναι ένας επαναληπτικός αλγόριθμος ο οποίος μειώνει τα σφάλματα σε κάθε νευρώνα, ακολουθώντας την αρνητική κλίση για κάθε μεμονωμένο δείγμα, αλλά όχι σε ολόκληρα πακέτα (batches). Με τη χρήση της SGD ενημερώνονται οι παράμετροι θ (βάρη και κατώφλια) της συνάρτησης κόστους, βάση του παρακάτω τύπου:

$$\theta = \theta - \alpha \nabla_{\theta} j(W, \theta, b; x, y)$$

όπου, α ο ρυθμός μάθησης που μπορεί να είναι η παράμετρος που μεταβάλλεται με την πάροδο των εποχών εκπαίδευσης.

Ωστόσο, η κλασικές SGD μέθοδοι, για τη διαδικασία της ελαχιστοποίησης, χρησιμοποιούν ολόκληρα πακέτα. Η ενημέρωση των παραμέτρων της γίνεται:

$$\theta = \theta - \alpha \nabla_{\theta} E[J(\theta)]$$

3.4.2.5 Ορμή Σύγκλισης (Momentum)

Για να αποφευχθεί η περίπτωση στην οποία το σύστημα να φτάσει σε ατέρμονα βρόγχο, προσθέτουμε στην SGD μια επιπλέον παράμετρο, την **Ορμή**. Αυτό γίνεται όταν η συνάρτηση ενεργοποίησης παίρνει ακραίες οριακές τιμές στο τοπικό μέγιστο, και κατευθύνει την SGD να ταλαντώνεται καθώς η αρνητική κλίση θα δείχνει συνέχεια προς το ελάχιστο, χωρίς να δίνει κάποιο αποτέλεσμα.

$$\mathbf{v} = \gamma \mathbf{v} + \alpha \nabla_{\theta} j(W, \theta, b; \mathbf{x}, y)$$

$$\theta = \theta - \mathbf{v}$$

όπου, γ ο αριθμός επαναλήψεων, ενώ το \mathbf{v} είναι η εκάστοτε ταχύτητα-διάνυσμα της διάστασης της παραμέτρου θ .

3.4.3 Διαδικασία Εκπαίδευσης (Μάθησης) Νευρωνικών Δικτύων

Για να κατασκευάσουμε ένα νευρωνικό δίκτυο, πρέπει πρώτα να ορίσουμε την τοπολογία του, δηλαδή τον αριθμό των νευρώνων που θα το απαρτίζουν και τη δομή τους. Στη συνέχεια, πρέπει να υπολογίσουμε τους συντελεστές βαρύτητας των συνάψεων (τη μνήμη του νευρωνικού δικτύου). Αυτό που αναφέρεται ως διαδικασία εκπαίδευσης, υπολογίζει τους συντελεστές αυτούς, συνήθως από ένα σύνολο παραδειγμάτων (δεδομένα εκπαίδευσης). Ανάλογα με το πόσες πληροφορίες έχουμε για το τελικό αποτέλεσμα, οι αλγόριθμοι της εκπαίδευσης (ή **μάθησης**) χωρίζονται στις εξής κατηγορίες:

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Το δίκτυο τροφοδοτείται με συγκεκριμένες εισόδους (labeled), οι οποίες συγκρίνονται με τις αναμενόμενες εξόδους. Ο αλγόριθμος κατασκευάζει ένα μοντέλο που αντιστοιχίζει τις δεδομένες εισόδους (labeled examples) σε γνωστές επιθυμητές εξόδους (σύνολο εκπαίδευσης), κάνοντας προβλέψεις και διορθώνοντας τις προβλέψεις σε περίπτωση λάθους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής για εισόδους με άγνωστη έξοδο (δεδομένα ελέγχου).
- **Μη Επιβλεπόμενη Μάθηση (unsupervised learning):** Η μονάδα εξόδου εκπαιδεύεται να ανταποκρίνεται σε ομάδες προτύπων που δίνονται στην είσοδο. Με αυτό το τρόπο το σύστημα ανακαλύπτει χαρακτηριστικά των προτύπων των εισόδων. Ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων χωρίς να γνωρίζει επιθυμητές εξόδους (unlabeled examples) για το σύνολο εκπαίδευσης εξάγοντας γενικούς κανόνες.
- **Ημι-επιτηρούμενη μάθηση (semi-supervised learning):** Αποτελεί συνδυασμό των παραπάνω προσεγγίσεων. Τα δεδομένα εκπαίδευσης είναι μια μίξη γνωστών και αγνώστων δειγμάτων (labeled & unlabeled examples), όπου υπάρχει ένα επιθυμητό πρόβλημα πρόβλεψης, αλλά το μοντέλο πρέπει να δημιουργήσει δομές για να οργανώσει τα δεδομένα και να κάνει προβλέψεις.

Ο αλγόριθμος εκπαίδευσης, χρησιμοποιεί ως είσοδο ένα σύνολο **δεδομένων εκπαίδευσης (training set)** για τη δημιουργία νέας γνώσης. Ωστόσο, μπορεί να χρησιμοποιήσει προ-υπάρχουσα γνώση, και πάνω σε αυτήν να «χτίσουμε» τη νέα. Την εκπαίδευση (training) ακολουθεί η φάση **ελέγχου (test)** της παραγόμενης νέας γνώσης. Συνήθως, ο έλεγχος πραγματοποιείται αρχικά από τον ίδιο τον αλγόριθμο, μέσω διαδικασιών ανάκλησης με τη βοήθεια **δεδομένων ελέγχου (test data)**.

3.4.4 Σύγχρονες Τάσεις στα Τεχνητά Νευρωνικά Δίκτυα

Θα μπορούσαμε τους αλγόριθμους των Τεχνητών Νευρωνικών Δικτύων να τους ομαδοποιήσουμε σε κατηγορίες (γενιές), σύμφωνα με την περίοδο που εμφανίστηκαν και την σπουδαιότητά τους:

- Στην **1^η γενιά ΤΝΔ**: εντάσσονται τα δίκτυα που παρουσιάστηκαν παραπάνω. Είναι αυτά που εστιάζονται, κυρίως, σε εφαρμογές στο N-διάστατο Ευκλείδειο χώρο, περιλαμβανομένων των εφαρμογών επεξεργασίας σημάτων και αναγνώρισης προτύπων.

Ακόμα, δίνουν ιδιαίτερη έμφαση σε ανάλυση κύριων συνιστωσών (PCA), σε ανάλυση ανεξάρτητων συνιστωσών (independent component analysis), κ.α.

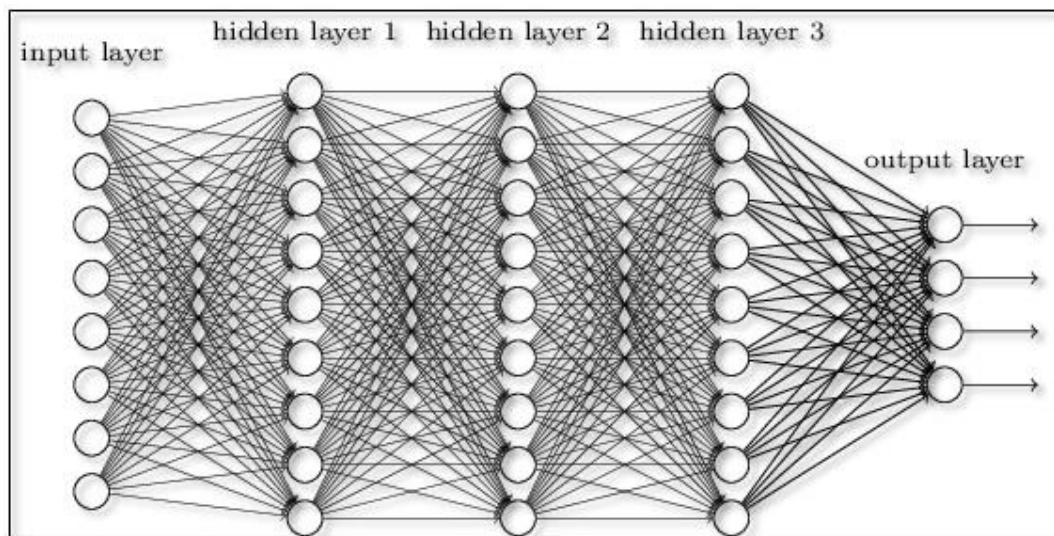
- Η 2^η γενιά ΤΝΔ: περιλαμβάνει αρχιτεκτονικές **Βαθιάς Μάθησης (deep learning)** που σήμερα είναι από τις πλέον δημοφιλείς στα ΤΝΔ. Από το 2006 και ύστερα [17], άρχισε μία σημαντική ανάπτυξη στο σχεδιασμό και τη δημιουργία ΤΝΔ με αρχιτεκτονικές βαθιάς μάθησης (Deep Learning Architectures). Το χαρακτηριστικό της βαθιάς μάθησης είναι η χρήση υπολογιστικών αρχιτεκτονικών ΤΝΔ με πολλαπλά στρώματα νευρώνων, (κρυφά στρώματα), ώστε να καθίσταται δυνατή η μάθηση αναπαραστάσεων με πολλαπλά επίπεδα. Να μπορούν να αλλάζουν τις εσωτερικές τους παραμέτρους υπολογίζοντας την αναπαράσταση κάθε στρώματος από την αναπαράσταση του προηγούμενου στρώματος. Έχουν προταθεί διάφορες αρχιτεκτονικές βαθιάς μάθησης, π.χ. τα **Συνελικτικά Νευρωνικά Δίκτυα** (convolutional neural networks), τα Αναδρομικά ΤΝΔ (RNN) κ.ά. Η πρακτική εφαρμογή τους έχει βελτιώσει σημαντικά την ταχύτητα καθώς και την εξόρυξη γνώσης σε μια ποικιλία προβλημάτων αναγνώρισης προτύπων. Για παράδειγμα, τα Συνελικτικά Νευρωνικά Δίκτυα έχουν επιφέρει βελτιώσεις κυρίως σε προβλήματα επεξεργασίας εικόνων, βίντεο, ομιλίας και ήχου, ενώ τα Αναδρομικά ΤΝΔ (RNN) έχουν επιδείξει αποτελεσματικότητα κυρίως στην επεξεργασία σειριακών δεδομένων (sequential data) και δενδρικές δομές σε εφαρμογές επεξεργασίας της φυσικής γλώσσας.
- Η 3^η γενιά ΤΝΔ: περιλαμβάνει αρχιτεκτονικές **ακιδωτών ΤΝΔ (spiking ANN)**, όπου τα σήματα μεταφέρονται με εκφορτίσεις χρονικών παλμών [38]. Σημασία δεν έχει μόνον η διάρκεια του κάθε χρονικού παλμού, αλλά και η συχνότητα των χρονικών παλμών μεταξύ των νευρώνων. Ένα σημαντικό χαρακτηριστικό των ακιδωτών ΤΝΔ είναι ότι προσομοιώνουν τον τρόπο λειτουργίας των βιολογικών κυττάρων του εγκεφάλου και υπάρχουν βάσιμες προσδοκίες για καλύτερη αναπαράσταση του χωρο-χρόνου.
- Η 4^η γενιά ΤΝΔ: αποτελεί μια μελλοντική προσέγγιση της εξέλιξης των ΤΝΔ. Μεταγενέστερα, ενδέχεται να υλοποιούν μη-συμβατικούς υπολογισμούς ή/και να εφαρμόζονται πέραν του Ν-διάστατου Ευκλείδειου χώρου. Κβαντικοί υπολογισμοί έχουν ήδη προταθεί για εκθετική επιτάχυνση των υπολογισμών [39]. Περαιτέρω, έχουν προταθεί υπολογισμοί με σημασιολογίες (semantics) οι οποίες παριστάνονται με μερικώς διατεταγμένες σχέσεις (partially ordered relations) [40] τέτοιου είδους υπολογισμοί, θα μπορούσαν να βελτιώσουν την αλληλεπίδραση ανθρώπου μηχανής.

3.4.5 Βαθιά Μάθηση (Deep Learning)

Η Βαθιά Μάθηση (Deep Learning) [18] είναι μέρος μιας ευρύτερης κατηγορίας μεθόδων μηχανικής μάθησης, βασιζόμενων σε αναπαραστάσεις δεδομένων μάθησης παρά σε αλγορίθμους συγκεκριμένων εργασιών.

Την τελευταία δεκαετία, προσεγγίσεις που βασίζονται σε αρχιτεκτονικές βαθιάς μάθησης έχουν ξεπεράσει σημαντικά τις τεχνικές ταξινόμησης με πιο απλές/ρηχές αρχιτεκτονικές. Ειδικότερα, οι αρχιτεκτονικές βαθιάς μάθησης αποτέλεσαν μία από τις σημαντικότερες ανακαλύψεις της τελευταίας δεκαετίας στον κλάδο της Τεχνητής Νοημοσύνης αφού πλέον έχει καταστεί δυνατή η εξαγωγή των πιο σύνθετων και εξεζητημένων χαρακτηριστικών, με πολύ μεγάλη αξιοπιστία.

Η βαθιά μηχανική μάθηση είναι η διαδικασία εφαρμογής «μαθησιακών διεργασιών» σε νευρωνικά δίκτυα πολλών επιπέδων, γνωστά και ως **Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks)**. Όπως προκύπτει και από την ίδια την έννοια, ένα βαθύ νευρωνικό δίκτυο διαφέρει από το απλό νευρωνικό λόγω του βάθους. Σαν βάθος ορίζουμε τον αριθμό των κρυφών επιπέδων (hidden layers) και επομένως, ως DNN ένα Νευρωνικό Δίκτυο πολλών κρυφών επιπέδων. Στο σχήμα 3.5, παρατηρούμε ένα βαθύ Νευρωνικό δίκτυο τριών κρυφών επιπέδων.

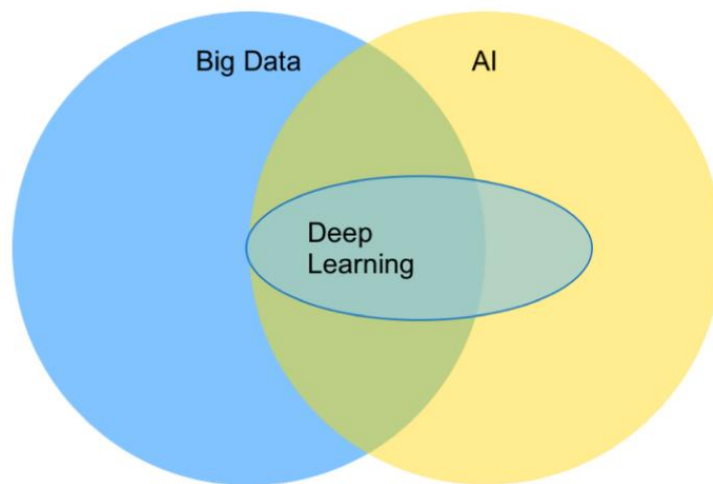


Σχήμα 3.4: Βαθύ Νευρωνικό δίκτυο (Deep Neural Network), με τρία κρυφά επίπεδα (Hidden Layers).

Έχουν προταθεί διάφορες αρχιτεκτονικές βαθιάς μάθησης, στις σημαντικότερες από τις οποίες περιλαμβάνονται τα **Συνελικτικά Νευρωνικά Δίκτυα (CNN)**. Ονομάζονται συνελικτικά διότι εφαρμόζουν τη (γραμμική) μαθηματική διαδικασία της συνέλιξης (convolution), αντί να χρησιμοποιήσουν πολλαπλασιασμό πινάκων, σε τουλάχιστον ένα από τα επίπεδα, που τα απαρτίζουν. Τα ΤΝΔ βαθιάς συνέλιξης έχουν επιφέρει βελτιώσεις κυρίως σε προβλήματα επεξεργασίας εικόνων, βίντεο, ομιλίας και ήχου. Στο κεφάλαιο που ακολουθεί γίνεται μια εκτενή παρουσίαση των ΣΝΔ, αφού αποτελούν αντικείμενο της παρούσας διπλωματικής.

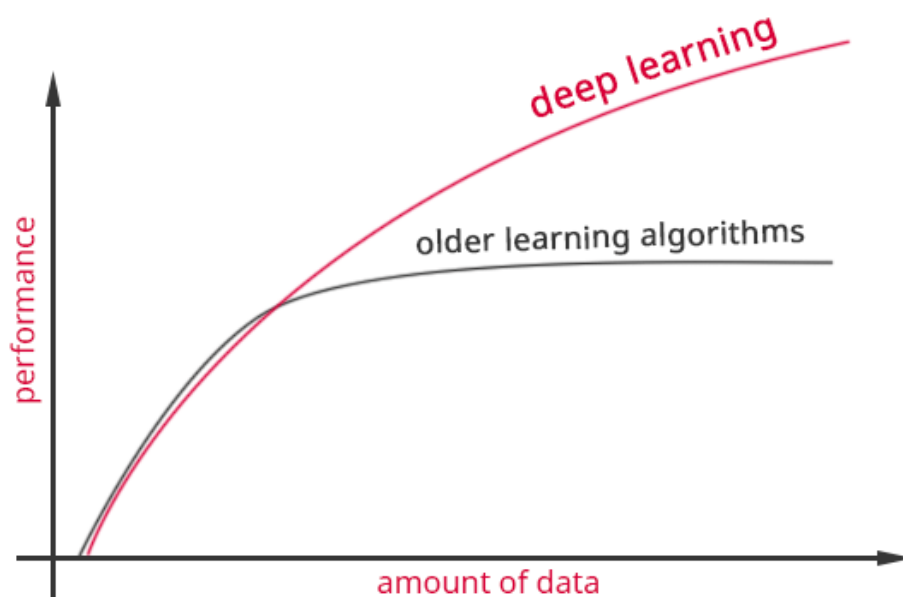
3.4.6 Μεγάλα Δεδομένα (Big Data)

Η μάθηση συναρτάται με δεδομένα εκπαίδευσης. Η βαθιά μάθηση συναρτάται απαραίτητως με πολλά δεδομένα, Μεγάλα δεδομένα (Big Data), όπως σήμερα λέγεται. Το Σχήμα 3.5 δείχνει χαρακτηριστικά τη συσχέτιση: Τεχνητή Νοημοσύνη-Μεγάλα Δεδομένα-Βαθιά Μάθηση (AI-Big Data-Deep Learning).



Σχήμα 3.5: Σχέση εξάρτησης μεταξύ Τεχνητής Νοημοσύνης - Μεγάλα Δεδομένα - Βαθιά Μάθηση (AI-Big Data-Deep Learning).

Η συμπεριφορά των συμβατικών αλγορίθμων εκμάθησης και των σύγχρονων αλγορίθμων (Βαθιά μάθηση) σε συνάρτηση με τον όγκο των δεδομένων εκπαίδευσης δείχνεται στο Σχήμα 3.6. Όπως φαίνεται, στους παλαιότερους αλγορίθμους μάθησης (machine learning) η συμπεριφορά των συστημάτων, από ένα επίπεδο και πλέον, γινόταν ανεξάρτητη του πλήθους των δεδομένων εκπαίδευσης. Αντίθετα, τα συστήματα βαθιάς μάθησης (deep learning) η συμπεριφορά τους συναρτάται με το πλήθος των δεδομένων. Δια τούτο προκύπτει η ανάγκη για χρήση πολλών δεδομένων ή μεγάλων δεδομένων (Big Data) στην διαδικασία της εκπαίδευσης. Έτσι, για την ολοκλήρωση ενός συστήματος βαθιάς μάθησης δεν επαρκούν οι καλοί αλγόριθμοι (βαθιά νευρωνικά δίκτυα), απαραίτητη προϋπόθεση αποτελεί και η ύπαρξη μεγάλων δεδομένων, κυρίως για την εκπαίδευση (training) αλλά και για τον έλεγχο (test) των συστημάτων.



Σχήμα 3.6: Συμπεριφορά συμβατικών αλγορίθμων μάθησης και αλγορίθμων Βαθιάς μάθησης, σε συνάρτηση με τον όγκο των δεδομένων εκπαίδευσης.

4 Συνελικτικά Νευρωνικά Δίκτυα

4.1 Εισαγωγή

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ), (Convolutional Neural Networks-CNN), αποτελούν μια ειδική κατηγορία νευρωνικών δικτύων, με αρχιτεκτονικές πρόσθιας τροφοδότησης (feedforward networks) [19], τα οποία βρίσκουν ευρεία εφαρμογή σε προβλήματα κατηγοριοποίησης και ταξινόμησης εικόνων, ήχων καθώς και στην επεξεργασία φυσικής γλώσσας. Ονομάζονται συνελικτικά διότι εφαρμόζουν τη (γραμμική) μαθηματική διαδικασία της συνέλιξης (convolution), σε τουλάχιστον ένα από τα επίπεδα, που τα απαρτίζουν αντί να χρησιμοποιήσουν πολλαπλασιασμό πινάκων.

Τα Συνελικτικά Νευρωνικά Δίκτυα έχουν αρχιτεκτονική πολλών επιπέδων ειδικά σχεδιασμένη για να αναγνωρίζουν δισδιάστατα σχήματα με υψηλό βαθμό ευαισθησίας. Η αρχιτεκτονική τους σχεδιάζεται έτσι ώστε να εκμεταλλεύεται τη 2D δομή των εικόνων εισόδου ή άλλων δισδιάστατων σημάτων. Ένα πρόσθετο πλεονέκτημα των ΣΝΔ είναι η ευκολότερη εκπαίδευση τους και έχουν πολύ λιγότερες παραμέτρους από τα πλήρως συνδεδεμένα ΝΔ με τον ίδιο αριθμό κρυφών επιπέδων. Σήμερα, τα ΣΝΔ, εντάσσονται στο state-of-the-art των τεχνικών αναγνώρισης προτύπων και ιδιαίτερα στις τεχνικές που αφορούν ταξινόμηση ιατρικών δεδομένων.

Στη συνέχεια του κεφαλαίου, παρουσιάζεται η δομή των Συνελικτικών Νευρωνικών Δικτύων (ΣΝΔ). Γίνεται ανασκόπηση των διαφόρων αρχιτεκτονικών ΣΝΔ που έχουν προταθεί τα τελευταία χρόνια, επιλογές των οποίων θα δοκιμαστούν κατά την πειραματική διαδικασία, στη συνέχεια αυτής της εργασίας.

4.2 Δομή των Συνελικτικών Νευρωνικών Δικτύων

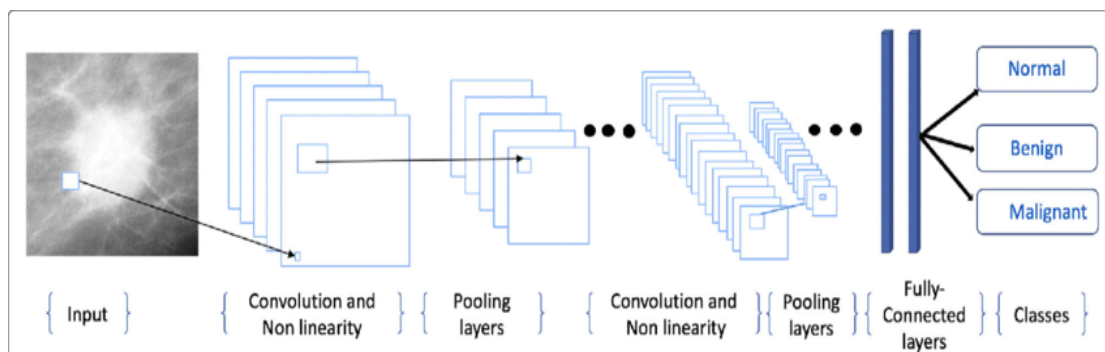
Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν ουσιαστικά μια ακολουθία επιπέδων όπου σε κάθε επίπεδο γίνεται ένας μετασχηματισμός του όγκου των δεδομένων. Για παράδειγμα, δέχονται για είσοδο εικόνες, σε τρισδιάστατες δομές, (όγκος δεδομένων) και ακολουθείται μια σειρά από συγκεκριμένα επίπεδα που μετασχηματίζουν τα δεδομένα αυτά ώστε να γίνουν πιο διαχειρίσιμα.

Τέσσερα είναι τα βασικά επίπεδα, που αποτελούν τα ΣΝΔ (Σχήμα 4.1):

- **Συνελικτικά Επίπεδα** (Convolutional Layers): οι νευρώνες εκτελούν συνέλιξη της εισόδου με φίλτρα που προκύπτουν από την εκμάθηση.

- **Συγκεντρωτικά Επίπεδα (Pooling Layers):** η κύρια λειτουργία των οποίων είναι η ελαχιστοποίηση των δεδομένων.
- **μη-Γραμμικά Επίπεδα (Non Linear Layers):** όπου με είσοδο μη γραμμικών ιδιοτήτων παρέχεται η δυνατότητα στο σύστημα να αποφεύγει χαοτικές συμπεριφορές, λόγω πολλαπλών παραμέτρων.
- **Πλήρως Συνδεδεμένα Επίπεδα (Fully Connected Layers):** τα συναντάμε κατά κανόνα στο τέλος κάθε ΣΝΔ.

Στο Σχήμα 4.1 φαίνονται τα βασικά μέρη που αποτελούν ένα ΣΝΔ. Η αρχιτεκτονική του έχει τη δομή στοίβας (stack) από **Συνελικτικά Επίπεδα (Convolutional layer)**, **μη-Γραμμικά Επίπεδα (Nonlinear layer, π.χ. ReLU)**, **Συγκεντρωτικά Επίπεδα (Pooling layer)** και **Πλήρως Συνδεδεμένα Επίπεδα**. Η έξοδος μπορεί να είναι μια συγκεκριμένη κατηγορία: π.χ. Κανονική (Normal), Καλοήθης (Benign), Κακοήθης (Malignant).



Σχήμα 4.1: Τα βασικά μέρη που αποτελούν τα ΣΝΔ. Στοίβα από τα διάφορα Επίπεδα. Η έξοδος οι κατηγορίες: Κανονική (Normal), Καλοήθης (Benign), Κακοήθης (Malignant).

4.2.1 Συνελικτικά Επίπεδα (Convolutional Layers)

Τα Συνελικτικά Επίπεδα αποτελούν την ουσία των ΣΝΔ. Αυτά είναι που δίνουν τη διαφορετικότητα και τις σπουδαίες ιδιότητες τους. Στο επίπεδο αυτό, με τη βοήθεια φίλτρων, σαρώνεται η εικόνα, και εξάγονται χαρακτηριστικά. Για να παραχθούν τα χαρακτηριστικά σαρώνεται ολόκληρη η εικόνα από φίλτρα και γίνεται η πράξη του εσωτερικού γινομένου μεταξύ τους. Αυτή η διαδικασία αποτελεί ουσιαστικά τη συνέλιξη μεταξύ των δύο σημάτων: του φίλτρου (filter ή kernel) και της εικόνας.

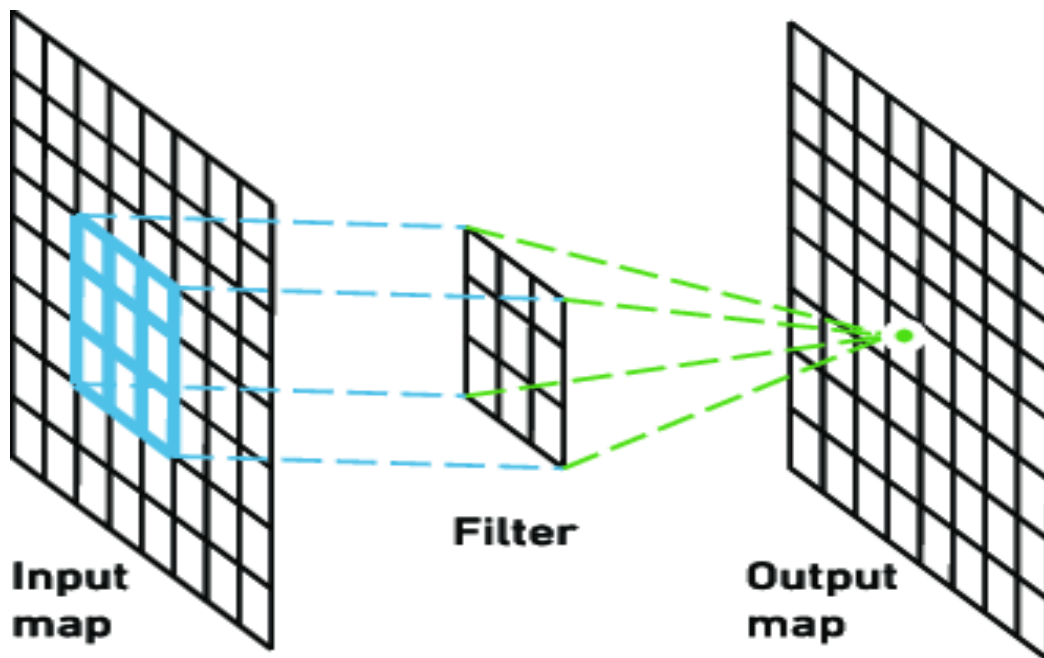
Με τον τρόπο αυτόν, το αποτέλεσμα του εσωτερικού γινομένου τοποθετείται σε κάποια θέση στο διδιάστατο πίνακα *χαρακτηριστικών* (feature map). Όταν γίνεται η πρόσθια διάδοση, όλα τα φίλτρα σαρώνουν την εικόνα σε όλο το πλάτος και το ύψος της εισόδου και υπολογίζουν το γινόμενο μεταξύ των αντίστοιχων θέσεων του φίλτρου και της εισόδου σε κάθε θέση. Έτσι το δίκτυο αντιλαμβάνεται σε πιο σημείο της εικόνας έχει ενεργοποιηθεί το αντίστοιχο φίλτρο, δηλαδή αντιλαμβάνεται το πότε θα ενεργοποιηθεί όταν έρχεται σε επαφή με ακμές ή γωνίες ή σε επίπεδα υφής (texture). Επιπλέον, σε κάθε επίπεδο, υπάρχουν πολλές ομάδες φίλτρων διαφορετικών μεγεθών, για να γίνεται δυνατή η επεξεργασία εισόδων διαφορετικών μοτίβων. Κάθε ομάδα φίλτρων παράγει έναν ξεχωριστό χάρτη ενεργοποίησης 2-διαστάσεων, τον χάρτη χαρακτηριστικών (feature map), όπου όταν στοιβάζονται αυτοί οι χάρτες ενεργοποίησης παράγουν την έξοδο. Ακόμα στο συνελικτικό επίπεδο για τη σωστή λειτουργία των φίλτρων χρησιμοποιούνται δύο **υπερπαραμέτροι: Stride (Αλμα)** και **Padding (Γέμισμα)**:

- **Stride** (Άλμα): στην ουσία έχει να κάνει με το πόσο ολισθαίνουν πάνω στην εικόνα τα φίλτρα. Ο βασικός λόγος που υπάρχει το stride είναι για να μειώνει τις διαστάσεις της εικόνας που βγαίνει από το συνελικτικό φίλτρο.
- **Padding** (γέμισμα περιθωρίου): χρησιμοποιείται για να ελέγχει το μέγεθος της εξόδου που βγαίνει από το συνελικτικό επίπεδο. Είναι ένα πλήθος από στοιχεία τα οποία τοποθετούνται στα άκρα της εικόνας, είτε κατά πλάτος είτε κατά ύψος.

Η σχέση που χρησιμοποιείται για να ορίσουμε τις διαστάσεις των χαρτών ενεργοποίησης είναι:

$$\frac{N + 2 * P - F}{S + 1}$$

όπου, N =διάσταση της εικόνας εισόδου, P =γέμισμα περιθωρίου με μηδενικά (Padding), F =διαστάσεις του φίλτρου και S = Άλμα (Stride).



Σχήμα 4.2: Στο επίπεδο συνέλιξης (Convolutional Layer).

4.2.2 Συγκεντρωτικά Επίπεδα (Pooling layers)

Η κύρια λειτουργία του Συγκεντρωτικού επιπέδου είναι η υπο-δειγματοληψία (Sub sampling) των δεδομένων και επομένως ο περιορισμός του αριθμού των απαιτούμενων πράξεων (Σχήμα 4.3). Στην ουσία μπορούμε να φανταστούμε το επίπεδο αυτό ως ένα πλέγμα νευρώνων όπου αντιστοιχεί μία ομάδα από εικονοστοιχεία σε μία μοναδική τιμή. Οι πιο κοινές κατηγορίες του επιπέδου pooling είναι:

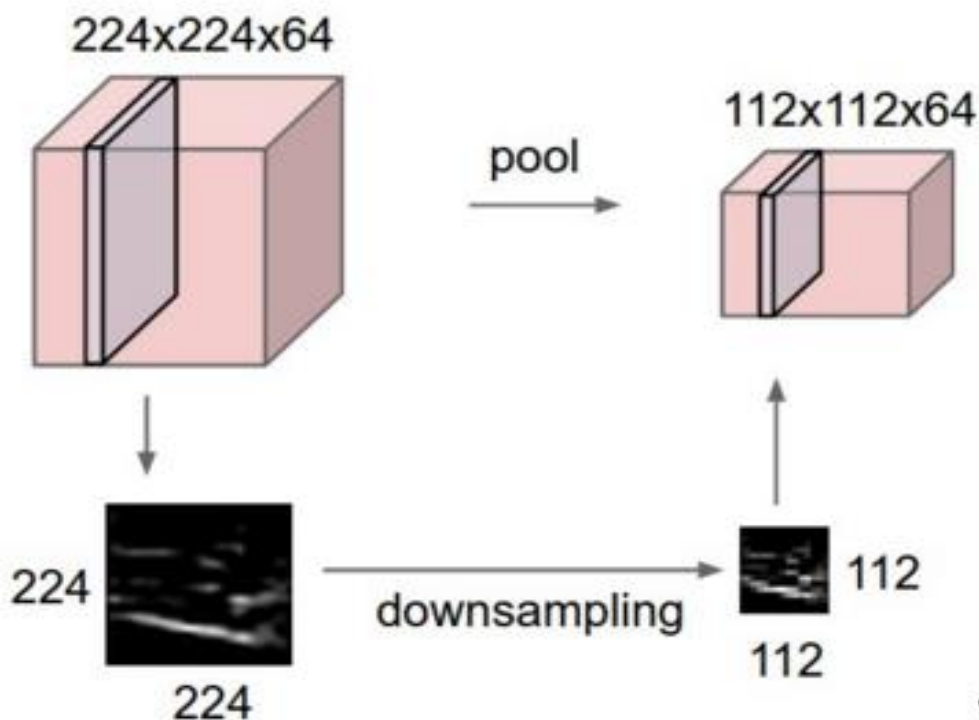
- **Maximum pooling** (Μεγίστου): Στην τεχνική αυτή σε κάθε περιοχή της εικόνας διατηρείται μόνο η τιμή του φίλτρου που είναι η μέγιστη. Επιπλέον σε κάθε περιοχή υπάρχουν και αδιάφορα χαρακτηριστικά τα οποία αγνοούνται, ώστε να δοθεί έμφαση μόνο στα χαρακτηριστικά που ενδιαφέρουν το κάθε φίλτρο. Η max pooling, βοηθάει και στην ταχύρρυθμη εκμάθηση του δικτύου αφού όλη η διαδικασία σαρώνει μόνο τους κόμβους που έχουν εντοπίσει το συγκεκριμένο χαρακτηριστικό στην εικόνα.

- **Average pooling** (Μέσου Όρου): Στην συγκεκριμένη τεχνική σε κάθε περιοχή της εικόνας διατηρείται μόνο ο μέσος όρος από την τιμή του φίλτρου. Όταν τα στοιχεία που θέλουμε να υπολογίσουμε υπάρχουν σε όλο το εύρος τη εικόνας τότε η δειγματοληψία μέσου όρου βρίσκει μεγάλη εφαρμογή. Η τεχνική αυτή, κυρίως, χρησιμοποιείται σε τεχνικές παλινδρόμησης (regression) και όχι κατηγοριοποίησης (classification) αφού το δίκτυο δεν διαλέγει κάποια κλάση ως έξοδο αλλά μέσω μιας συνάρτησης κόστους υπολογίζεται η απόσταση από τη πραγματική.
- **Median pooling** (Μεσαία τιμή): Η τεχνική αυτή διαφέρει από την αντίστοιχη του μέσου όρου αφού αντί να υπολογίζει το μέσο όρο επιλέγεται η μεσαία τιμή. Το θετικό χαρακτηριστικό της συγκεκριμένης τεχνικής είναι ότι δεν επηρεάζεται από ακραίες αποκρίσεις του φίλτρου.

Η σχέση η οποία χρησιμοποιείται για την έξοδο, μετά την εφαρμογή της Μέγιστης Συγκέντρωσης (Max Pooling) είναι η εξής:

$$(N - F) / S + 1$$

όπου, N =διάσταση της εικόνας εισόδου του επιπέδου συγκέντρωσης, F =διαστάσεις του φίλτρου και S =Αλμα (Stride).



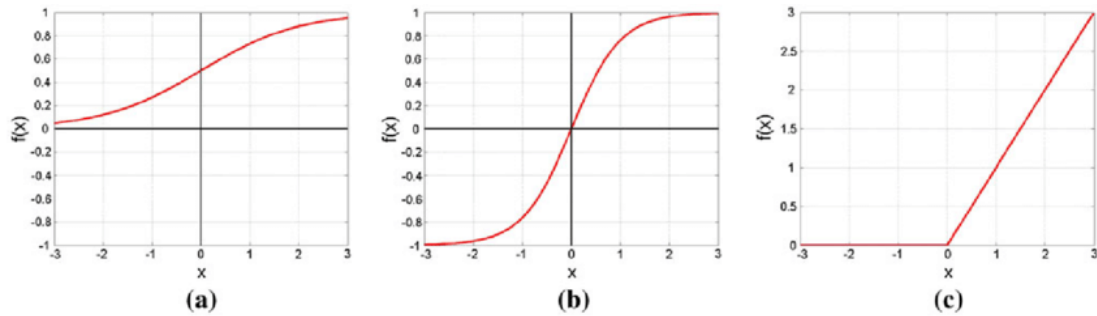
Σχήμα 4.3: Επίπεδο υπο-δειγματοληψίας (Pooling layers).

4.2.3 Μη-Γραμμικά Επίπεδα (Non Linearity)

Το μη-Γραμμικό επίπεδο (Non Linearity Layer) βοηθάει ώστε να αποφευχθεί μια χαοτική συμπεριφορά, λόγω των πολύ μεγάλων τιμών και των πολλαπλών παραμέτρων. Με την είσοδο μη γραμμικών ιδιοτήτων στο σύστημα αποφεύγεται τέτοιου είδους δυσλειτουργία, καθώς φράζεται η κάθε τιμή. Στην ουσία το μη-Γραμμικό επίπεδο αποτελεί έναν κόμβο που

τοποθετείται στο τέλος ή ανάμεσα στα επίπεδα των νευρωνικών δικτύων και βοηθάει στο να πυροδοτεί έναν νευρώνα ή όχι.

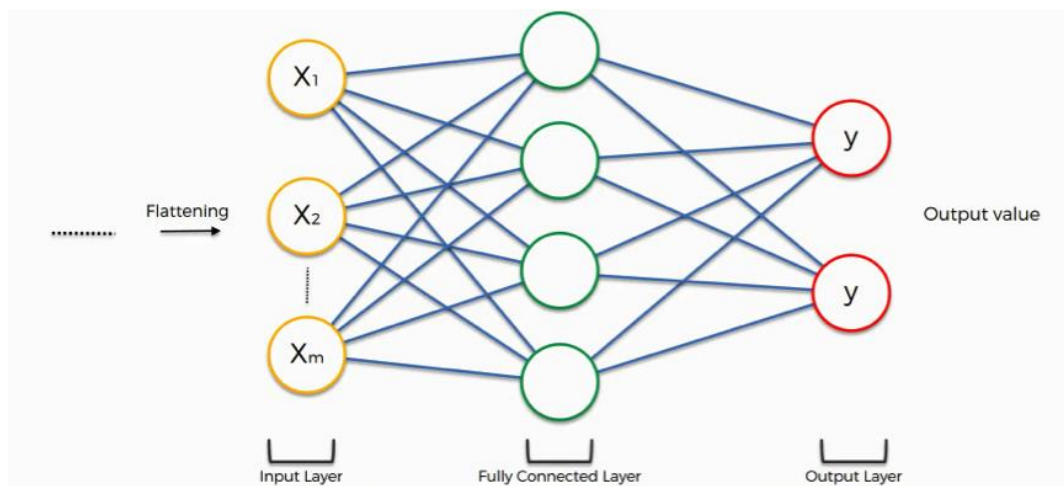
Μερικά παραδείγματα μη γραμμικότητας (Σχήμα 4.4) αποτελούν οι συναρτήσεις: *sigmoid*, *tan(h)* και η *ReLU* (Rectified Linear Unit).



Σχήμα 4.4: Επίπεδα μη-Γραμμικότητας (Non Linearity). (a) Sigmoid, (b) tanh, (c) **ReLU** (Rectified Linear Unit).

4.2.4 Πλήρως Συνδεδεμένα Επίπεδα (Fully Connected Layers)

Στα πλήρως συνδεδεμένα επίπεδα, κάθε νευρώνας των κρυφών επιπέδων συνδέεται με την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου. Είναι μια κλασική συνδεσμολογία πολυεπίδεδων νευρώνων (Multilayer Perceptron) η οποία χρησιμοποιεί μια συνάρτηση απώλειας (loss function) για ενεργοποίηση. Ο κύριος στόχος της είναι να χρησιμοποιεί τα χαρακτηριστικά για να κάνει κατηγοριοποίηση του συνόλου των εικόνων εισόδου στις διάφορες κατηγορίες που θα υπάρχουν, με βάση το πρότυπο εκπαίδευσης που χρησιμοποιήθηκε.



Σχήμα 4.5: Πλήρως Συνδεδεμένα Επίπεδα (Fully Connected Layers).

4.3 Τεχνικές Προσαρμογής των Δεδομένων στο Σύστημα

Από τα πιο κεντρικά προβλήματα που προκύπτουν στον τομέα της μηχανικής μάθησης, αποτελεί το πώς να κάνουμε τον αλγόριθμό μας να βρει μια τέλεια εφαρμογή όχι μόνο στο σετ εκπαίδευσης αλλά και στο σύνολο των δοκιμών. Όταν ένας αλγόριθμος έχει καλή απόδοση στο σετ εκπαίδευσης αλλά δεν έχει καλή απόδοση στο σετ των δοκιμών, ο αλγόριθμος λέγεται ότι είναι *overfitted* (υπερεκπαιδευμένος) στα δεδομένα εκπαίδευσης. Στη συνέχεια αναφέρονται τεχνικές αντιμετώπισης σχετικών προβλημάτων.

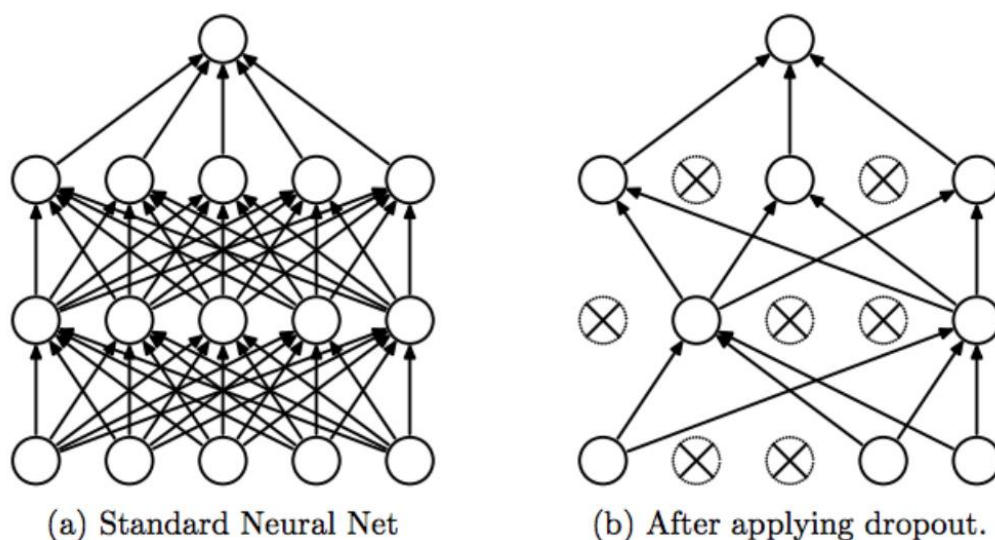
4.3.1 Κανονικοποίηση (Regularization)

Η μέθοδος της κανονικοποίησης (Regularization), είναι μια διαδικασία που βοηθάει στην αντιμετώπιση του προβλήματος *overfitting* (υπερεκπαδευσης). Η τεχνική αυτή λειτουργεί με την επιβολή «ποινών» στα βάρη του δικτύου μέσω της συνάρτησης κόστους, ώστε να αρχίσουν να φθίνουν γραμμικά.

Η πιο γνωστή των μεθόδων Regularization είναι η **L_2 -regularization**, η οποία ονομάστηκε έτσι επειδή για κάθε βάρος του δικτύου υπολογίζουμε την L_2 νόρμα του εκάστοτε παραθύρου, που είναι το άθροισμα των τετραγώνων όλων των στοιχείων του πίνακα. Στη συνέχεια, το αποτέλεσμα το πολλαπλασιάζουμε με μια σταθερά και σε κάθε οπισθοδρόμηση προσθέτουμε αυτή την τιμή στη συνάρτηση κόστους. Επιπλέον, λόγω των συγκεκριμένων αλληλεπιδράσεων που υπάρχουν μεταξύ των βαρών και των συνάψεων κάνουν το δίκτυο να χρησιμοποιεί όλες τις εισόδους από λίγο.

4.3.2 Εγκατάλειψη (Dropout)

Η διαδικασία Dropout έχει ως σκοπό λειτουργίας να απενεργοποιεί ομάδες νευρώνων (κρυμμένων ή ορατών) μαζί με τις αντίστοιχες συνάψεις τους, κατά τη φάση εκπαίδευσης, με τυχαίο τρόπο και να μην τους υπολογίζει στη φάση της πρόσθιας τροφοδότησης. Αποτελεί ένα είδος μεθόδου κανονικοποίησης αφού πρακτικά αποτρέπει στο να σχηματίζονται αλληλεξαρτήσεις μεταξύ των νευρώνων. Είναι ένας ευρέως διαδεδομένος τρόπος για την αντιμετώπιση του προβλήματος του Overfitting [20].



Σχήμα 4.5: Πριν και μετά την εφαρμογή Dropout (εγκατάλειψη) σε Νευρωνικό Δίκτυο.

Στο πλήρως συνδεδεμένο επίπεδο, ο αριθμός των παραμέτρων είναι πολύ μεγάλος και σε συνδυασμό με τις πολλές επαναλήψεις, οι νευρώνες αναπτύσσουν σχέσεις αλληλεξάρτησης μεταξύ τους και οδηγούνται σε overfitting των δεδομένων εκπαίδευσης. Πιο τεχνικά, σε κάθε στάδιο εκπαίδευσης, μεμονωμένοι κόμβοι είτε εγκαταλείπονται από το δίκτυο (με πιθανότητα $1-p$) είτε διατηρούνται (με πιθανότητα p), έτσι ώστε να μείνει ένα μειωμένο δίκτυο. Οι εισερχόμενες και εξερχόμενες άκρες σε έναν αποκλεισμένο κόμβο αφαιρούνται επίσης.

4.3.3 Επαύξηση των Δεδομένων (Data Augmentation)

Στις εφαρμογές Βαθιάς Μάθησης, απαιτούνται πολλά δεδομένα. Τα δεδομένα αυτά θα πρέπει να έχουν καλή ποικιλομορφία, καθώς το αντικείμενο ενδιαφέροντος πρέπει να υπάρχει σε διάφορα μεγέθη, συνθήκες φωτισμού και πόζες, αν θέλουμε το δίκτυο μας να γενικεύσει καλά κατά τη φάση εκπαίδευσης. Για να ξεπεράσουμε το πρόβλημα περιορισμένης ποσότητας και περιορισμένης ποικιλίας δεδομένων, παράγουμε (κατασκευάζουμε) τα δικά μας δεδομένα από τα δεδομένα που έχουμε. Αυτή η μεθοδολογία δημιουργίας δεδομένων είναι γνωστή ως **Επαύξηση των Δεδομένων** (Data augmentation).

Κατά τη διαδικασία αυτή χρησιμοποιούνται μετασχηματισμοί που εφαρμόζονται στα δεδομένα ώστε να παραχθούν από αυτά τα νέα δεδομένα. Αυτό μας δίνει τη δυνατότητα να επεκτείνουμε το μέγεθος του dataset αφού εκτός των αρχικών δεδομένων θα υπάρχουν και τα νέα μετασχηματισμένα δεδομένα.

Ακόμα αν υπάρχουν αρκετές μεμονωμένες περιπτώσεις σε κάθε κλάση μπορούμε να χρησιμοποιήσουμε μετασχηματισμούς που δεν επηρεάζουν την πιστότητα του μοντέλου και δημιουργούν νέες περιπτώσεις από τις αρχικές.

4.3.4 Batch Normalization

Το πρόβλημα που επιδιώκει να αντιμετωπίσει η τεχνική κανονικοποίησης Batch-Normalization [21] είναι η αλλαγή της κατανομής των ενεργοποιήσεων σε ένα νευρωνικό δίκτυο, που προκύπτει κάθε φορά που μεταβάλλονται οι παράμετροι ενός μοντέλου κατά την εκπαίδευση. Έχει φανεί πως αυτή η μεταβολή ενδέχεται να συγκρατεί τις τιμές των ενεργοποιήσεων στην περιοχή κορεσμού με αποτέλεσμα οι κλίσεις να είναι μικρές, με αποτέλεσμα η εκπαίδευση να

καθυστερεί σημαντικά. Για να περιοριστεί το παραπάνω πρόβλημα επιδιώκεται να σταθεροποιηθεί η κατανομή των εισόδων x κάθε στρώματος.

Στη διαδικασία της κανονικοποίησης των εισόδων, μετασχηματίζουμε τις τιμές των δεδομένων ώστε να εκτείνονται σε ένα συγκεκριμένο όριο (συνήθως από -1 έως 1). Επιπλέον, η συγκεκριμένη μέθοδος κανονικοποιεί τα δεδομένα ακόμα και στα κρυφά επίπεδα ενώ σε αυτά οι τιμές τους μεταβάλλονται συνέχεια. Η συγκεκριμένη τεχνική αρχικά συμβάλει στο να μειώνεται σημαντικά ο χρόνος για την εκπαίδευση του δικτύου. Ακόμα, η μέθοδος αυτή συντελεί στη μείωση της διασποράς των παραμέτρων των κρυφών στρωμάτων με αποτέλεσμα τα στρώματα του δικτύου να μπορούν να εκπαιδεύονται ανεξάρτητα το ένα από το άλλο. Τέλος, η διαδικασία βοηθάει στο να αποφεύγουμε περιπτώσεις ακραίων τιμών, δηλαδή ασταθείς περιπτώσεις.

4.3.5 Principal Components Analysis (PCA)

Σκοπός της Ανάλυσης Πρωτευουσών Συνιστωσών (PCA) είναι να βρει από ένα πλήθος p μεταβλητών ένα νέο πλήθος λιγότερων μεταβλητών. Οι νέες μεταβλητές έχουν την ιδιότητα να είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών και παράλληλα να μη συσχετίζονται μεταξύ τους, [47]. Το κύριο χαρακτηριστικό του PCA είναι ότι μειώνει τις διαστάσεις πολυδιάστατων συνόλων δεδομένων για περαιτέρω ανάλυση.

Η Ανάλυση Πρωτευουσών Συνιστωσών είναι ένας τρόπος αναγνώρισης προτύπων στα δεδομένα και εμφάνισης των δεδομένων αυτών με τρόπο που να δίνεται έμφαση στις ομοιότητες και τις διαφορές τους. Δεδομένου ότι τα πρότυπα αναζήτησης είναι δύσκολο να εντοπιστούν, για μεγάλο όγκο δεδομένων, η τεχνική PCA είναι ένα δυνατό εργαλείο για την προεπεξεργασία τους.

Από τεχνικής άποψης η PCA είναι ένας ορθογώνιος γραμμικός μετασχηματισμός που μετασχηματίζει τα δεδομένα σε ένα νέο σύστημα συντεταγμένων έτσι ώστε η μέγιστη διαφορά από οποιαδήποτε προβολή των στοιχείων έρχεται να βρεθεί στην πρώτη συντεταγμένη (πρώτη πρωτεύουσα συνιστώσα), η δεύτερη μέγιστη διαφορά στη δεύτερη συντεταγμένη, και ούτω καθ' εξής. Τα βήματα εφαρμογής της τεχνικής PCA μπορούν να συνοψιστούν στα ακόλουθα: Αρχικά αφαιρείται η μέση τιμή κάθε διάστασης από κάθε στοιχείο της αντίστοιχης διάστασης του δείγματος. Έτσι παίρνουμε ένα νέο σύνολο δεδομένων που η μέση τιμή του είναι μηδέν. Στη συνέχεια υπολογίζουμε τον πίνακα συνδιασποράς και αφού ο πίνακας συνδιασποράς είναι τετραγωνικός μπορούμε να υπολογίσουμε τα ιδιοδιανύσματα και τις ιδιοτιμές του. Από τα ιδιοδιανύσματα μπορούμε να εξάγουμε γραμμές οι οποίες χαρακτηρίζουν τα δεδομένα μας. Τώρα, παρατηρώντας τα ιδιοδιανύσματα και τις ιδιοτιμές βλέπουμε ότι οι ιδιοτιμές έχουν διαφορετικές τιμές μεταξύ τους. Τα ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές είναι οι πρωτεύουσες συνιστώσες του συνόλου δεδομένων μας. Αυτό που κάνουμε είναι να διατάξουμε τα ιδιοδιανύσματα σύμφωνα με τις ιδιοτιμές τους από τη μεγαλύτερη στη μικρότερη. Έτσι παίρνουμε τις συνιστώσες με τη σειρά σημαντικότητας τους.

4.3.6 Zero Components Analysis (ZCA)

Ένας άλλος τρόπος μείωσης των διαστάσεων είναι ο Zero Components Analysis (ZCA) [48] και έχει χαρακτηριστεί ως ο ξάδελφος της PCA. Λευκαίνει (whitening) τα δεδομένα με την έννοια ότι τα χαρακτηριστικά θα πρέπει να έχουν μοναδιαία συσχέτιση με τα ίδια και μηδενική συσχέτιση με κάθε άλλο, διαιρώντας τον μετασχηματισμό PCA με τις αντίστροφες μοναδικές

τιμές. Γι' αυτό εφαρμόζεται με την προσδοκία ότι θα μπορούσε να βελτιώσει τα αποτελέσματά μας, αλλά ταυτόχρονα ήμασταν προετοιμασμένοι και για χειρότερα αποτελέσματα.

Οι μικροσκοπικές εικόνες εμφανίζουν ισχυρούς συσχετισμούς μεταξύ κοντινών εικονοστοιχείων. Συγκεκριμένα, αμφίδρομοι συσχετισμοί είναι αρκετά ισχυροί. Όταν μαθαίνεται ένα στατιστικό μοντέλο εικόνων, μπορεί να είναι ωραίο να αναγκάσει το μοντέλο να επικεντρωθεί σε συσχετίσεις υψηλότερης τάξης παρά να αποσπάται η προσοχή από τη μοντελοποίηση αμφίδρομων συσχετισμών. Η υπόθεση είναι ότι αυτό μπορεί να κάνει το μοντέλο πιο πιθανό να ανακαλύψει ενδιαφέρον κανονικότητες στις εικόνες και όχι απλώς να μάθουν ότι τα κοντινά εικονοστοιχεία είναι παρόμοια.

Ένας τρόπος για να αναγκάσετε το μοντέλο να αγνοήσει τη δομή δεύτερης τάξης είναι να το αφαιρέσετε. Ευτυχώς αυτό μπορεί να γίνει με ένα βήμα προεπεξεργασίας δεδομένων που αποτελείται απλώς από τον πολλαπλασιασμό της μήτρας δεδομένων με έναν πίνακα λεύκανσης (whitening matrix).

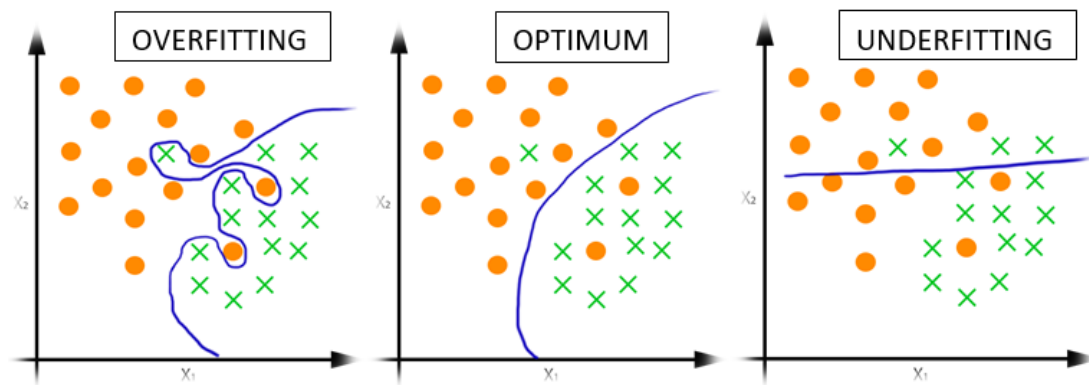
4.3.7 Overfitting, Underfitting

Η **υπερεκπαίδευση** (Overfitting), είναι ένα από πιο συχνά προβλήματα που αντιμετωπίζουμε στη διαδικασία εκπαίδευσης ενός συστήματος. Είναι το φαινόμενο κατά το οποίο ένα ΤΝΔ χειρίζεται με μεγάλη ακρίβεια όλα τα παραδείγματα του συνόλου εκπαίδευσης χωρίς να παρουσιάζει καλές επιδόσεις στα άγνωστα. Δηλαδή, το συνολικό σφάλμα για κάποια παραδείγματα εκπαίδευσης γίνεται πάρα πολύ μικρό, αλλά γίνεται υπερβολικά μεγάλο σε άλλα παραδείγματα. Αυτό σημαίνει ότι το νευρωνικό δίκτυο δεν *γενικεύει* (generalizes) καλά. Όταν το δίκτυο, μέσω της αναπροσαρμογής των βαρών, εκπαιδεύεται σε συγκεκριμένα δεδομένα μαθαίνει να αναγνωρίζει χαρακτηριστικά και πρότυπα. Όταν όμως παρουσιαστεί πρόβλημα το σύστημα, συγκεκριμενοποιεί τα χαρακτηριστικά στα οποία εκπαιδεύτηκε και δεν μπορεί να γενικεύσει όταν έρχεται σε επαφή με καινούργια δεδομένα. Οι λόγοι που δημιουργούν το πρόβλημα του overfitting μπορούν να συνοψιστούν στους ακόλουθους:

- Περιορισμένος αριθμός δεδομένων.
- Ανισοκατανομή των δεδομένων. Δεν είναι ισο-μοιρασμένα στις διάφορες κατηγορίες ταξινόμησης.
- Οι εποχές που εκπαιδεύουμε το δίκτυο είναι υπερβολικά πολλές.
- Η αρχιτεκτονική του νευρωνικού είναι υπερβολικά σύνθετη.

Το **Underfitting** είναι το ακριβώς αντίθετο του overfitting. Εδώ το σύστημα δυσκολεύεται να μάθει να αναγνωρίζει χαρακτηριστικά, τόσο από τα δεδομένα εκπαίδευσης (training data), όσο και από τα δεδομένα ελέγχου (testing data). Συνήθως προκαλείται από κακή σχεδίαση του προβλήματος της κατηγοριοποίησης.

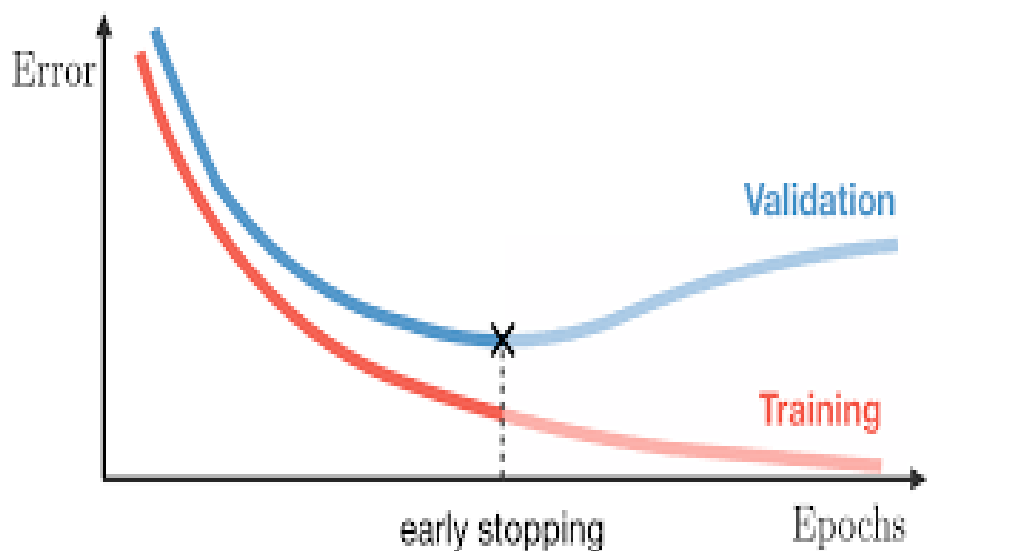
Στο Σχήμα 4.6, παρουσιάζονται σχηματικά οι τρεις περιπτώσεις εκπαίδευσης: η υπερεκπαίδευση (Overfitting), η ιδανική εκπαίδευση (Optimum) και η υποεκπαίδευση (Underfitting).



Σχήμα 4.6: Υπερεκπαίδευση (overfitting), Ιδανική εκπαίδευση (optimum) και Υποεκπαίδευση (Underfitting).

4.3.8 Πρόωρη Διακοπή (Early stopping)

Η πρόωρη διακοπή είναι μια τεχνική που εφαρμόζεται για να αποφευχθεί η υπερφόρτωση του δικτύου κατά τη διαδικασία της εκπαίδευσης. Αυτό γίνεται με την παρακολούθηση των παραμέτρων του δικτύου, (επικύρωση ακρίβειας απώλειας, ακρίβειας ανάκλησης), ώστε να ελέγχεται αν σε κάθε εποχή οι τιμές υπερβαίνουν τα όρια που έχουν τεθεί. Το σημαντικό είναι να βρεθεί το σημείο στο οποίο αρχίζει να επιδεινώνονται οι παράμετροι του δικτύου και άρα η μάθηση του δικτύου αρχίζει χειροτερεύει κατά το πέρασμα των εποχών. Η τεχνική πρόωρης διακοπής δεν έχει ως αποτέλεσμα την κατασκευή καλύτερου νευρωνικού δικτύου, αλλά την ανίχνευση προβλημάτων γενίκευσης κατά την εκπαίδευση με οποιαδήποτε μέθοδο, άρα μπορεί να αποτελέσει ένδειξη για την επιλογή μιας άλλης μεθόδου.



Σχήμα 4.7: Πρόωρη Διακοπή (Early stopping).

4.3.9 Μεταφερόμενη Μάθηση και Συντονισμός

Τα διαθέσιμα δεδομένα που απαιτούνται, για την ολοκλήρωση ενός ΣΝΔ διαιρούνται σε δεδομένα εκπαίδευσης (training data) και δεδομένα ελέγχου (testing data). Ένας συνήθης ποσοστιαίος διαχωρισμός στο πλήθος των δύο κατηγοριών είναι 50-50. Η εκπαίδευση γίνεται μόνο με τα δεδομένα εκπαίδευσης και ο έλεγχος με τα δεδομένα ελέγχου.

Ωστόσο, για την εκπαίδευση ενός ολόκληρου Συνελικτικού Νευρωνικού Δικτύου μεγάλου βάθους από την αρχή με τυχαίες αρχικοποιήσεις των βαρών, σχεδόν ποτέ δεν διατίθεται ένα σύνολο δεδομένων ικανοποιητικού μεγέθους που απαιτεί το τελικό νευρωνικό δίκτυο. Η απάντηση στο πρόβλημα αυτό είναι η μεταφερόμενη μάθηση. Η μεταφορά μάθησης μας δίνει την δυνατότητα να μεταφέρουμε τη γνώση που έχουμε δημιουργήσει από μια ή περισσότερες εφαρμογές και να τη χρησιμοποιήσουμε για να βελτιώσουμε τη μάθηση σε άλλη παρόμοια εφαρμογή.

Σήμερα, τα περισσότερα προ-εκπαιδευμένα ΣΝΔ είναι αρχικά εκπαιδευμένα στο τεράστιο σύνολο δεδομένων του ImageNet. Ένα σύνολο δεδομένων που περιέχει πάνω από 1 εκατομμύριο εικόνες, κατηγοριοποιημένες σε 1.000 κατηγορίες.

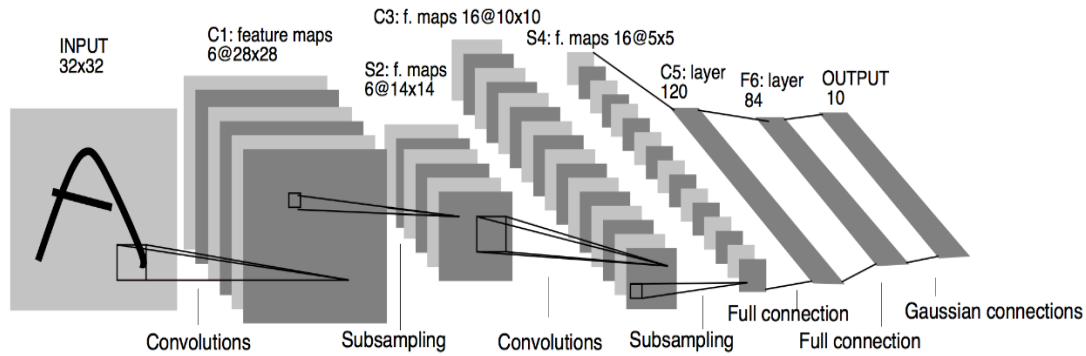
Για να επιτύχει η τεχνική της μεταφερόμενης μάθησης πρέπει να ακολουθείται από τη στρατηγική του συντονισμού. Η στρατηγική αυτή κάνει τροποποιήσεις στα βάρη ενός ήδη εκπαιδευμένου Συνελικτικού Νευρωνικού Δικτύου, συνεχίζοντας τον αλγόριθμο της οπισθοδιάδοσης. Ακόμα, μπορούμε να αλλάξουμε τα βάρη σε όλα τα επίπεδα της αρχιτεκτονικής ή μόνο σε κάποιο από τα τελευταία επίπεδα του δικτύου, κρατώντας σταθερά κάποια από τα αρχικά στρώματα.

4.4 Αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων

Η επιλογή μιας αρχιτεκτονικής είναι πολύ περίπλοκη διαδικασία και αυτό είναι περισσότερο μηχανική από μια επιστημονική προσέγγιση. Είναι επομένως σημαντικό να μελετηθούν οι αρχιτεκτονικές που έχουν αποδειχθεί αποτελεσματικές και να αντληθεί έμπνευση από αυτά τα σημαντικά παραδείγματα [22].

4.4.1 LeNet (1990)

Το μοντέλο LeNet-5, αποτελεί την πρώτη επιτυχημένη εφαρμογή συνελικτικών νευρωνικών δικτύων που αναπτύχθηκε τη δεκαετία του 1990 [23]. Αποτελείται από μόνο λίγα επίπεδα και λίγα φίλτρα, λόγω των περιορισμών της υπολογιστικής δύναμης εκείνη την περίοδο. Η συγκεκριμένη αρχιτεκτονική χρησιμοποιήθηκε κυρίως για αναγνώριση κωδικών, ψηφίων κ.α. Στο σχήμα 4.8 φαίνεται το μοντέλο LeNet-5, στην εφαρμογή αναγνώριση χαρακτήρων.

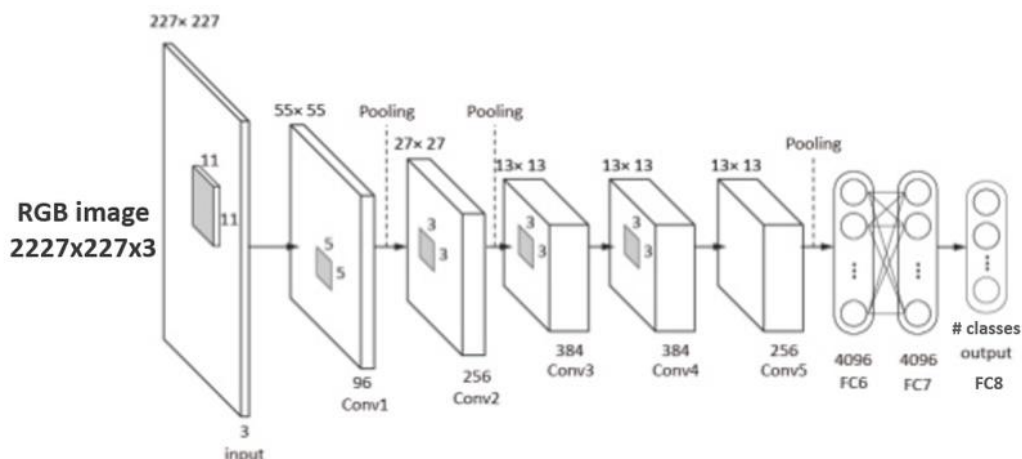


Σχήμα 4.8: Το συνελκτικό νευρωνικό δίκτυο μοντέλο LeNet-5, στην αναγνώριση χαρακτήρων.

4.4.2 AlexNet (2012)

Το συνελκτικό νευρωνικό δίκτυο AlexNet αναπτύχθηκε το 2012 [24] και κέρδισε την πρώτη θέση στο διαγωνισμό ImageNet (LSVRC2012), που έλαβε μέρος. Ήταν το πρώτο το οποίο έκανε τα συνελκτικά δίκτυα διάσημα στο χώρο της υπολογιστικής όρασης (computer vision). Το δίκτυο αυτό είχε μία πολύ παρόμοια αρχιτεκτονική με αυτή του LeNet, ωστόσο ήταν βαθύτερο (περισσότερα κρυφά επίπεδα), μεγαλύτερο και είχε πολλά συνελκτικά επίπεδα. Συγκεκριμένα, αποτελείται από οκτώ (8) επίπεδα, τα 5 συνελκτικά, ακολουθούμενα από επίπεδα συγκέντρωσης αλλά και ενεργοποίησης και στο τέλος 3 πλήρως συνδεδεμένα επίπεδα.

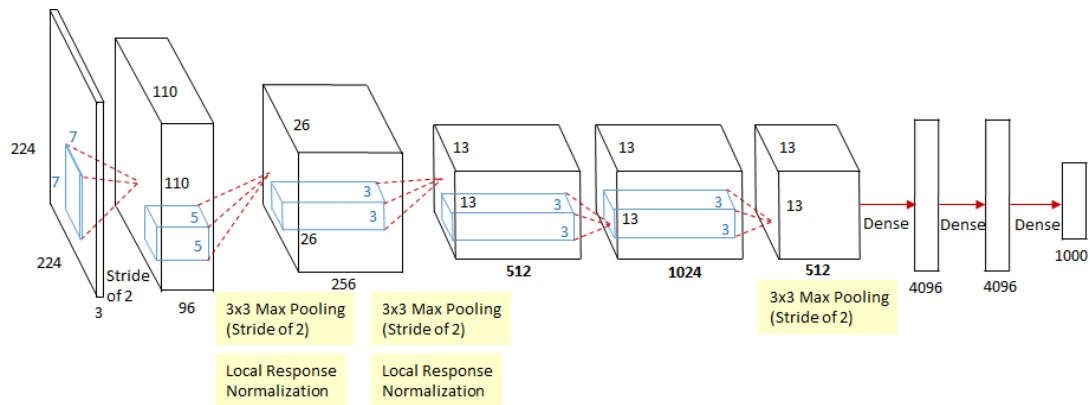
Στο Σχήμα 4.9, φαίνεται ένα παράδειγμα δικτύου με είσοδο RGB εικόνα ($227 \times 227 \times 3$). Για υποδειγματοληψία χρησιμοποιεί την τεχνική maxpooling, με βήμα μετατόπισης 2. Σε όλα τα επίπεδα χρησιμοποιεί συνάρτηση ενεργοποίησης ReLu, εκτός από το τελευταίο που χρησιμοποιεί softmax.



Σχήμα 4.9: Αρχιτεκτονική του Συνελκτικού Νευρωνικού Δικτύου AlexNet. Περιλαμβάνει 5 συνελκτικά στρώματα και 3 πλήρως συνδεδεμένα στρώματα.

4.4.3 ZFNet (2013)

Το Συνελκτικό Νευρωνικό Δίκτυο ZFNet [25] συμμετείχε και νίκησε στο διαγωνισμό ILSVRC (ImageNet Large Scale Visual Recognition Competition) 2013. Αυτό το μοντέλο αποτελεί ουσιαστικά μια βελτίωση του AlexNet, με αλλαγές στις παραμέτρους και στην αρχιτεκτονική. Γενικότερα, επέκτειναν το μέγεθος των μεσαίων συνελκτικών επιπέδων και ρύθμισαν το άλμα (stride) και το μέγεθος του φίλτρου του πρώτου επιπέδου να είναι μικρότερο. Η αρχιτεκτονική του μοντέλου ZFNet φαίνεται στο σχήμα 4.10.

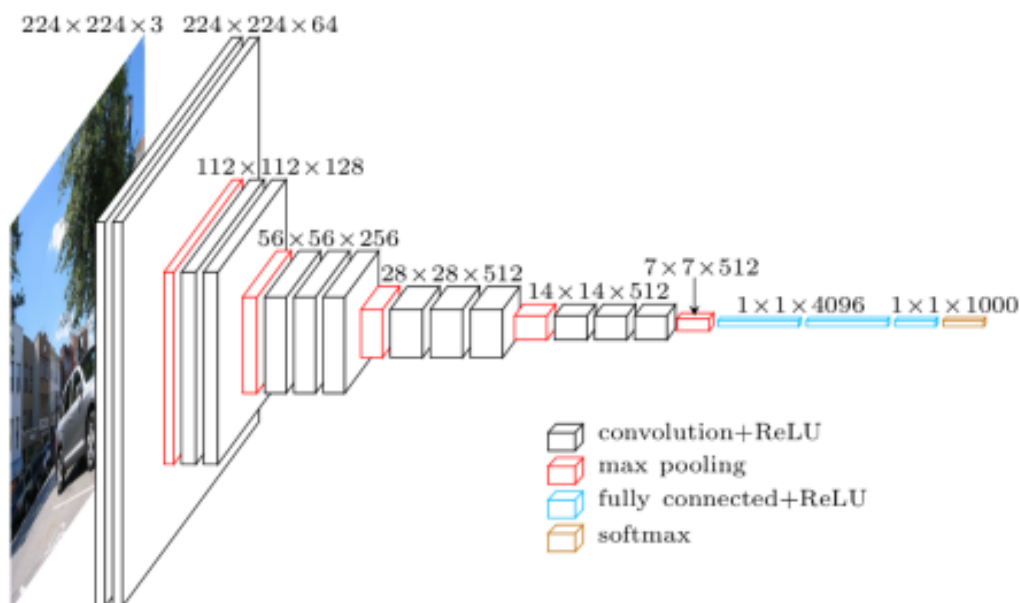


Σχήμα 4.10: Σχηματική Απεικόνιση του μοντέλου ZFNet.

4.4.4 VGGNet (2014)

Το VGGNet είναι το CNN που έχει δείξει τη σημασία του βάθους του δικτύου στην απόδοση του μοντέλου. Κυκλοφόρησαν το 2014 από την ομάδα VGG του πανεπιστημίου της Οξφόρδης [25]. Έχει παρόμοια φιλοσοφία με το δίκτυο AlexNet, με βελτιώσεις, κυρίως, τη δομή του δικτύου [26]. Η βασική συνεισφορά του ήταν το μεγάλο βάθος του δικτύου που αποτελεί κρίσιμο στοιχείο για την επίτευξη καλύτερης αναγνώρισης/ταξινόμησης στα DNN. Πρωτοπαρουσιάστηκε στο διαγωνισμό του ILSVRC 2014 και κατέκτησε την δεύτερη θέση (όπως και το GoogleNet).

Έχει 16 Συνελκτικά/Πλήρη επίπεδα, και το χαρακτηρίζει μία εξαιρετικά ομογενή αρχιτεκτονική που εκτελεί μόνο 3x3 συνελίξεις και 2x2 συγκεντρώσεις από την αρχή μέχρι το τέλος (Σχήμα 4.11). Το μειονέκτημα αυτού του μοντέλου βρίσκεται στο ότι χρησιμοποιεί περισσότερη μνήμη και παραμέτρους (138×10^6), οι περισσότερες από τις οποίες βρίσκονται στο πρώτο πλήρως συνδεδεμένο επίπεδο.



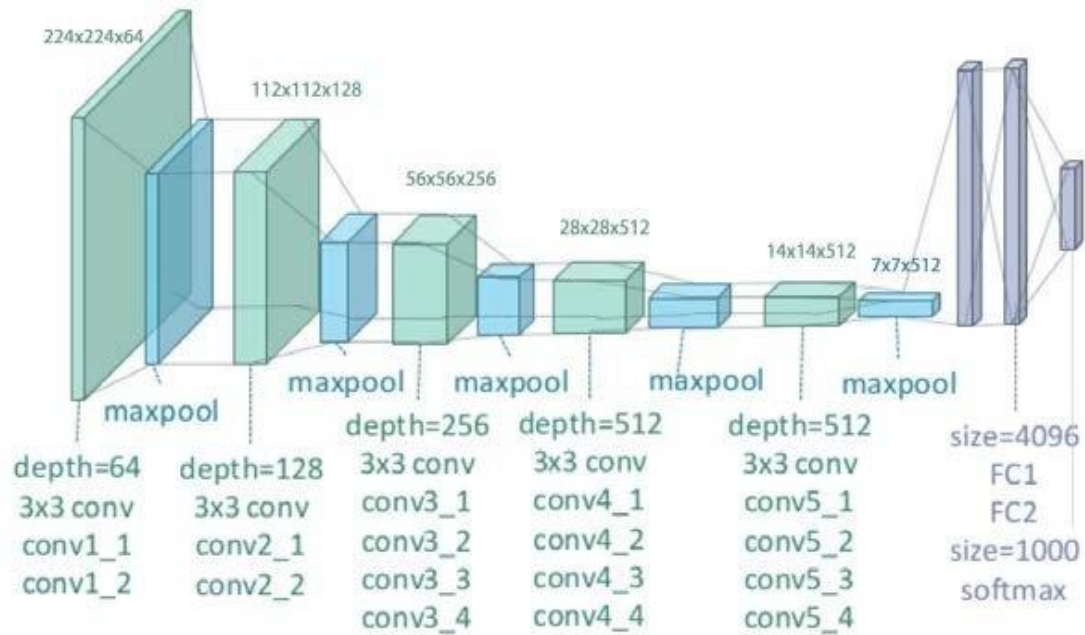
Σχήμα 4.11: Το Συνελικτικό Νευρωνικό Δίκτυο VGGNet.

Τα "16" και "19" στο VGGxx, αντιπροσωπεύουν τον αριθμό των στρωμάτων βάρους στο δίκτυο. Το VGG19 έχει μόλις 3 επιπλέον επίπεδα.

Αρχικά εγκαταλείφθηκε η χρήση μεγάλων φίλτρων (11x11 και 5x5) και στη θέση τους εισήχθησαν μόνο 3x3 φίλτρα σε μεγαλύτερο βάθος δικτύου. Το δίκτυο VGG16 εκπαιδεύτηκε πάνω στο σύνολο δεδομένων **ImageNet**¹. Στη δομή του VGG16 υπήρχε η καινοτομία, ότι κάνει πολλαπλές συνελιξεις στο ίδιο επίπεδο με διαφορετικούς πυρήνες πριν να εισάγει το pooling layer 2x2. Αυτό έχει ως αποτέλεσμα να αυξάνει τη μη γραμμικότητα, αφού σε ένα επίπεδο υπάρχουν 2 ή και 3 μη γραμμικές συναρτήσεις που εφαρμόζονται.

Το VGG19, (Σχήμα 4.12), όπως προ είπαμε, έχει τρία περισσότερα στρωμάτων βάθους στο δίκτυο, σε σχέση με το VGG16. Η αρχιτεκτονική του VGG19 έχει ως εξής: Ως είσοδο δέχεται εικόνες προκαθορισμένου μεγέθους (224x224 εικονοστοιχείων, σε RGB). Η μόνη απαραίτητη προ επεξεργασία είναι η αφαίρεση της μέσης τιμής στο RGB για το σύνολο εκπαίδευσης από κάθε εικονοστοιχείο της εικόνας. Το επίπεδο εισόδου ακολουθείται από 16 συνελικτικά επίπεδα με πυρήνες μεγέθους 3x3. Η υποδειγματοληψία γίνεται από 5 επίπεδα max pooling που βρίσκονται μετά το 2^ο, 4^ο, 8^ο, 12^ο και 16^ο συνελικτικό επίπεδο, αντίστοιχα. Μετά το συνελικτικό τμήμα του δικτύου, βρίσκονται 3 πλήρως συνδεδεμένα επίπεδα. Τα δύο πρώτα αποτελούνται από 2¹² νευρώνες ενώ το τελευταίο από 2¹⁰ (όσες και οι κλάσεις στο σύνολο εκπαίδευσης του δικτύου). Όλα τα κρυφά επίπεδα κάνουν χρήση της μη-γραμμικής συνάρτησης ενεργοποίησης Relu.

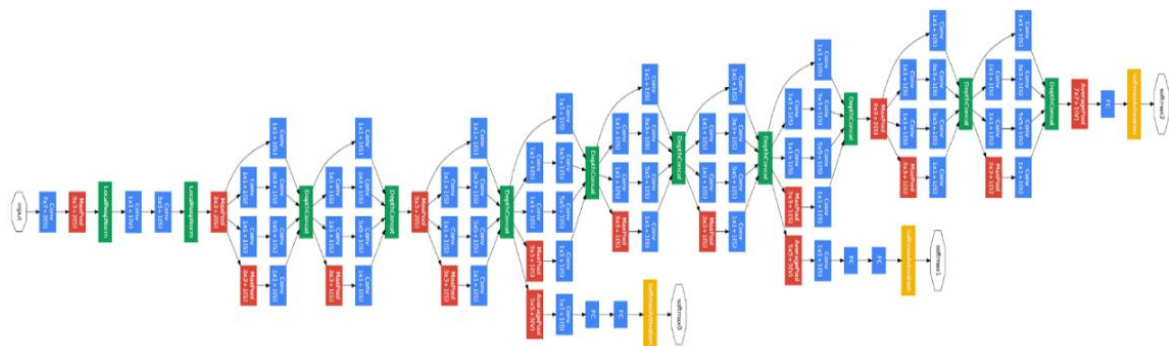
¹ **ImageNet** είναι ένα σύνολο δεδομένων με εκατομμύρια ετικετοποιημένες (labeled) εικόνες υψηλής ανάλυσης που ανήκουν περίπου σε 22k κατηγορίες. Οι εικόνες συλλέχθηκαν από το Διαδίκτυο και επισημάνθηκαν από ανθρώπους. Από το 2010, ως μέρος της πρόκλησης. Το ILSVRC χρησιμοποιεί ένα υποσύνολο του ImageNet με περίπου 1.000 εικόνες σε κάθε μία από τις 1.000 κατηγορίες. Υπάρχουν περίπου 1,2 εκατομμύρια εικόνες εκπαίδευσης (training), 50k επικύρωση (validation) και 150k εικόνες δοκιμών (testing).



Σχήμα 4.12: Το Συνελικτικό Νευρωνικό Δίκτυο VGG-19.

4.4.5 GoogLeNet(InceptionV1) (2014)

Το δίκτυο GoogLeNet (InceptionV1) κατέκτησε τη δεύτερη θέση στο διαγωνισμό LSVRC [27] το 2014. Η βασική συνεισφορά του ήταν η ανάπτυξη του λεγόμενου “InceptionModule” το οποίο μείωσε δραστικά τον αριθμό των παραμέτρων του δικτύου, από 60 εκατομμύρια που είχε το AlexNet σε 4 εκατομμύρια. Επιπρόσθετα, χρησιμοποιεί μέση συγκέντρωση (average pooling) αντί για πλήρως συνδεδεμένα επίπεδα, στην κορυφή του ΣΝΔ, εξαλείφοντας έτσι μεγάλο πλήθος παραμέτρων οι οποίες δε δείχνουν τελικά να έχουν και τόσο μεγάλη σημασία. Στο δίκτυο προστέθηκαν δύο βοηθητικές εξόδους κατά μήκος του, όπως φαίνεται στο Σχήμα 4.13. Αργότερα διαπιστώθηκε, ότι η πρώτη βοηθητική έξοδος δεν προσέδιδε ιδιαίτερη διαφορά στην ποιότητα του τελικού αποτελέσματος του δικτύου. Αυτή η προσθήκη των βοηθητικών εξόδων αρχικά ωφέλησε τις τελικές επιδόσεις του μοντέλου, συγκλίνοντας σε μια ελαφρώς καλύτερη τιμή σε σχέση με το ίδιο δίκτυο που δεν χρησιμοποιούσε βοηθητικές εξόδους.



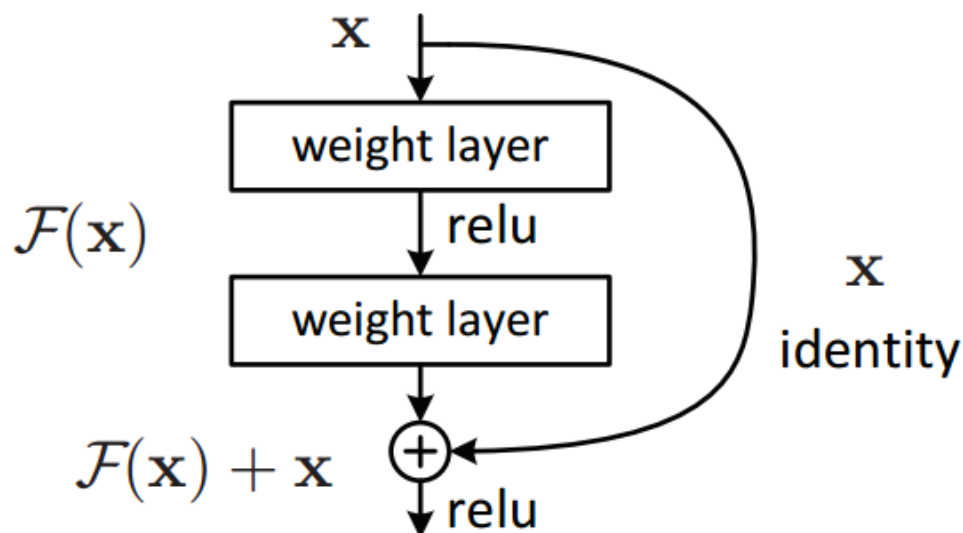
Σχήμα 4.13: Σχηματική Απεικόνιση του μοντέλου GoogLeNet.

4.4.6 ResNet (2015)

Το δίκτυο ResNet είναι η συντόμευση του Residual Network, και όπως υποδηλώνει το όνομα του δικτύου, η νέα ορολογία που εισάγει είναι η **υπολειμματική μάθηση** (residual learning). Από το AlexNet, η σύγχρονη αρχιτεκτονική CNN πηγαίνει σε όλο και πιο βαθιά νευρωνικά δίκτυα. Ενώ το AlexNet είχε μόνο 5 επίπεδα, το δίκτυο VGG και το GoogleNet είχαν 19 και 22 επίπεδα αντίστοιχα το ResNet φτάνει σε εκατοντάδες επίπεδα. Το **ResNet** [28], πρωτοπαρουσιάστηκε στο διαγωνισμό ILSVRC του 2015 [37], από τη Microsoft και ήταν και ο νικητής του, επιτυγχάνοντας την καλύτερη μέχρι τότε ακρίβεια. Για να το επιτύχει αυτό ξεπέρασε κατά πολύ την αντίληψη που υπήρχε για το βάθος των Νευρωνικών Δικτύων. Το ResNet αποτελεί σήμερα το κορυφαίο Βαθύ Νευρωνικό Δίκτυο και καθιστά δυνατή την εκπαίδευση έως και εκατοντάδων ή και χιλιάδων επιπέδων και επιτυγχάνει ακόμα εντυπωσιακή απόδοση.

Η βασική ιδέα του ResNet είναι η εισαγωγή μιας λεγόμενης «**σύνδεσης συντόμευσης**» που παραλείπει ένα ή περισσότερα επίπεδα, όπως φαίνεται στο δομικό στοιχείο του σχήματος 4.14. Σε αυτό παρατηρούμε ότι ξεφεύγουμε από τη συμβατική ροή επεξεργασίας των δικτύων στην οποία το κάθε επίπεδο λαμβάνει μία είσοδο x και παράγει μία έξοδο $F(x)$, με την $F()$ να εξαρτάται από το είδος του Επιπέδου. Εδώ, το κάθε Residual Block (δομικό στοιχείο) παράγει μία έξοδο $F(x) + x$. Έτσι, με αυτόν τον τρόπο αντί μετά από κάθε επίπεδο να παράγουμε μία αντιπροσώπευση των δεδομένων η οποία δεν συνδέεται καθόλου με την προηγούμενη, τώρα κάθε Block παράγει μία μεταβολή της εισόδου, η οποία προστίθεται στην πρώτη για να παραχθεί τροποποιημένη έκδοση της εισόδου.

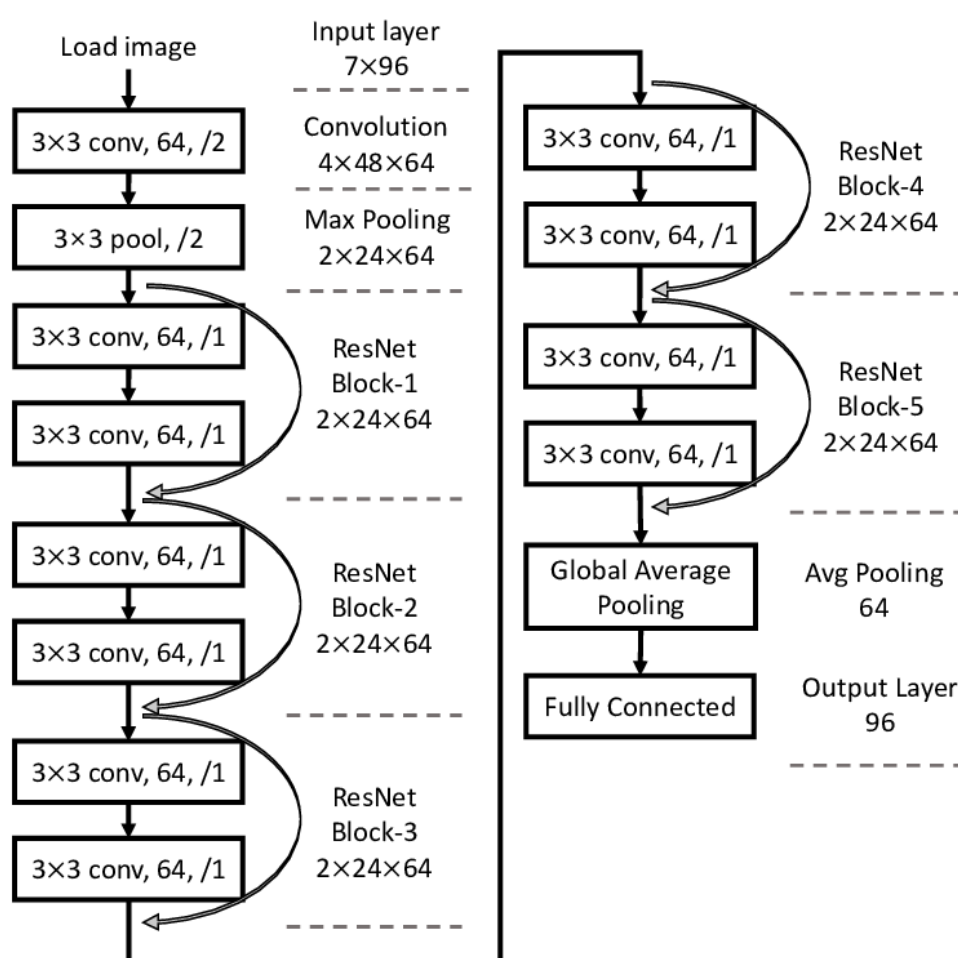
Η διατύπωση του $F(x) + x$ μπορεί να πραγματοποιηθεί μέσω εμπρόσθιας τροφοδότησης με συνδέσεις συντόμευσης (απευθείας σύνδεση εισόδου του $n^{\text{στou}}$ στρώματος σε κάποιο $(n + x)^{\text{στο}}$ στρώμα, το οποίο εμφανίζεται ως καμπύλο βέλος).



Σχήμα 4.14: Residual Learning (υπολειμματική μάθηση): ένα δομικό στοιχείο (Residual Block).

Με τη χρήση του residual network, υπάρχουν πολλά προβλήματα που μπορούν να επιλυθούν όπωσ. Τα ResNets είναι εύκολο να βελτιστοποιηθούν, σε σύγκριση με τα «απλά» δίκτυα (που απλώς στοιβάζουν στρώματα) και εμφανίζουν υψηλότερο σφάλμα εκπαίδευσης κατά την αύξηση του βάθους. Επίσης, μπορούν εύκολα να αποκτήσουν ακρίβεια από πολύ αυξημένο βάθος παράγοντας έτσι αποτελέσματα που είναι καλύτερα από τα προηγούμενα δίκτυα. Ακόμα, μέσω αυτής της αρχιτεκτονικής, επιτρέπεται η ανάπτυξη πολύ βαθύτερων δικτύων, με εκατοντάδες επίπεδα, σε σχέση με τα μέχρι πρότινος προτεινόμενα δίκτυα που είχαν κάποιες δεκάδες επίπεδα. Θα μπορούσαμε να πούμε ότι το ResNet είναι η εξ ορισμού υλοποίηση των βαθιών ΣΝΔ.

Το Residual δίκτυο έχει τρεις εκδόσεις, οι οποίες διαφέρουν κυρίως στον αριθμό των επιπέδων τους: (α) Το ResNet των 50 επιπέδων, (β) το ResNet των 101, και (γ) το ResNet των 152 Επιπέδων. Στο σχήμα 4.15, φαίνεται η δομή ενός ResNet-12.



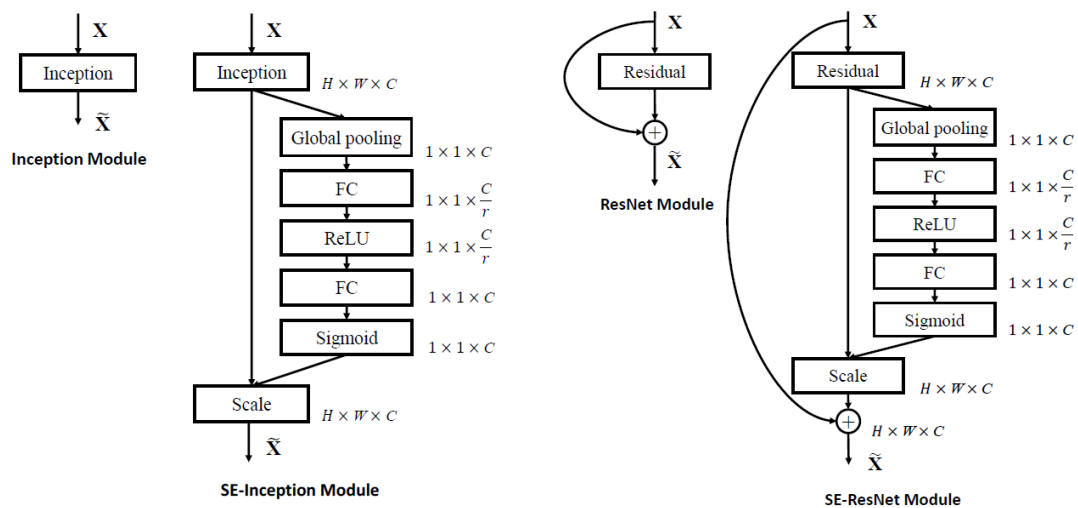
Σχήμα 4.15: Η δομή του ResNet-12.

Το ResNets, έχει αποδείξει ότι η υπολειμματική μάθηση μπορεί να βελτιώσει την απόδοση της εκπαίδευσης μοντέλων, ειδικά όταν το μοντέλο έχει βαθύ δίκτυο με περισσότερα από 20 επίπεδα, και επίσης περιστρέφει το πρόβλημα της υποβαθμισμένης ακρίβειας σε βαθιά δίκτυα.

4.4.7 SENet

Το 2017 παρουσιάστηκε το μοντέλο του SENet [29] νικώντας στο διαγωνισμό ILSVRC2017. Αυτό που εισάγεται με το SENet μοντέλο είναι το “Squeeze-and-Excitation block”, μία αρχιτεκτονική μονάδα που σχεδιάστηκε για να βελτιώνει την αντιπροσωπευτική ισχύ ενός δικτύου, επιτρέποντάς του την επαναβαθμονόμηση των χαρακτηριστικών, κανάλι-προς-κανάλι (channel-wise feature recalibration). Μέσα από πολλά πειράματα αποδείχτηκε η αποτελεσματικότητα των SENets, η οποία αγγίζει τις βέλτιστες αποδόσεις για πολλά και διάφορα σύνολα δεδομένων και εφαρμογές.

Στο σχήμα 4.16 φαίνεται η φιλοσοφία του “Squeeze-and-Excitation” και πως προσαρμόζεται σε γνωστά δίκτυα, βελτιώνοντας έτσι την απόδοσή τους. Αριστερά, δείχνεται ένα αυθεντικό Inception module (GoogLeNet) και δίπλα του το SE-Inception Module. Δεξιά, δείχνεται ένα αυθεντικό Residual module (ResNet) και δίπλα του το SE-ResNet Module.



Σχήμα 4.16: Αριστερά το αυθεντικό Inception module (GoogLeNet) και δίπλα του το SE-Inception Module. Δεξιά ένα αυθεντικό μοντέλο ResNet και δίπλα του ένα SE-ResNet Module.

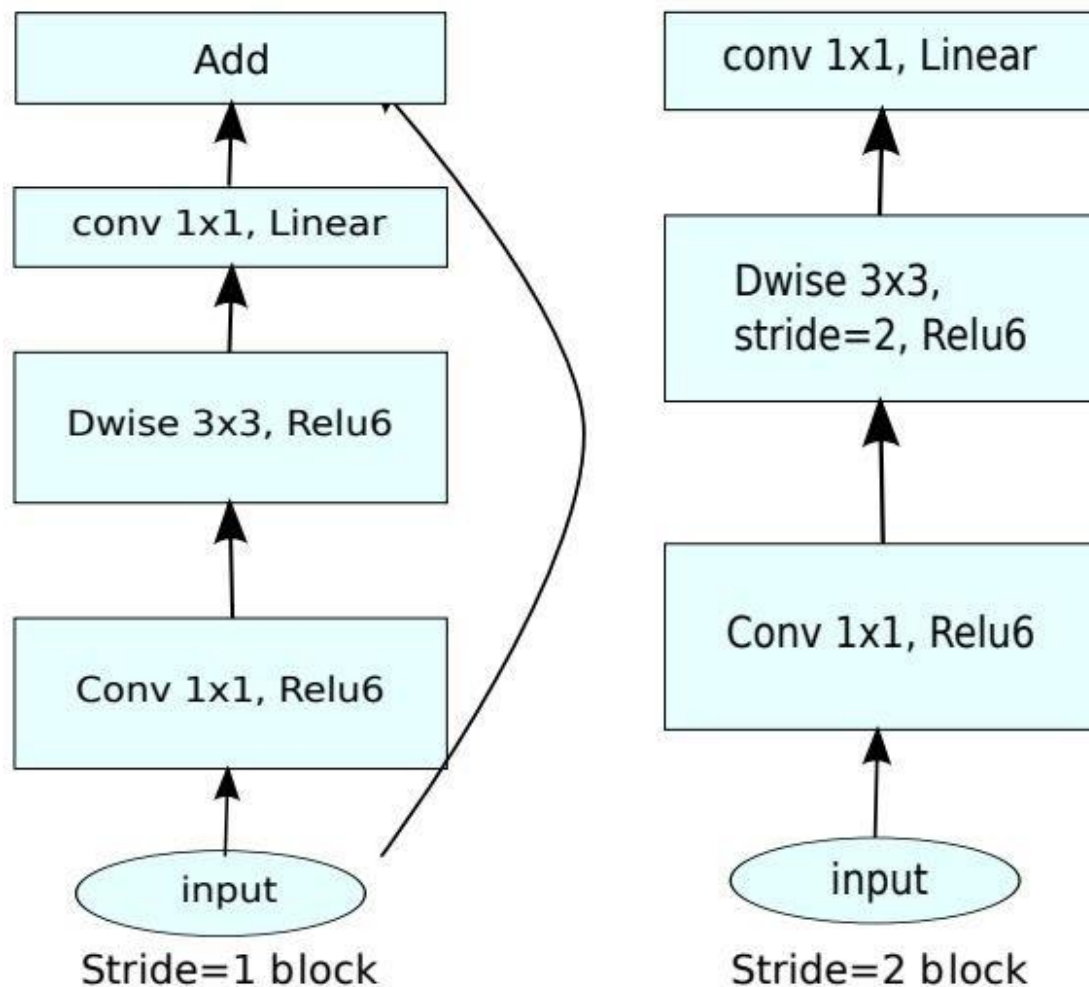
4.4.8 MobileNet

Το MobileNet δημιουργήθηκε τον Απρίλιο του 2017 από την Google [46] και αποτελεί ένα αρκετά αποτελεσματικό Νευρωνικό Δίκτυο που στοχεύει σε εφαρμογή σε κινητά και Ενσωματωμένα Συστήματα (Embedded Systems) που διαθέτουν ισχυρούς περιορισμούς σε ζητήματα μνήμης και υπολογιστικής ικανότητας.

Το MobileNet για να αντιμετωπίσει τους περιορισμούς των Ενσωματωμένων Συστημάτων διαχωρίζει τα Συνελκτικά Επίπεδα σε δύο αντίστοιχα, τα οποία υλοποιούν ακριβώς την ίδια λειτουργία χρησιμοποιώντας λιγότερες παραμέτρους. Ένα συμβατικό Συνελκτικό Επίπεδο υλοποιεί δύο ενέργειες: 1) Φιλτράρει την Είσοδο πολλές φορές και 2) Συνενώνει όλα τα Φιλτραρίσματα σε μία Δομή Εξόδου.

Η αξία του MobileNet έγκειται στο γεγονός ότι κατάφερε να μειώσει σε μεγάλο βαθμό το μέγεθος των δικτύων και να δείξει ότι η χρήση της Βαθιάς Μηχανικής Μάθησης σε Ενσωματωμένα Συστήματα είναι προσιτός στόχος. Ταυτόχρονα, με τις τεχνικές που

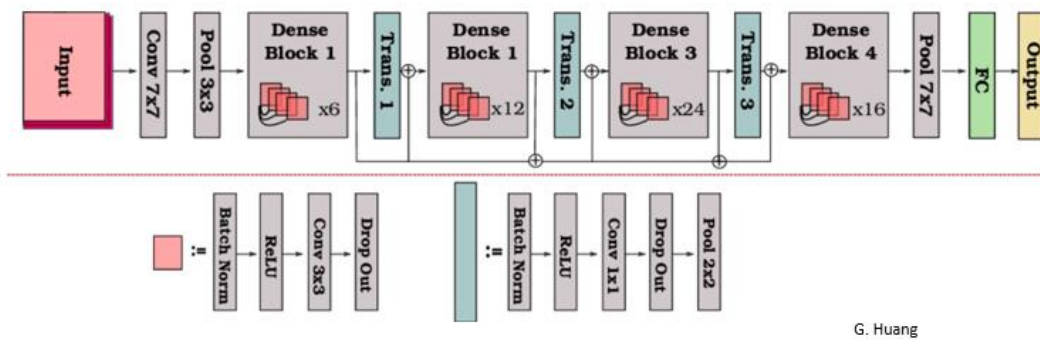
χρησιμοποιεί αυξάνει την ταχύτητα εκτέλεσης κάτι που είναι πολύ σημαντικό για πολλές εφαρμογές των συγκεκριμένων συστημάτων.



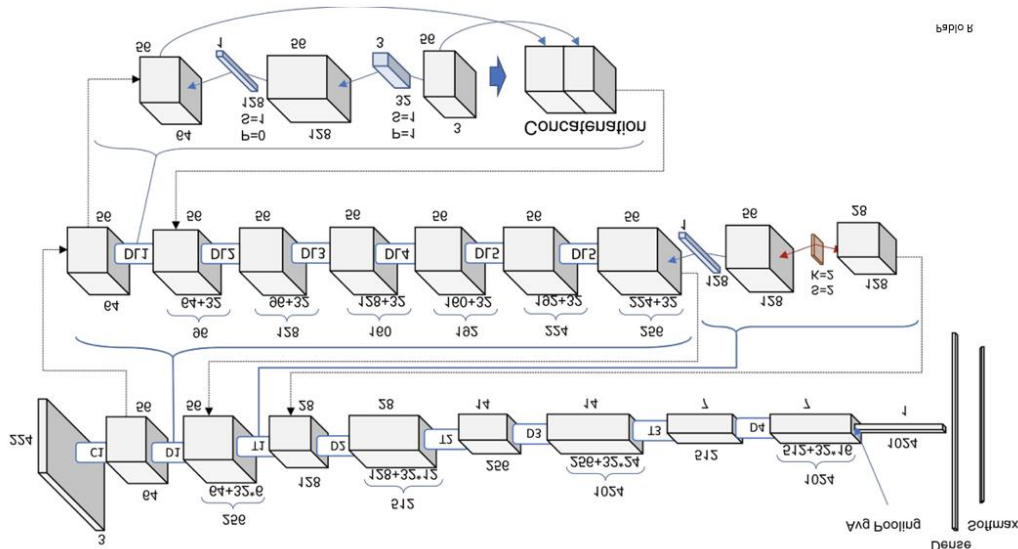
Σχήμα 4.17: Το MobileNet V2.

4.4.9 DenseNets

Τα Πυκνά Νευρωνικά Δίκτυα (Convolutional Networks-**DenseNets**) εντάσσονται στα πιο εξελιγμένα ΣΝΔ, και αποτελούν το επόμενο βήμα στην εμβάθυνση των Συνελκτικών Δικτύων. Είναι αρκετά παρόμοιο με το ResNet με ορισμένες θεμελιώδεις διαφορές. Το ResNet χρησιμοποιεί μια πρόσθετη μέθοδο που συγχωνεύει το προηγούμενο επίπεδο με το επόμενο επίπεδο, ενώ το DenseNet συνενώνει την έξοδο του προηγούμενου στρώματος με το επόμενο επίπεδο. Το DenseNet αναπτύχθηκε ειδικά για να βελτιώσει την φθίνουσα ακρίβεια που προκαλείται, λόγω της μεγαλύτερης διαδρομής μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου, όπου οι πληροφορίες εξαφανίζονται πριν φτάσουν στον προορισμό τους. Δηλαδή, εμφανίζονται προβλήματα με τα CNNs όταν προχωρούν βαθύτερα. Αυτό συμβαίνει γιατί το μονοπάτι των πληροφοριών από το στρώμα εισόδου μέχρι το στρώμα εξόδου γίνεται τόσο μεγάλο, που μπορεί οι πληροφορίες να εξαφανιστούν πριν να φτάσουν στην άλλη μεριά [49]. Στο Σχήμα 4.17 φαίνεται το DenseNet, και μια πλήρη αρχιτεκτονική, σε αφαιρετική μορφή, δείχνεται στο Σχήμα 4.18.



Σχήμα 4.18: Χαρακτηριστικό διάγραμμα του DenseNet [47].



Σχήμα 4.19: Η πλήρης αρχιτεκτονική του DenseNet, σε αφαιρετική μορφή.

4.5 Παράμετροι αξιολόγησης νευρωνικού δικτύου

Οι παράμετροι αξιολόγησης [30-32] της απόδοσης ενός νευρωνικού δικτύου, πρέπει να αναλύονται ώστε να κατανοούμε το ποσοστό της απόδοσης της εξόδου του δικτύου στην εκάστοτε εφαρμογή ταξινόμησης. Επιπλέον όλες οι αξιολογήσεις όπου βγαίνουν είναι εξαρτημένες μεταξύ τους, αλλά και ο τρόπος που εξάγονται προκύπτει από τις τιμές εξόδου του κάθε μοντέλου. Η επιλογή της σωστής παραμέτρου και τις σωστής αξιολόγησης αφορά στη φύση του προβλήματος που αντιμετωπίζουμε και είναι δύσκολη, λόγω των πολλών διαφορετικών παραμέτρων αλλά και προβλημάτων κατηγοριοποίησης όπου υπάρχουν. Αν ο ταξινομητής διαχωρίζει δύο κλάσεις (Positive / Negative), χρησιμοποιούνται τα παρακάτω μέτρα αξιολόγησης:

- **Αληθώς θετικό (True Positive-TP):** Το αποτέλεσμα της εξόδου του νευρωνικού δικτύου είναι θετικό, σε κάποιο άτομο όπου πάσχει από κάποια πάθηση. Επομένως το αποτέλεσμα είναι σωστό.

- **Αληθώς αρνητικό (True Negative-TN):** Το αποτέλεσμα της εξόδου του νευρωνικού δικτύου είναι αρνητικό, σε κάποιο άτομο το οποίο είναι υγιές. Επομένως το αποτέλεσμα είναι σωστό.
- **Ψευδώς θετικό (False Positive-FP):** Το αποτέλεσμα της εξόδου του νευρωνικού δικτύου είναι θετικό, σε κάποιο άτομο το οποίο είναι υγιές. Επομένως το αποτέλεσμα είναι λάθος.
- **Ψευδώς αρνητικό (False Negative-FN):** Το αποτέλεσμα της εξόδου του νευρωνικού δικτύου είναι αρνητικό, σε κάποιο άτομο όπου πάσχει από κάποια πάθηση. Επομένως το αποτέλεσμα είναι λάθος.

πρόβλεψη/πάθηση	Θετικό στην πάθηση	Αρνητικό στην πάθηση
Θετικό στην πρόβλεψη	True Positive(TP)	False Positive (FP)
Αρνητικό στην πρόβλεψη	False Negative (FN)	False Negative (FN)

Πίνακας 4.1: Πίνακας πιθανοτήτων για τη μέτρηση της ακρίβειας μεθόδων ταξινόμησης

Σε συνδυασμό με τις παραπάνω ποσότητες, οι πιο συνήθεις μετρικές που προτείνονται είναι οι ακόλουθες μετρικές:

- **Accuracy (Ορθότητα):** είναι η παράμετρος όπου μετρά το ποσοστό των περιπτώσεων όπου το δίκτυο ταξινόμησε σωστά προς τις συνολικές ταξινομήσεις.

$$(accuracy) = \frac{TP+TN}{TP+TN+FP+FN},$$

- **Precision (Ακρίβεια):** είναι η παράμετρος όπου μετρά το ποσοστό των περιπτώσεων όπου σωστά ταξινομήθηκε από το δίκτυο ως θετικό:

$$(precision) = \frac{TP}{TP+FP},$$

- **Recall ή sensitivity (Ευαισθησία):** είναι η παράμετρος όπου μετρά το ποσοστό των σωστών κατηγοριοποιήσεων του δικτύου:

$$(sensitivity) = \frac{TP}{TP+FN}$$

- **F-score (F-measure):** είναι η παράμετρος όπου μετρά το ποσοστό της συνολικής ακρίβειας της ταξινόμησης (συνολική απόδοση του δικτύου):

$$(F - measure) = \frac{2TP}{2TP+FN+FP},$$

- **Specificity (Προσδιοριστικότητα):** είναι η παράμετρος όπου μετρά το ποσοστό των περιπτώσεων όπου σωστά δεν κατηγοριοποιήθηκε σε κάποια κατηγορία. Λόγος των ορθών αρνητικών προβλέψεων, επί του συνόλου των αρνητικών προτύπων:

$$(specificity) = \frac{TN}{FP+TN}$$

- **fall-out (Πιθανότητα λάθους συναγερμού):** είναι η παράμετρος όπου μετρά το ποσοστό των περιπτώσεων της λάθος πυροδότησης του συναγερμού δηλαδή να ταξινομηθεί λάθος ένα περιστατικό σε κάποια από τις κατηγορίες:

$$(fall - out) = \frac{FP}{FP + TN}$$

-
- **miss-rate (Βαθμός αστοχίας):** είναι η παράμετρος όπου μετρά το ποσοστό των περιπτώσεων όπου ένα περιστατικό έπρεπε να τοποθετηθεί ως θετικό αλλά το δίκτυο το ταξινομήσε ως αρνητικό:

$$(\text{miss} - \text{rate}) = \frac{FN}{TP+FN},$$

Προφανές είναι ότι η επιλογή των κατάλληλων μετρικών εξαρτάται από την εφαρμογή, το δίκτυο, τα δεδομένα, αλλά και τους στόχους που θέλουμε να εξυπηρετήσουμε με την τρέχουσα αξιολόγηση.

5 Πειραματική Διαδικασία

5.1 Εισαγωγή

Το κεφάλαιο αυτό αναφέρεται στην πειραματική διαδικασία που ακολουθήθηκε στα πλαίσια αυτής της διπλωματικής. Αυτό περιλαμβάνει την παρουσίαση των ταξινομητών που δοκιμάστηκαν, τις βάσεις δεδομένων που χρησιμοποιήθηκαν και τους πίνακες και διαγράμματα που προέκυψαν.

Οι διαδικασίες ταξινόμησης που δοκιμάστηκαν αναφέρονται σε τεχνικές ανάλυσης και βελτιστοποίησης βαθιών νευρωνικών δικτύων (DNN) και συγκεκριμένα στα Συνελκτικά Νευρωνικά Δίκτυα (ΣΝΔ) με μεγάλο βάθος. Τα συνελκτικά νευρωνικά δίκτυα (CNN) αποτελούν μια ξεχωριστή κατηγορία δικτύων μηχανικής μάθησης, τα οποία κυριαρχούν σήμερα στις τεχνικές αναγνώρισης προτύπων και ιδιαίτερα στις τεχνικές που αφορούν ταξινόμηση ιατρικών δεδομένων. Τα ΣΝΔ έχουν διάφορες αρχιτεκτονικές, με διαφορετικό βάθος, διαφορετικό πλήθος συνελίξεων, διαφορετικές συναρτήσεις ενεργοποίησης και πολλές άλλες παραμέτρους και τεχνολογίες. Από τις πολλές αρχιτεκτονικές, που περιλαμβάνονται στην οικογένεια των Συνελκτικών Νευρωνικών Δικτύων (Κεφ. 4.), επελέγησαν και δοκιμάστηκαν χχχ διαφορετικές αρχιτεκτονικές.

Κατά τη διαδικασία της εκπαίδευσης και ελέγχου των δικτύων, χρησιμοποιήθηκε η Ιστοπαθολογική Βάση Δεδομένων Καρκίνου του Μαστού, BreaKHis. Μισά από τα δεδομένα της βάσης χρησιμοποιήθηκαν για την εκπαίδευση των συστημάτων και τα άλλα μισά για τον έλεγχο, κατά τον οποίο έπρεπε να ταξινομηθούν τα δεδομένα σε θετικές (κακοήθη) και αρνητικές (καλοήθη), στην ύπαρξη όγκου, εικόνες.

5.2 Βάση Δεδομένων

Στην παρούσα εργασία χρησιμοποιήθηκε η Ιστοπαθολογική Βάση Δεδομένων, καρκίνου του μαστού Breast Cancer Histopathological Database (BreaKHis) [33-34]. Η βάση δεδομένων περιέχει συνολικά 7.909 (με στόχο τις 9.109) εικόνες, ιστού όγκου μαστού, που συλλέχθηκαν από 82 ασθενείς: 58 που πάσχουν από καρκίνο και 24 που έχουν κάποια καλοήθη νεοπλασία. Μέχρι σήμερα, περιέχει 2.480 καλοήθη και 5.429 κακοήθη δείγματα, των 700x460 εικονοστοιχείων, RGB, 3 καναλιών, βάθος 8-bit σε κάθε κανάλι, μορφή PNG.

Η BreaKHis, δημιουργήθηκε σε συνεργασία των εργαστηρίων Visão Robótica e Imagem Laboratório de Pesquisa ligado ao Departamento de Informática και P&D Laboratory –

Pathological Anatomy and Cytopathology, Parana, Brazil [12]. Η BreaKHis βάση δεδομένων αποτελεί χρήσιμο εργαλείο, δεδομένου ότι καθιστά δυνατή τη μελλοντική συγκριτική αξιολόγηση και αξιολόγηση.

5.2.1 Χαρακτηριστικά της Βάσης Δεδομένων

Το σύνολο των δεδομένων της Ιστοπαθολογικής Βάσης Δεδομένων BreaKHis χωρίζεται σε δύο κύριες ομάδες: *καλοήθεις όγκους* και *κακοήθεις όγκους*:

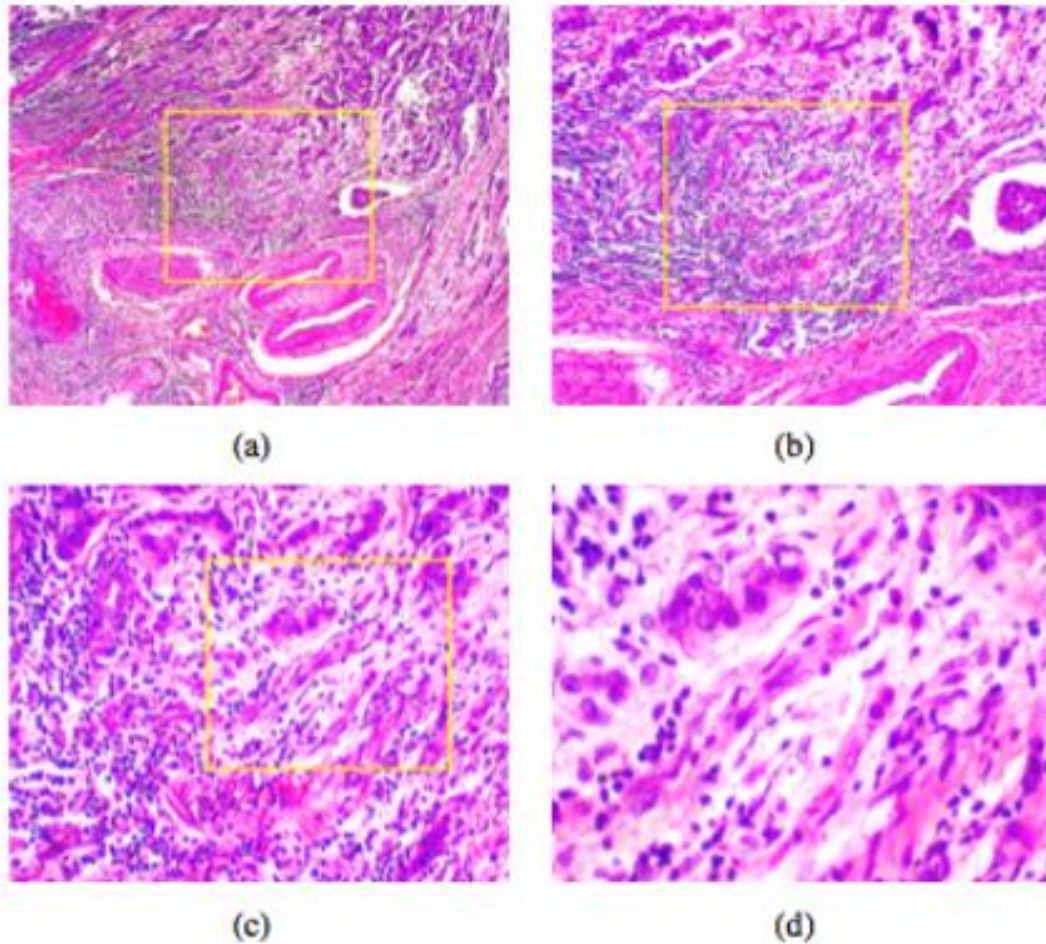
- **Καλοήθης όγκος (benign tumors):** αναφέρεται σε μια βλάβη που δεν ταιριάζει με κανένα κριτήριο κακοήθειας, π.χ., χαρακτηρισμένη κυτταρική ατυπία, μίτωση, διαταραχή των βασικών μεμβρανών, μετάσταση κ.λπ. Κανονικά, οι καλοήθεις όγκοι είναι σχετικά «αθώοι», παρουσιάζουν αργή ανάπτυξη και παραμένουν εντοπισμένοι.
- **Κακοήθης όγκος (malignant tumors):** είναι συνώνυμο του καρκίνου: η βλάβη μπορεί να εισβάλει και να καταστρέψει παρακείμενες δομές (τοπικά επεμβατικές) και να εξαπλωθεί σε απομακρυσμένες περιοχές (μετάσταση) για να προκαλέσει θάνατο.

Τα δείγματα που υπάρχουν στο σύνολο των δεδομένων συλλέχθηκαν με τη μέθοδο της **μερικής μαστεκτομής** ή **εκτομής βιοψίας, (SOB)**. Αυτός ο τύπος διαδικασίας, σε σύγκριση με οποιαδήποτε άλλη μέθοδο βιοψίας βελόνας, αφαιρεί το μεγαλύτερο μέγεθος δείγματος ιστού και γίνεται σε νοσοκομείο με γενικό αναισθητικό.

Μεγέθυνση (Εστίαση)	Καλοήθεις (Benign)	Κακοήθεις (Malignant)	Σύνολο
40x	652 (24)	1.370 (58)	1.995
100x	644 (24)	1.437 (58)	2.081
200x	623 (24)	1.390 (58)	2.013
400x	588 (24)	1.232 (58)	1.820
Σύνολο	2.480	5.429	7.909

Πίνακας 5.1: Η δομή της Ιστοπαθολογικής Βάσης Δεδομένων, Καρκίνου του Μαστού, BreaKHis 1.0.

Η δομή της βάσης δεδομένων BreaKHis 1.0, την οποία χρησιμοποιήσαμε στα πειράματα στην παρούσα διπλωματική, φαίνεται στον Πίνακα 5.1. Οι εικόνες είναι χωρισμένες ανάλογα με την μεγέθυνση του μικροσκοπίου σε 4 κατηγορίες: 40x, 100x, 200x και 400x. Οι ετικέτες αυτού του συνόλου δεδομένων δεν περιορίζονται μόνο σε καλοήθεις και κακοήθεις αλλά επεκτείνονται σε 4 υποκατηγορίες για την κάθε περίπτωση. Συνολικά δηλαδή αντιστοιχίζονται σε 8 διαφορετικές κλάσεις. Στην Εικόνα 5.1, απεικονίζεται Ιστοπαθολογική εικόνα κακοήθους όγκου του μαστού σε 4 διαφορετικούς συντελεστές μεγέθυνσης: (α) 40x, (β) 100x, (γ) 200x και (δ) 400x.



Εικόνα 5.1: Μια εικόνα κακοήθους όγκου του μαστού (βάφτηκε με HE), σε 4 διαφορετικούς συντελεστές μεγέθυνσης: (α) 40X, (β) 100X, (γ) 200X και (δ) 400X.

Τόσο οι καλοήθειες όσο και οι κακοήθειες όγκοι του μαστού μπορούν να ταξινομηθούν σύμφωνα με τα νεοπλασματικά κύτταρα που φαίνονται κάτω από το μικροσκόπιο. Οι διάφοροι τύποι όγκων του μαστού μπορεί να έχουν διαφορετικές προγνώσεις και επιπτώσεις στη θεραπεία. Το σύνολο δεδομένων περιέχει:

- τέσσερις ιστολογικούς διακριτούς τύπους καλοηθών όγκων του μαστού: *αδένωση* (adenosis - **A**), *ινωδοδένωμα* (fibroadenoma - **F**), *όγκος φυλλωδών* (phyllodes tumor- **PT**) και *σωληνοειδές αδένωμα* (tubular adenoma - **TA**), και
- τέσσερις κακοήθεις όγκους (καρκίνος του μαστού): *καρκίνωμα* (carcinoma- **DC**), *λοβιακό καρκίνωμα* (lobular carcinoma - **LC**), *βλεννώδες καρκίνωμα* (mucinous carcinoma - **MC**) και *θηλώδες καρκίνωμα* (papillary carcinoma - **PC**).

Κάθε όνομα αρχείου εικόνας αποθηκεύει πληροφορίες για την ίδια την εικόνα: **μέθοδος της διαδικασίας της βιοψίας, κατηγορία όγκου, τύπος όγκου, ταυτοποίηση ασθενούς και συντελεστής μεγέθυνσης**. Για παράδειγμα, το:

SOB_B_TA-14-4659-40-001.png

αφορά την εικόνα 1, με συντελεστή μεγέθυνσης 40X, ενός καλοήθους όγκου τύπου σωληνοειδούς αδενώματος, πρωτότυπα από τη διαφάνεια 14-4659 που συλλέχθηκε με τη διαδικασία SOB.

Πιο τυπικά, η μορφή του ονόματος αρχείου εικόνας δίνεται από την σημειογραφία BNF του Πίνακα 5.2.

```
<BIOPSY_PROCEDURE>_<TUMOR_CLASS>_<TUMOR_TYPE>-<YEAR>-  
<SLIDE_ID>-<MAG>-<SEQ>  
<BIOPSY_PROCEDURE>::=SOB  
<TUMOR_CLASS>::=M|B  
<TUMOR_TYPE>::=<BENIGN_TYPE>|<MALIGNANT_TYPE>  
<BENIGN_TYPE>::=A|F|PT|TA  
<MALIGNANT_TYPE>::=DC|LC|MC|PC  
<YEAR>::=<DIGIT><DIGIT>  
<PATIENT_ID>::=<NUMBER><SEC>  
<SEQ>::=<NUMBER>  
<MAG>::=40|100|200|400  
<NUMBER>::=<NUMBER><DIGIT>|<DIGIT>  
<SEC>::=<SEC>::<LETTER>|<LETTER>  
<DIGIT>::=0|1|...|9  
<LETTER>::=A|B|...|Z
```

Πίνακας 5.2: Σημειογραφία BNF, για τη μορφή του ονόματος αρχείου εικόνας.

5.2.2 Προεπεξεργασία της Βάσης Δεδομένων

Πρώτο και σημαντικό στάδιο στη διαδικασία της εκπαίδευσης και ελέγχου, ενός συστήματος αποτελεί το στάδιο της προεπεξεργασίας των δεδομένων. Η διαδικασία αυτή, στην πλήρη της εφαρμογή, μπορεί να μας απαλλάξει από πολλά σοβαρά προβλήματα, όπως η διαφορά στη φωτεινότητα που μπορεί να παρουσιάζουν κάποιες εικόνες (πιο σκοτεινές ή πιο φωτεινές), οι αποκρύψεις διάφορων στοιχείων που μπορεί να υπάρχουν (το ζητούμενο σε διαφορετικό μέρος της εικόνας), να παρουσιάζονται οι εικόνες υπό διαφορετικές γωνίες κ.α. Από τα βασικά βήματα που ακολουθούνται σε κάθε στάδιο προεπεξεργασίας είναι τα ακόλουθα:

- Αναδιαμόρφωση του μεγέθους των εικόνων (resize): Στο αρχικό dataset, οι εικόνες έχουν διαστάσεις 700*460*3. σε 224x224x3, διότι τα δίκτυα που θα δοκιμαστούν στη συνέχεια δέχονται εικόνες τέτοιου μεγέθους. Να ταιριάζει με τις διαστάσεις του πρώτου επιπέδου (layer) των συνελκτικών νευρωνικών που θα χρησιμοποιηθεί. Αφού εφαρμοστεί το παραπάνω βήμα, αποθηκεύουμε τις νέες μας εικόνες σε έναν καινούργιο φάκελο (με τη βοήθεια του preprocessing.py).
- Κανονικοποίηση χρωματικού εύρους: Σε κάποιες εικόνες παρατηρούμε ότι οι τιμές που παίρνουν τα χρώματα, ειδικά το μοβ χρώμα, έχει μεγαλύτερο εύρος τιμών. Οπότε, αυτό πρέπει να προσαρμοστεί στο εύρος τιμών των άλλων χρωμάτων (να περιοριστεί το εύρος τιμών του).
- Μη ισορροπημένα δεδομένα (Imbalanced data): Το αρχικό dataset αποτελείται από 7.909 φωτογραφίες, από τις οποίες οι 5.429 είναι Malignant και 2.480 Benign. Αυτή η μεγάλη διαφορά (περισσότερα από διπλάσια) μπορεί να δημιουργήσει πρόβλημα τόσο στην εκπαίδευση όσο και στον έλεγχο. Μπορεί να δημιουργείται πιθανά κάποιο Imbalanced data.
- Επαύξηση δεδομένων εκπαίδευσης και ελέγχου: Ο όγκος των δεδομένων του αρχικού dataset, δεν είναι ιδιαίτερα μεγάλος. Μέσω της χρήσης, γεννητριών συναρτήσεων (εργαλείο του framework) μπορεί να επεκταθεί το αρχικό dataset, με επιπλέον δεδομένα (φωτογραφίες), τροποποιημένα μέσω μετασχηματισμών.
- Τελευταίο βήμα στο στάδιο της προεπεξεργασίας είναι ο διαχωρισμός του συνόλου των δεδομένων σε 2 υποσύνολα (split dataset):

- ο **Δεδομένα Εκπαίδευσης** (trainnig set): βάση των οποίων θα εκπαιδευτεί το δίκτυο.
- ο **Δεδομένα Ελέγχου** (testing set): βάση των οποίων θα αξιολογηθεί η αποτελεσματικότητα του δικτύου. Αποτελείται από εικόνες που δεν έχει ξαναδεί το δίκτυο και ουσιαστικά καλείται να αναγνωρίσει την ετικέτα κάθε δείγματος αυτού του συνόλου.

5.3 Αξιολόγηση των Συστημάτων

Κατά την πειραματική διαδικασία, η αξιολόγηση των συστημάτων γίνεται μέσω διαφόρων μετρικών αξιολόγησης, αρκετές εκ των οποίων παρουσιάσαμε στο προηγούμενο κεφάλαιο. Για τα δικά μας πειράματα, επιλέχθηκαν οι πλέον διαδεδομένες μετρικές [35]: *Accuracy*, *Precision*, *Recall*, και *F1*.

Η μετρική *Accuracy* δείχνει πόσες σωστές προβλέψεις έκανε το μοντέλο σε σχέση με το σύνολο των προβλέψεων που έκανε. Δηλαδή, το ποσοστό των περιπτώσεων όπου το δίκτυο ταξινόμησε σωστά ως προς τις συνολικές ταξινομήσεις:

$$Accuracy = [TruePositive + TrueNegative] / TotalPopulation$$

Η λογική πίσω από την επιλογή της μετρικής *precision* είναι πόσα είναι τα σωστά μεταξύ αυτών που ο αλγόριθμος μας χαρακτήρισε ως σωστά:

$$precision = True\ Positive / (True\ Positive + False\ Positive)$$

Η *recall*, υπολογίζει το πόσα σωστά μας επέστρεψε ο αλγόριθμος σε σχέση με όλα τα σωστά στο αρχικό δείγμα:

$$recall = True\ Positive / (True\ Positive + False\ Negative)$$

Ωστόσο, τόσο το *precision* όσο και το *recall*, ανεξάρτητα, παρουσιάζουν αδυναμίες. Για παράδειγμα, αν το *Precision* εμφάνιζε ότι στα σωστά είναι μόνο ένα στοιχείο αυτό που είχε μεγαλύτερο σκορ τότε πιθανότατα θα είναι 1 (αφού τα FP θα ήταν 0). Αντίστοιχα, αν στην *Recall*, υποθεθεί ότι όλα τα δείγματα ανήκουν στην κατηγορία ενδιαφέροντος, τότε το *recall* είναι 1. Οπότε, ούτε η μία ούτε η άλλη, ξεχωριστά, μπορεί να χρησιμοποιηθεί για την αξιολόγηση ενός αλγόριθμου μηχανικής μάθησης. Όμως, ο συνδυασμός και των δύο παραμέτρων μαζί βγάζουν τα σωστά αποτελέσματα. Ένας τέτοιος συνδυασμός είναι η μετρική *F1* (ή *F-measure*) που υπολογίζεται ως:

$$F1 = 2 / \{ (1/precision) + (1/recall) \}$$

και παίρνει τιμές από 0 έως 1.

Η εκπαίδευση ενός μοντέλου σημαίνει μάθηση των σωστών τιμών, για τα **βάρη** (weights) και τις **προκαταλήψεις** (bias), από επισημασμένα (labeled) παραδείγματα. Ο στόχος λοιπόν της εκπαίδευσης ενός μοντέλου είναι να βρει ένα σύνολο βαρών και προκαταλήψεων που να δίνουν χαμηλή απώλεια (**loss**), κατά μέσο όρο, σε όλα τα παραδείγματα. Απώλεια (**Loss**) είναι ένας αριθμός που δείχνει πόσο κακή ήταν η πρόβλεψη του μοντέλου σε ένα μόνο παράδειγμα. Εάν η πρόβλεψη του μοντέλου είναι τέλεια, η απώλεια είναι μηδέν. Διαφορετικά, η απώλεια είναι μεγαλύτερη.

Το μοντέλο αξιολογείται (evaluated), αναφέροντας την ακρίβεια ταξινόμησης των δεδομένων εκπαίδευσης και ελέγχου (train and test sets) αντίστοιχα. Να σημειωθεί ότι τα αποτελέσματα ενδέχεται να διαφέρουν ανάλογα με τη στοχαστική φύση του αλγορίθμου εκπαίδευσης.

Αναλυτικότερα, τα πειραματικά αποτελέσματα τα παρουσιάζουμε σε πίνακες, με τις μέσες τιμές των μετρικών αξιολόγησης που προαναφέραμε και με σχηματικές απεικονίσεις, σε συνάρτηση με τις εποχές (epoch).

Οι σχηματικές απεικονίσεις (plots) δείχνουν δύο γραφικές παραστάσεις:

- μία για τις καμπύλες εκμάθησης της απώλειας (**loss**), στην εκπαίδευση (train set) και τον έλεγχο (test set), (Loss vs. Epoch), και
- μία για την ταξινόμηση στα sets εκπαίδευσης και τα ελέγχου (Accuracy vs. Epoch).

5.4 Βιβλιοθήκες

5.4.1 Tensorflow

Το Tensorflow, είναι μια συλλογή βιβλιοθηκών ανοιχτού κώδικα, που σχεδιάστηκε από την Google, με μεγάλη απήχηση τόσο στον ερευνητικό όσο και στον βιομηχανικό τομέα. Έχει μεγάλη εφαρμογή στον σχεδιασμό και στην υλοποίηση συστημάτων μηχανικής μάθησης. Επιπλέον ο τρόπος με τον οποίο έχει σχεδιαστεί, το συγκεκριμένο λογισμικό, δίνει την δυνατότητα η κάθε εφαρμογή να μπορεί να αναπτυχθεί σε οποιαδήποτε πλατφόρμα, από επεξεργαστές, κάρτες γραφικών ακόμα και servers.

Η ονομασία του tensorflow προέρχεται από τους τένσορες (tensor) αλλά και τα γραφήματα ροής δεδομένων (data flow graphs). Τένσορας είναι ένας πίνακας n-διαστάσεων που περιέχει μέσα κάποιους τύπους δεδομένων. Όταν κάποιος προγραμματιστής μέσω της Python χρησιμοποιεί την συγκεκριμένη βιβλιοθήκη για κάποια εφαρμογή, στην ουσία δημιουργείται ένα γράφημα ροής δεδομένων, όπου κάθε κόμβος είναι ένας τένσορας. Επιπλέον, σε κάθε κόμβο γίνονται εσωτερικά κάποιοι υπολογισμοί και το τελικό αποτέλεσμα μεταφέρεται στον επόμενο κόμβο. Επομένως, όλοι οι κόμβοι έχουν μεταξύ τους κάποια εξάρτηση, η οποία καθορίζει το πως μεταφέρεται η πληροφορία από τον έναν κόμβο στον άλλο.

5.4.2 Keras

Το keras [36] είναι μια ανοιχτού λογισμικού βιβλιοθήκη γραμμένη σε python, που τρέχει πάνω στο tensorflow. Είναι πολύ φιλικό, καθώς ελαχιστοποιεί τον αριθμό των ενεργειών του χρήστη και δίνει τη δυνατότητα να χρησιμοποιηθούν μεγάλα νευρωνικά δίκτυα (π.χ. CNN). Ακόμα, προσφέρει αρκετές τεχνικές για την προεπεξεργασία των δεδομένων, με ελάχιστες γραμμές κώδικα. Επιπλέον, όπως και το tensorflow, δίνει τη δυνατότητα σε κάθε εφαρμογή να μπορεί να αναπτυχθεί σε οποιαδήποτε πλατφόρμα.

Η βιβλιοθήκη Keras, προσφέρει μια πιο υψηλού επιπέδου δυνατότητα για εύκολη κατασκευή μεγάλου βάθους νευρωνικών δικτύων, ανεξάρτητα του πάνω σε ποια βιβλιοθήκη βρίσκεται (Tensorflow ή Theano ή κάποια άλλη). Ακόμα, περιέχει πολλές υλοποιήσεις από κοινά χρησιμοποιούμενα blocks νευρωνικών δικτύων, όπως τα επίπεδα (layers), οι συναρτήσεις απώλειας (loss functions), οι συναρτήσεις ενεργοποίησης (activation functions), οι αλγόριθμοι βελτιστοποίησης (optimizers) και αρκετά εργαλεία για να κάνει την επεξεργασία εικόνων και κειμένων ευκολότερη. Συμπεριλαμβάνει και άλλες λειτουργίες επιπέδων όπως: το Dropout, το Batch, το Normalization και το Pooling. Επιπλέον δίνει τη δυνατότητα στους χρήστες να μπορούν να κάνουν την εκπαίδευση των μοντέλων βαθιάς μάθησης, με τη χρήση καρτών γραφικών, πράγμα το οποίο εξοικονομεί πολύ χρόνο σε σχέση με τη χρήση επεξεργαστή.

Τέλος, δίνεται η δυνατότητα στο χρήστη να μπορεί να φορτώσει στο πρόγραμμα του κάποια γνωστή αρχιτεκτονική. Είναι πολλές οι επιλογές που παρέχει η βιβλιοθήκη Keras, μεταξύ των οποίων είναι αρχιτεκτονικές ΣΝΔ: Xception, VGG16, VGG19, ResNet, ResNetV2, ResNeXt, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet, NASNet. Ακόμα, υπάρχει η επιλογή αυτά τα δίκτυα να φορτωθούν είτε με τυχαία βάρη είτε με βάρη που προέκυψαν κατά από προεκπαίδευση με την βάση δεδομένων εικόνων, ImageNet, που προαναφέραμε.

Με τη χρήση επίσης των τεχνικών Transfer Learning & Fine Tuning συνδυαστικά με τις προαναφερθείσες δυνατότητες που παρέχει η Keras, μπορούν να επιτευχθούν γρήγορα και εύκολα πολύ υψηλά ποσοστά ακρίβειας πρόβλεψης.

5.5 Πειραματικά Αποτελέσματα

Κατά την πειραματική διαδικασία δοκιμάστηκαν γνωστές αρχιτεκτονικές ΣΝΔ συστημάτων. Οι βιβλιοθήκες **Tensorflow** και **Keras** [36] παρείχαν τη δυνατότητα χρησιμοποίησης σχετικών προγραμμάτων.

Είναι πολλές οι επιλογές που παρέχει η βιβλιοθήκη Keras, σε αρχιτεκτονικές δικτύων ΣΝΔ. Υπάρχει η επιλογή αυτά τα δίκτυα να φορτωθούν είτε με τυχαία βάρη είτε με βάρη από προεκπαίδευση. Η επιλογή με προεκπαιδευμένα βάρη είναι εξαιρετικά σημαντικό χαρακτηριστικό των προσφερόμενων δικτύων. Αυτό μας δίνει τη δυνατότητα να χρησιμοποιήσουμε τα δεδομένα της εκπαίδευσής μας, ουσιαστικά, για fine tuning των παραμέτρων των δικτύων. Αυτός είναι ένας από τους βασικούς λόγους που επιλέχθηκε αυτή η βιβλιοθήκη.

Στην συνέχεια παραθέτουμε τα πειραματικά αποτελέσματα από τις διάφορες αρχιτεκτονικές ΣΝΔ που χρησιμοποιήσαμε στην παρούσα εργασία. Όλα τα πειράματα επαναλήφθηκαν προσθέτοντας στάδια προεπεξεργασίας στα δίκτυα. Συγκεκριμένα, όλα τα πειράματα έγιναν τρεις φορές: **χωρίς** προεπεξεργασία, με προεπεξεργασία **PCA**, και με προεπεξεργασία **ZCA**.

5.5.1 Πειραματικά Αποτελέσματα Δικτύων, Χωρίς Προεπεξεργασία

Κατά την πειραματική διαδικασία δοκιμάστηκαν οκτώ (8) διαφορετικές αρχιτεκτονικές ΣΝΔ. Τα πειραματικά αποτελέσματα των δικτύων χωρίς προεπεξεργασία, παρουσιάζονται στη συνέχεια αυτής της ενότητας.

Συγκεκριμένα, παραθέτουμε πίνακες με μετρήσεις μέσων τιμών (mean) των μετρικών, *Accuracy*, *Precision*, *Recall*, *F1* και *Loss*, αλλά και της τυπικής απόκλισης (*standard deviation*) των τιμών αυτών. Οι μετρήσεις αυτές αναφέρονται στις τέσσερις (4) κατηγορίες εστίασης (**40x**, **100x**, **200x**, **400x**). Ακόμα, δείχνονται σχηματικές απεικονίσεις (plots) δύο γραφικών παραστάσεων: μία με τις καμπύλες απώλειας (**loss**), για εκπαίδευση (train set) και έλεγχο (test set), σε συνάρτηση με τις εποχές (Loss vs. Epoch), και μία για την *Accuracy*, για sets εκπαίδευσης και ελέγχου, σε συνάρτηση με τις εποχές (Accuracy vs. Epoch).

5.5.1.1 VGG16

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **VGG16**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.1, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης

(40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

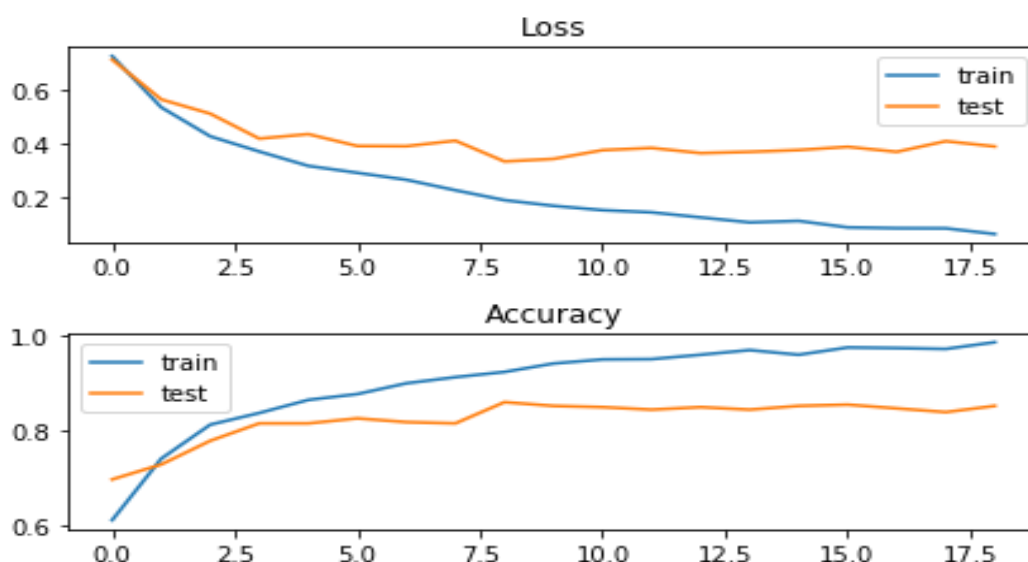
<u>VGG16</u>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	86.235%(0.0103)	42.339%(0.0689)	90.010%(0.0061)	87.189%(0.0106)	93.184%(0.0056)
100x	88.286%(0.0067)	49.289%(0.0144)	92.584%(0.0082)	88.786%(0.0089)	96.101%(0.0105)
200x	87.784%(0.0085)	52.226%(0.0319)	92.631%(0.0055)	90.374%(0.0029)	95.142%(0.0118)
400x	86.257%(0.0091)	41.645%(0.0120)	90.123%(0.0212)	90.676%(0.0170)	89.807%(0.0279)

Πίνακας 5.1: Μετρικές αποτελεσμάτων του μοντέλου *VGG-16*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.1, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλύτερες τιμές είναι:

- *Accuracy*=**88.286%** (για 100x), *Loss*=**41.645%** (για 400x), *F1*=**92.631%** (για 200x), *Precision*=**90.676%** (για 400x), *Recall*=**96.101%** (για 100x).

Στο Σχήμα 5.1, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες *Accuracy* (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **40x**.



Σχήμα 5.1: *Vgg16_40x*: *Accuracy*=87.071, *Loss*=37.716, *F1-score*=90.834, *Precision*=88.168, *Recall*=93.797.

5.5.1.2 VGG19

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **VGG19**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.2, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

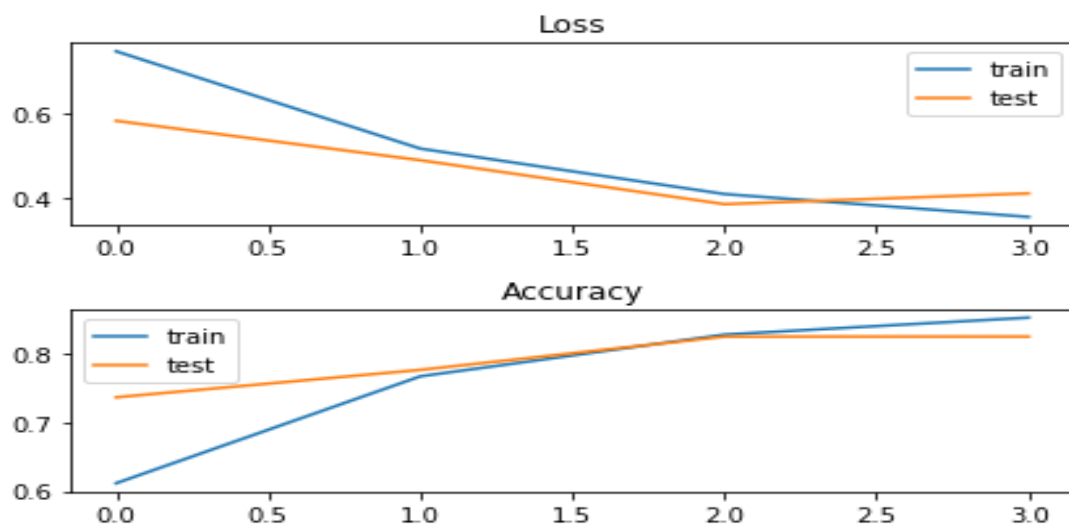
<u>VGG19</u>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	86.708%(0.0064)	38.956%(0.057)	90.457%(0.0027)	87.700%(0.0049)	93.805%(0.0076)
100x	85.666%(0.002)	42.198%(0.0143)	90.711%(0.0062)	86.969%(0.0114)	94.061%(0.0050)
200x	86.257%(0.0116)	40.427%(0.0656)	83.438%(0.0807)	80.685%(0.0769)	86.902%(0.0838)
400x	86.148%(0.0084)	35.044%(0.0462)	90.394%(0.0073)	89.890%(0.0187)	90.897%(0.0253)

Πίνακας 5.2: Μετρικές αποτελεσμάτων του μοντέλου **VGG-19**, για εστίαση **40x**, **100x**, **200x**, **400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.2, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**86.708%** (για 40x), *Loss*=**35.044%** (για 400x), *F1*=**90.711%** (για 100x), *Precision*=**89.890%** (για 400x), *Recall*=**94.061%** (για 100x).

Στο Σχήμα 5.2, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες *Accuracy* (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **400x**.



Σχήμα 5.2: VGG19_400x: Accuracy=87.143, Loss=35.005, F1=91.031, Precision=90.831, Recall=91.369.

5.5.1.3 Resnet50

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου Resnet50, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.3, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

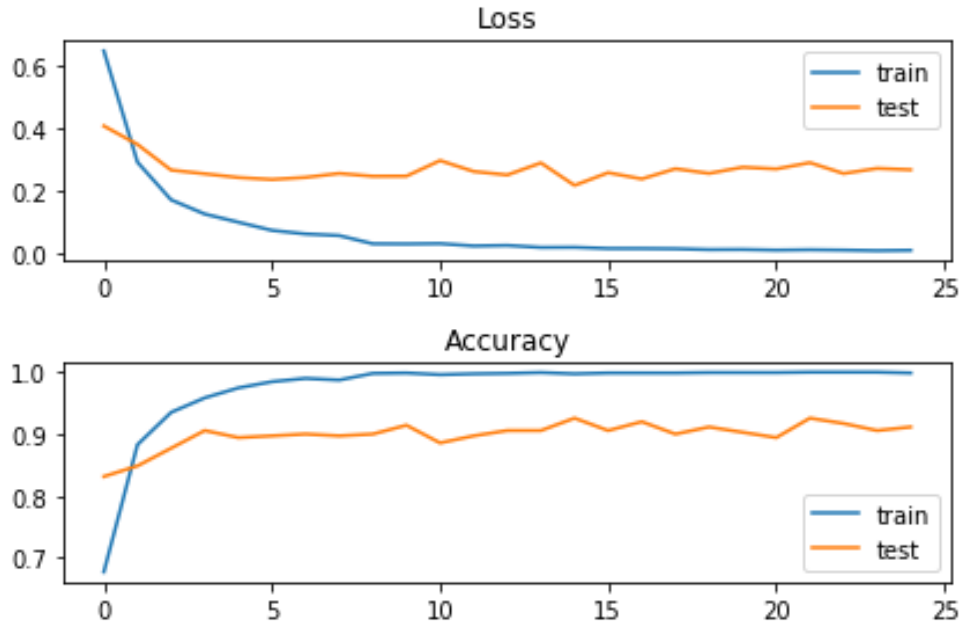
<i>Resnet50</i>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	88.210%(0.0062)	32.653%(0.0263)	91.114%(0.0121)	90.003%(0.0101)	91.967%(0.0163)
100x	90.335%(0.0092)	29.282%(0.0162)	92.892%(0.0102)	92.609%(0.0130)	92.998%(0.0233)
200x	90.007%(0.0051)	28.896%(0.0229)	90.952%(0.0625)	89.016%(0.0685)	92.926%(0.0558)
400x	90.297%(0.0095)	26.370%(0.0158)	92.305%(0.0142)	93.437%(0.0064)	91.723%(0.0256)

Πίνακας 5.3: Μετρικές αποτελεσμάτων του μοντέλου *Resnet50*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.3, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- Accuracy=**90.335%** (για 100x), Loss=**26.370%** (για 400x), F1=**92.892%** (για 100x), Precision=**93.437%** (για 400x), Recall=**92.998%** (για 100x).

Στο Σχήμα 5.3, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες Accuracy (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **400x**.



Σχήμα 5.3: resnet50_400x: Accuracy=91.429, Loss=23.626, F1-score=92.966, Precision=93.887, Recall=92.144.

5.5.1.4 Resnet101

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **Resnet101**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.4, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

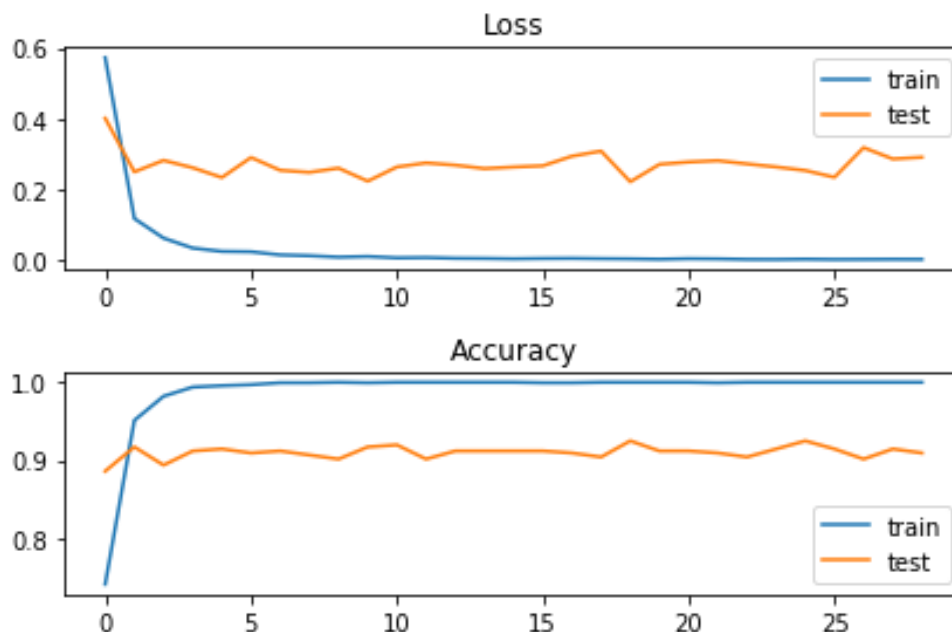
<i>Resnet101</i>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	90.932%(0.0116)	20.917%(0.0254)	92.787%(0.0075)	95.641%(0.0066)	90.365%(0.0143)
100x	92.655%(0.0075)	22.561%(0.0164)	94.821%(0.0099)	93.945%(0.0109)	95.622%(0.0092)
200x	90.831%(0.0041)	26.335%(0.0077)	92.247%(0.0032)	93.728%(0.0086)	94.970%(0.0085)
400x	88.431%(0.0204)	31.247%(0.075)	91.348%(0.0153)	92.221%(0.0129)	90.834%(0.0182)

Πίνακας 5.4: Μετρικές αποτελεσμάτων του μοντέλου **Resnet101**, για εστίαση **40x**, **100x**, **200x**, **400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.4, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**92.655%** (για 100x), *Loss*=**20.917%** (για 40x), *F1*=**94.821%** (για 100x), *Precision*=**95.641%** (για 40x), *Recall*=**95.622%** (για 100x).

Στο Σχήμα 5.4, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες *Accuracy* (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **200x**.



Σχήμα 5.4: **resnet101_200x:** Accuracy=91.169, Loss=27.108, F1-score=94.724, Precision=94.496, Recall=95.023.

5.5.1.5 Resnet152

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **Resnet152**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.5, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<u>Resnet152</u>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	92.259%(0.0144)	21.963%(0.0311)	94.549%(0.0144)	95.845%(0.008)	92.439%(0.0197)
100x	93.423%(0.0143)	23.423%(0.0525)	95.323%(0.0125)	94.920%(0.0213)	96.713%(0.0132)
200x	88.844%(0.0489)	30.095%(0.0359)	84.506%(0.0766)	83.712%(0.0788)	85.571%(0.0852)
400x	90.412%(0.0076)	29.348%(0.0256)	92.684%(0.0085)	93.333%(0.0124)	93.505%(0.0132)

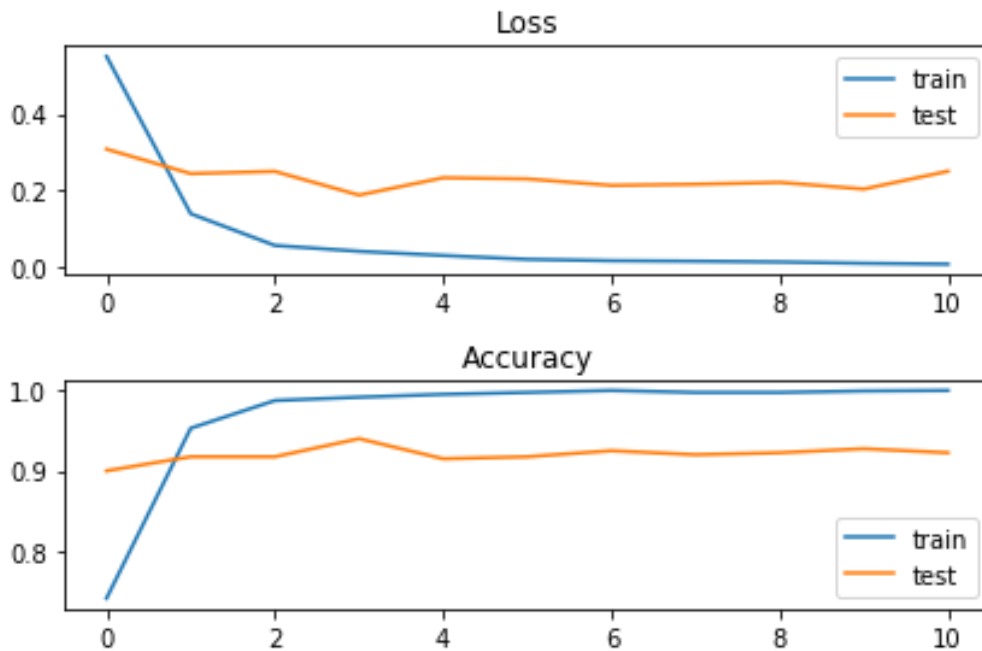
Πίνακας 5.5: Μετρικές αποτελεσμάτων του μοντέλου **Resnet152**, για εστίαση **40x**, **100x**, **200x**, **400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.5, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**93.423%%** (για 100x), *Loss*=**21.963%%** (για 40x), *F1*=**95.323%** (για 100x), *Precision*=**95.845%** (για 40x), *Recall*=**96.713%** (για 100x).

Στο Σχήμα 5.5, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες

Accuracy (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **100x**.



Σχήμα 5.5: Resnet152_100x: Accuracy=91.000, Loss=32.731, F1-score=93.175, Precision=91.447, Recall=95.056

5.5.1.6 MobileNet

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **Mobilenet**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.6, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

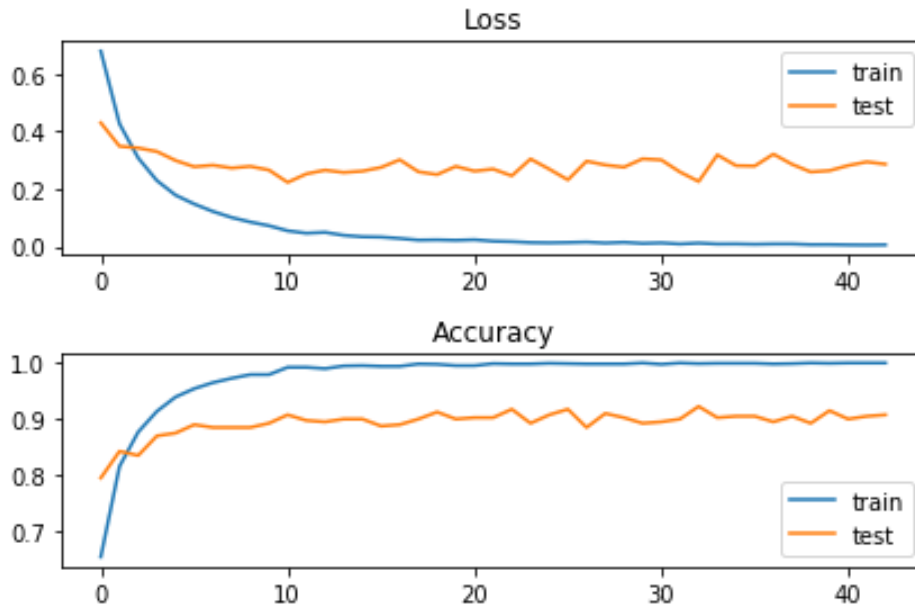
<i>MobileNet</i>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	89.167%(0.019)	27.804%(0.0524)	91.610%(0.017)	91.765%(0.024)	92.091%(0.0105)
100x	91.150%(0.0023)	30.863%(0.0103)	93.640%(0.0086)	90.547%(0.0089)	96.760%(0.0081)
200x	88.861%(0.0022)	35.413%(0.0374)	92.867%(0.0076)	91.419%(0.0166)	95.1810%(0.0077)
400x	87.072%(0.0023)	34.129%(0.0285)	90.852%(0.0040)	90.529%(0.0081)	91.420%(0.0100)

Πίνακας 5.6: Μετρικές αποτελεσμάτων του μοντέλου **Mobilenet**, για εστίαση **40x,100x,200x,400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.6, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- Accuracy=**91.150%** (για 100x), Loss=**27.804%** (για 40x), F1=**90.852%** (για 400x), Precision=**90.529%** (για 400x), Recall=**96.760%** (για 100x).

Στο Σχήμα 5.6, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες Accuracy (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση 100x.



Σχήμα 5.6: MobileNet_100x: Accuracy=91.500, Loss=32.120, F1-score=93.951, Precision=90.528, Recall=97.760.

5.5.1.7 MobileNetV2

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **MobilenetV2**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.7, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

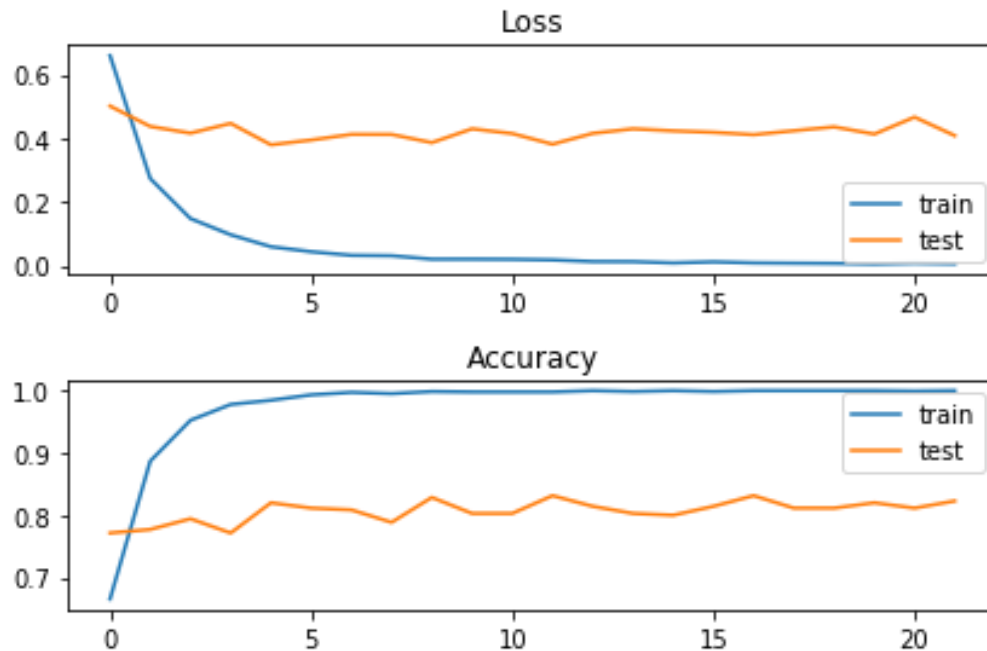
<i>MobileNetV2</i>	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	82.604%(0.0062)	49.635%(0.016)	87.379%(0.0076)	84.116%(0.0098)	92.169%(0.0129)
100x	87.255%(0.0055)	41.163%(0.028)	91.965%(0.0060)	91.345%(0.0209)	93.129%(0.0101)
200x	84.340%(0.0129)	43.983%(0.047)	85.816%(0.0734)	83.638%(0.0699)	86.025%(0.0671)
400x	82.631%(0.0069)	44.585%(0.0123)	87.176%(0.0124)	86.225%(0.0107)	89.176%(0.0284)

Πίνακας 5.7: Μετρικές αποτελεσμάτων του μοντέλου **MobilenetV2**, για εστίαση 40x, 100x, 200x, 400x.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.7, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- Accuracy=**87.255%** (για 100x), Loss=**41.163%** (για 100x), F1=**91.965%** (για 100x), Precision=**91.345%** (για 100x), Recall=**93.129%** (για 100x).

Στο Σχήμα 5.7, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες Accuracy (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση **400x**.



Σχήμα 5.7: MobileNetV2_400x: Accuracy=81.429, Loss=46.727, F1-score=86.148, Precision=85.510, Recall= 86.986.

5.5.1.8 DenseNet121

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **DenseNet121**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.8, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

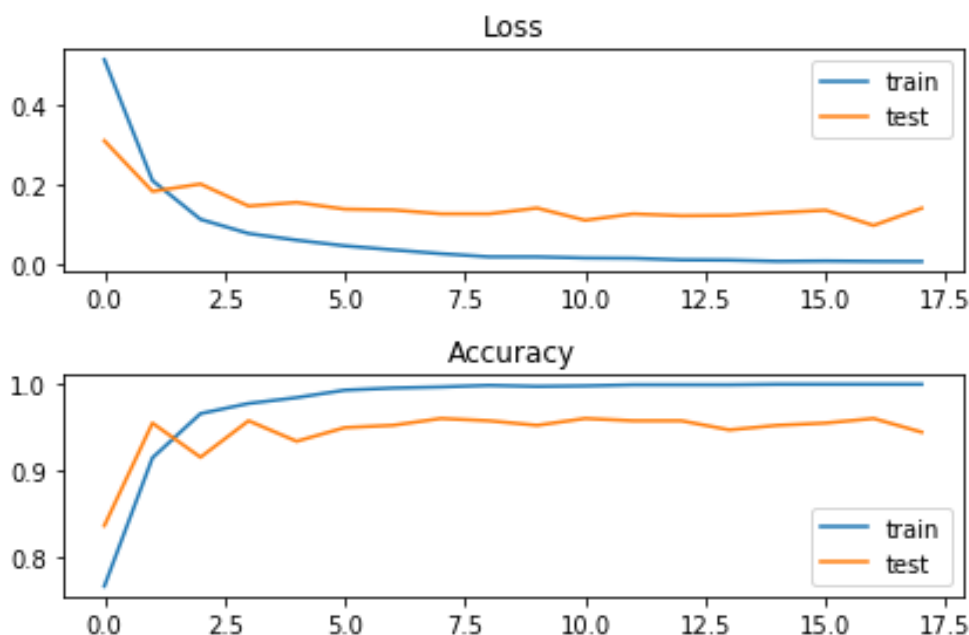
<i>DenseNet</i> <i>121</i>	Accuracy mean(std)	loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	94.833%(0.0036)	14.676%(0.0175)	95.812%(0.0063)	97.471%(0.0087)	94.985%(0.0075)
100x	93.845%(0.0089)	19.770%(0.0175)	95.566%(0.0092)	95.811%(0.0091)	95.954%(0.0110)
200x	91.056%(0.0068)	23.898%(0.0391)	89.264%(0.0701)	87.685%(0.0778)	90.894%(0.0718)
400x	90.143%(0.0137)	27.674%(0.0293)	92.066%(0.0108)	92.793%(0.0081)	91.973%(0.0115)

Πίνακας 5.8: Μετρικές αποτελεσμάτων του μοντέλου **DenseNet121**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.8, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- Accuracy=**94.833%** (για 40x), Loss=**14.676%** (για 40x), F1=**95.812%** (για 40x), Precision=**97.471%** (για 40x), Recall=**95.954%** (για 100x).

Στο Σχήμα 5.8, δείχνονται σχηματικές απεικονίσεις δύο γραφικών παραστάσεων: καμπύλες απώλειας (**Loss vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), και καμπύλες Accuracy (**Accuracy vs. Epoch**), στην εκπαίδευση (train) και τον έλεγχο (test), για εστίαση 40x.



Σχήμα 5.8: DenseNet121_40x: Accuracy=95.251, Loss=12.582f1-score=96.55, Precision=98.845, Recall= 94.449.

5.5.2 Πειραματικά Αποτελέσματα Δικτύων, με χρήση ZCA

Η πιο πάνω πειραματική διαδικασία, κατά την οποία δοκιμάστηκαν οκτώ αρχιτεκτονικές ΣΝΔ, επαναλήφθηκε έχοντας προσθέσει στάδιο προεπεξεργασίας ZCA σε κάθε ένα από τα πειράματα.

Συγκεκριμένα, δοκιμάστηκαν και τα οκτώ δίκτυα, έχοντας προσθέσει στάδιο προεπεξεργασίας ZCA σε κάθε ένα από αυτά. Τα αποτελέσματα αυτών των πειραμάτων, δηλαδή οι μετρήσεις των μέσων τιμών των μετρικών, Accuracy, Precision, Recall, F1, Loss, και της τυπικής απόκλισης (standard deviation-std) των τιμών αυτών, παρατίθενται σε πίνακες. Οι μετρήσεις αυτές αναφέρονται στις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), που προαναφέραμε.

5.5.2.1 VGG16, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **VGG16**, με χρήση **ZCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.9, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

VGG16 με ZCA	Accuracy mean(std)	loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	89.391%(0.0080)	25.170%(0.0355)	91.740%(0.0151)	88.350%(0.0087)	95.611%(0.0332)
100x	89.192%(0.0133)	29.210%(0.0178)	90.903%(0.0078)	89.115%(0.0099)	94.742%(0.0139)
200x	89.149%(0.0084)	33.581%(0.0335)	93.236%(0.0096)	92.200%(0.0160)	95.675%(0.0046)
400x	89.313%(0.0127)	28.012%(0.0347)	89.485%(0.0039)	89.372%(0.0047)	91.184%(0.0163)

Πίνακας 5.9: Μετρικές αποτελεσμάτων του μοντέλου *VGG16*, με χρήση *ZCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.9, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**89.391%%** (για 40x), *Loss*=**25.170%%** (για 40x), *F1*=**93.236%** (για 200x), *Precision*=**92.200%** (για 200x), *Recall*=**95.675%** (για 200x).

5.5.2.2 VGG19 με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *VGG19*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.10, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

VGG19 με ZCA	Accuracy mean(std)	loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	89.988%(0.0061)	30.268%(0.0233)	91.951%(0.0082)	90.863%(0.0103)	93.024%(0.0100)
100x	90.374%(0.0095)	32.624%(0.0132)	92.481%(0.0098)	92.150%(0.0116)	92.444%(0.0155)
200x	88.745%(0.0234)	32.453%(0.0567)	90.029%(0.0051)	88.787%(0.0076)	91.564%(0.0184)
400x	88.675%(0.0101)	29.987%(0.0429)	91.765%(0.0054)	90.456%(0.0072)	91.345%(0.0113)

Πίνακας 5.10: Μετρικές αποτελεσμάτων του μοντέλου *VGG19*, με χρήση *ZCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.10, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**90.374%(0.0095)** (για 100x), *Loss*=**29.987%(0.0429)** (για 400x), *F1*=**92.481%(0.0098)** (για 100x), *Precision*=**92.150%(0.0116)** (για 100x), *Recall*=**93.024%(0.0100)** (για 40x).

5.5.2.3 Resnet50, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet50*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.11, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών

(Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>Resnet50</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	90.743%(0.0093)	30.702%(0.0297)	91.452%(0.0121)	91.457%(0.0098)	92.011%(0.0178)
100x	91.065%(0.0093)	27.992%(0.0198)	92.477%(0.0145)	90.876%(0.0091)	92.774%(0.0112)
200x	90.132%(0.0074)	28.500%(0.0201)	91.453%(0.0334)	90.772%(0.0123)	92.032%(0.0198)
400x	90.654%(0.0104)	25.972%(0.0177)	91.690%(0.0076)	94.005%(0.0119)	92.867%(0.0277)

Πίνακας 5.11: Μετρικές αποτελεσμάτων του μοντέλου *Resnet50*, με χρήση **ZCA**, για εστίαση **40x, 100x, 200x, 400x**

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.11, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**91.065%(0.0093)** (για 100x), *Loss*=**25.972%(0.0177)** (για 400x), *F1*=**92.477%(0.0145)** (για 100x), *Precision*=**94.005%(0.0119)** (για 400x), *Recall*=**92.867%(0.0277)** (για 400x).

5.5.2.4 *Resnet101*, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet101*, με χρήση **ZCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.12, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>Resnet101</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	91.277%(0.0096)	20.774%(0.0165)	93.323%(0.0141)	95.443%(0.0093)	92.551%(0.0223)
100x	92.774%(0.0107)	22.491%(0.0173)	95.011%(0.0132)	92.734%(0.0096)	92.867%(0.0277)
200x	91.222%(0.0044)	25.675%(0.0081)	93.767%(0.0258)	93.345%(0.0143)	94.282%(0.0176)
400x	90.077%(0.0084)	30.765%(0.0111)	91.051%(0.0049)	93.519%(0.0087)	90.156%(0.0235)

Πίνακας 5.12: Μετρικές αποτελεσμάτων του μοντέλου *Resnet101*, με χρήση **ZCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.12, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**92.774%(0.0107)** (για 100x), *Loss*=**20.774%(0.0165)** (για 40x), *F1*=**95.011%(0.0132)** (για 100x), *Precision*=**95.443%(0.0093)** (για 40x), *Recall*=**94.282%(0.0176)** (για 200x).

5.5.2.5 Resnet152, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet152*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.13, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>Resnet152</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	92.544%(0.0122)	21.712%(0.0153)	94.077%(0.0233)	92.699%(0.0347)	93.841%(0.0086)
100x	94.111%(0.0046)	22.998%(0.0256)	94.776%(0.0049)	93.519%(0.0087)	90.156%(0.0235)
200x	90.156%(0.0194)	28.749%(0.0221)	90.777%(0.0076)	89.898%(0.0069)	90.546%(0.0078)
400x	90.867%(0.0170)	28.991%(0.0111)	93.756%(0.0143)	92.576%(0.0156)	93.654%(0.0091)

Πίνακας 5.13: Μετρικές αποτελεσμάτων του μοντέλου *Resnet152*, με χρήση *ZCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.13, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**94.111%(0.0046)** (για 100x), *Loss*=**21.712%(0.0153)** (για 40x), *F1*=**94.776%(0.0049)** (για 100x), *Precision*=**93.519%(0.0087)** (για 100x), *Recall*=**93.841%(0.0086)** (για 40x).

5.5.2.6 Mobilenet, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Mobilenet*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.14, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>Mobilenet</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	90.544%(0.0084)	25.892%(0.0252)	93.744%(0.0169)	92.541%(0.056)	91.784%(0.0076)
100x	90.442%(0.087)	32.435%(0.0239)	92.769%(0.0046)	93.665%(0.0128)	92.434%(0.0342)
200x	88.143%(0.0156)	33.665%(0.0166)	93.776%(0.0112)	90.991%(0.0058)	91.666%(0.0128)
400x	88.663 %(0.0066)	32.343%(0.0222)	90.786%(0.0097)	92.667%(0.0039)	90.119%(0.0174)

Πίνακας 5.14: Μετρικές αποτελεσμάτων του μοντέλου *Mobilenet*, με χρήση *ZCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.14, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=90.544\%(0.0084)$ (για 40x), $Loss=25.892\%(0.0252)$ (για 40x), $F1=93.776\%(0.0112)$ (για 200x), $Precision=93.665\%(0.0128)$ (για 100x), $Recall=92.434\%(0.0342)$ (για 100x).

5.5.2.7 MobilenetV2, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Mobilenet*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.15, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>MobilenetV2</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	87.756%(0.0049)	34.444%(0.0083)	90.661%(0.0091)	92.443%(0.0159)	90.665%(0.0145)
100x	88.978%(0.0096)	31.434%(0.0199)	90.003%(0.0086)	91.576%(0.0126)	92.749%(0.0185)
200x	86.441%(0.0065)	32.545%(0.0083)	91.778%(0.0156)	90.719%(0.0134)	91.786%(0.0075)
400x	86.443%(0.0179)	33.334%(0.0123)	90.483%(0.0096)	92.159%(0.0098)	89.768%(0.0129)

Πίνακας 5.15: Μετρικές αποτελεσμάτων του μοντέλου *MobilenetV2*, με χρήση *ZCA*, για εστίαση **40x**, **100x**, **200x**, **400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.15, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=88.978\%(0.0096)$ (για 100x), $Loss=31.434\%(0.0199)$ (για 100x), $F1=90.661\%(0.0091)$ (για 40x), $Precision=92.443\%(0.0159)$ (για 40x), $Recall=92.749\%(0.0185)$ (για 100x).

5.5.2.8 DenseNet121, με χρήση ZCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Densenet121*, με χρήση *ZCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.16, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (Accuracy, Precision, Recall, F1 και Loss), αλλά και της τυπικής απόκλισης (std), των τιμών αυτών.

<i>DenseNet121</i> με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	93.456%(0.0049)	15.449%(0.0084)	90.435%(0.0091)	92.557%(0.0132)	94.202%(0.0134)
100x	94.654%(0.0086)	18.476%(0.0106)	91.443%(0.0046)	93.223%(0.0087)	91.887%(0.0195)
200x	92.435%(0.0062)	22.449%(0.0097)	90.003%(0.0076)	93.232%(0.0207)	92.667%(0.0149)
400x	92.447%(0.0062)	23.223%(0.0153)	91.233%(0.0092)	92.323%(0.0156)	91.404%(0.0176)

Πίνακας 5.16: Μετρικές αποτελεσμάτων του μοντέλου *Densenet121*, με χρήση *ZCA*, για εστίαση **40x**, **100x**, **200x**, **400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.16, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=94.654\%(0.0086)$ (για 100x), $Loss=15.449\%(0.0084)$ (για 40x), $F1=91.443\%(0.0046)$ (για 100x), $Precision=93.232\%(0.0207)$ (για 200x), $Recall=94.202\%(0.0134)$ (για 40x).

5.5.3 Πειραματικά Αποτελέσματα Δικτύων, με χρήση PCA

Κατά την τελευταία πειραματική διαδικασία, δοκιμάστηκαν οι οκτώ αρχιτεκτονικές ΣΝΔ, έχοντας προσθέσει στάδιο προεπεξεργασίας **PCA** σε κάθε ένα από τα πειράματα.

Συγκεκριμένα, δοκιμάστηκαν και τα οκτώ δίκτυα, έχοντας προσθέσει στάδιο προεπεξεργασίας PCA σε κάθε ένα από αυτά. Τα αποτελέσματα αυτών των πειραμάτων, παραθέτονται σε πίνακες. Οι μετρήσεις αυτές αναφέρονται στις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), που προαναφέραμε.

5.5.3.1 VGG16, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **VGG16**, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.17, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

VGG16 με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	89.445%(0.0076)	33.754%(0.0079)	89.765%(0.0105)	88.116%(0.0145)	91.190%(0.0280)
100x	89.854%(0.0128)	30.005%(0.0363)	92.786%(0.0067)	91.786%(0.0234)	93.545%(0.0145)
200x	90.567%(0.0123)	29.965%(0.0174)	90.567%(0.0158)	89.886%(0.0098)	91.765%(0.0212)
400x	90.113%(0.0195)	32.635%(0.0164)	88.765%(0.0176)	84.116%(0.0098)	92.169%(0.0129)

Πίνακας 5.17: Μετρικές αποτελεσμάτων του μοντέλου **Vgg16**, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.17, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=90.567\%(0.0123)$ (για 200x), $Loss=29.965\%(0.0174)$ (για 200x), $F1=92.786\%(0.0067)$ (για 100x), $Precision=91.786\%(0.0234)$ (για 100x), $Recall=93.545\%(0.0145)$ (για 100x).

5.5.3.2 VGG19, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **VGG16**, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.18, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών

(*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>VGG19</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	87.442%(0.0143)	33.453%(0.0245)	90.376%(0.0134)	89.553%(0.0087)	91.881%(0.0210)
100x	88.453%(0.0062)	34.567%(0.056)	88.564%(0.0076)	89.548%(0.0256)	90.455%(0.0134)
200x	89.234%(0.0058)	33.6453%(0.018)	89.3548%(0.012)	91.458%(0.0081)	91.453%(0.0102)
400x	88.657%(0.0156)	34.675%(0.0201)	89.345%(0.0082)	90.234%(0.0073)	91.567%(0.080)

Πίνακας 5.18: Μετρικές αποτελεσμάτων του μοντέλου *Vgg16*, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.18, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**89.234%(0.0058)** (για 200x), *Loss*=**33.453%(0.0245)** (για 40x), *F1*=**90.376%(0.0134)** (για 40x), *Precision*=**91.458%(0.0081)** (για 200x), *Recall*=**91.881%(0.0210)** (για 40x).

5.5.3.3 *Resnet50*, με χρήση **PCA**

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet50*, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.19, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>Resnet50</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	91.564%(0.0132)	32.635%(0.0049)	89.756%(0.0078)	91.435%(0.0119)	93.434%(0.0223)
100x	92.101%(0.0049)	29.734%(0.0201)	89.379%(0.0076)	84.116%(0.0098)	92.169%(0.0129)
200x	92.112%(0.0049)	24.567%(0.0152)	90.221%(0.0124)	92.516%(0.0088)	93.161%(0.0111)
400x	92.333%(0.0058)	24.567%(0.0171)	90.344%(0.0066)	93.221%(0.0123)	91.564%(0.0243)

Πίνακας 5.19: Μετρικές αποτελεσμάτων του μοντέλου *Resnet50*, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.19, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**92.333%(0.0058)** (για 400x), *Loss*=**24.567%(0.0171)** (για 400x), *F1*=**90.344%(0.0066)** (για 400x), *Precision*=**93.221%(0.0123)** (για 400x), *Recall*=**93.434%(0.0223)** (για 40x).

5.5.3.4 Resnet101, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet101*, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.20, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>Resnet101</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	92.3454%(0.0049)	25.498%(0.0234)	91.343%(0.0076)	92.421%(0.0143)	92.897%(0.0098)
100x	91.123%(0.0113)	26.765%(0.0161)	91.434%(0.0124)	92.332%(0.0345)	94.223%(0.0087)
200x	91.765%(0.0045)	26.223%(0.0132)	92.265%(0.0103)	91.763%(0.0076)	93.431%(0.0156)
400x	90.854%(0.0056)	23.675%(0.0104)	90.567%(0.0108)	92.435%(0.0083)	93.765%(0.0112)

Πίνακας 5.20: Μετρικές αποτελεσμάτων του μοντέλου *Resnet101*, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.20, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**92.3454%(0.0049)** (για 40x), *Loss*=**23.675%(0.0104)** (για 400x), *F1*=**92.265%(0.0103)** (για 200x), *Precision*=**92.435%(0.0083)** (για 400x), *Recall*=**94.223%(0.0087)** (για 100x).

5.5.3.5 Resnet152, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Resnet152*, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.21, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>Resnet152</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	92.675%(0.0052)	26.657%(0.0179)	91.559%(0.0045)	92.545%(0.0078)	93.723%(0.0223)
100x	93.457%(0.0156)	25.113%(0.0146)	92.478%(0.0098)	94.445%(0.0352)	93.337%(0.0087)
200x	92.779%(0.0056)	24.675%(0.0201)	93.465%(0.0092)	94.354%(0.0146)	92.556%(0.0176)
400x	92.354%(0.0048)	23.221%(0.0134)	92.484%(0.0078)	92.332%(0.0154)	93.689%(0.0181)

Πίνακας 5.21: Μετρικές αποτελεσμάτων του μοντέλου *Resnet152*, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.21, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**93.457%(0.0156)** (για 100x), *Loss*=**23.221%(0.0134)** (για 400x), *F1*=**93.465%(0.0092)** (για 200x), *Precision*=**94.445%(0.0352)** (για 100x), *Recall*=**93.723%(0.0223)** (για 40x).

5.5.3.6 Mobilenet, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *Mobilenet*, με χρήση *PCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.22, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>Mobilenet</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	87.768%(0.0066)	26.777%(0.0056)	90.300%(0.0143)	92.227%(0.0134)	95.763%(0.0043)
100x	88.673%(0.0078)	28.534%(0.0202)	91.445%(0.0187)	92.253%(0.0056)	92.756%(0.0043)
200x	89.332%(0.0075)	27.419%(0.0194)	89.776%(0.0086)	92.444%(0.0107)	91.998%(0.0167)
400x	87.678%(0.0046)	28.334%(0.0161)	91.234%(0.0064)	92.657%(0.0112)	91.111%(0.0174)

Πίνακας 5.22: Μετρικές αποτελεσμάτων του μοντέλου *Mobilenet*, με χρήση *PCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.22, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- *Accuracy*=**89.332%(0.0075)** (για 200x), *Loss*=**26.777%(0.0056)** (για 40x), *F1*=**91.445%(0.0187)** (για 100x), *Precision*=**92.657%(0.0112)** (για 400x), *Recall*=**95.763%(0.0043)** (για 40x).

5.5.3.7 MobilenetV2, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου *MobilenetV2*, με χρήση *PCA*, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.23, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>MobilenetV2</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	85.654%(0.0071)	31.059%(0.0184)	88.991%(0.0143)	85.992%(0.0073)	90.434%(0.0141)
100x	87.659%(0.0084)	34.221%(0.0205)	89.814%(0.0044)	88.786%(0.0154)	90.443%(0.0147)
200x	86.343%(0.0066)	35.785%(0.0143)	91.653%(0.0101)	90.002%(0.0075)	91.334%(0.0206)
400x	85.112%(0.0045)	36.422%(0.0194)	88.379%(0.0073)	91.224%(0.0145)	90.254%(0.0070)

Πίνακας 5.23: Μετρικές αποτελεσμάτων του μοντέλου *MobilenetV2*, με χρήση *PCA*, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.23, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=87.659\%(0.0084)$ (για 100x), $Loss=31.059\%(0.0184)$ (για 40x), $F1=91.653\%(0.0101)$ (για 200x), $Precision=91.224\%(0.0145)$ (για 400x), $Recall=91.334\%(0.0206)$ (για 200x).

5.5.3.8 Densenet121, με χρήση PCA

Τα συγκεντρωτικά πειραματικά αποτελέσματα, του μοντέλου **Densenet121**, με χρήση **PCA**, παρουσιάζονται στη συνέχεια. Στον Πίνακα 5.24, δείχνονται τα αποτελέσματα, για τις τέσσερις κατηγορίες εστίασης (40x, 100x, 200x, 400x), των μέσων τιμών (mean) των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1* και *Loss*), αλλά και της τυπικής απόκλισης (*std*), των τιμών αυτών.

<i>Densenet121</i> με PCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
40x	93.773%(0.0033)	18.002%(0.0202)	94.334 % (0.008)	91.552%(0.0093)	94.992%(0.0254)
100x	92.764%(0.0086)	20.445%(0.0079)	92.442%(0.0106)	93.112%(0.0104)	93.005%(0.0206)
200x	91.087%(0.0073)	20.332%(0.0105)	91.665%(0.0059)	92.475%(0.0176)	93.612%(0.0196)
400x	89.745%(0.0111)	26.754%(0.0148)	92.564%(0.0082)	91.554%(0.0156)	92.774%(0.0043)

Πίνακας 5.24: Μετρικές αποτελεσμάτων του μοντέλου **Densenet121**, με χρήση **PCA**, για εστίαση **40x, 100x, 200x, 400x**.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.24, διαπιστώνουμε ότι οι μέσες (mean) τιμές των μετρικών δεν συναρτώνται με την αύξηση των τιμών της εστίασης των εικόνων (40x, 100x, 200x, 400x). Συγκεκριμένα, οι καλλίτερη τιμές είναι:

- $Accuracy=93.773\%(0.0033)$ (για 40x), $Loss=18.002\%(0.0202)$ (για 40x), $F1=94.334\%(0.008)$ (για 40x), $Precision=93.112\%(0.0104)$ (για 100x), $Recall=94.992\%(0.0254)$ (για 40x).

5.5.4 Συγκεντρωτικοί Πίνακες

Προκειμένου να συγκριθούν τα επιμέρους δίκτυα και οι τεχνικές προεπεξεργασίας, δημιουργήσαμε συγκεντρωτικούς πίνακες, στους οποίους τοποθετούμε (α) τα καλύτερα αποτελέσματα ανά μετρική κατηγορία, ανεξάρτητα από την εστίαση, για κάθε δίκτυο, (β) τα καλύτερα δίκτυα, ανά μετρική κατηγορία, για κάθε κατηγορία εστίασης.

5.5.4.1 Συγκεντρωτικοί Πίνακες των δικτύων, χωρίς προεπεξεργασία

Από τους Πίνακες 5.1-5.8, συγκεντρώσαμε τα καλλίτερα αποτελέσματα των μέσων τιμών των μετρικών (*Accuracy*, *Precision*, *Recall*, *F1*, *Loss*), για κάθε ένα από τα οκτώ δίκτυων που δοκιμάσαμε, ανεξάρτητα από την κατηγορία εστίασης των εικόνων (40x, 100x, 200x, 400x). Τα αποτελέσματα αυτά φαίνονται στον Πίνακα 5.25, (σε παρένθεση δίνεται η εστίαση στην οποία αναφέρεται το αποτέλεσμα).

Καλύτερες επιδόσεις	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
<u>Vgg16</u> (Πίνακας 5.1)	88.286%(0.0067) (100x)	41.645%(0.0120) (400x)	92.631%(0.0055) (200x)	90.676%(0.0170) (400x)	96.101%(0.0105) (100x)
<u>Vgg19</u> (Πίνακας 5.2)	86.708%(0.0064) (40x)	35.044%(0.0462) (400x)	90.711%(0.0062) (100x)	89.890%(0.0187) (400x)	94.061%(0.0050) (100x)
<u>ResNet50</u> (Πίνακας 5.3)	90.335%(0.0092) (100x)	26.370%(0.0158) (400x)	92.892%(0.0102) (100x)	93.437%(0.0064) (400x)	92.998%(0.0233) (100x)
<u>ResNet101</u> (Πίνακας 5.4)	92.655%(0.0075) (100x)	20.917%(0.0254) (40x)	94.821%(0.0099) (100x)	95.641%(0.0066) (40x)	95.622%(0.0092) (100x)
<u>ResNet152</u> (Πίνακας 5.5)	93.423%(0.0143) (100x)	21.963%(0.0311) (40x)	95.323%(0.0125) (100x)	95.845%(0.008) (40x)	96.713%(0.0132) (100x)
<u>MobileNet</u> (Πίνακας 5.6)	91.150%(0.0023) (100x)	27.804%(0.0524) (40x)	90.852%(0.0040) (400x)	90.529%(0.0081) (400x)	96.760%(0.0081) (100x)
<u>MobileNetV2</u> (Πίνακας 5.7)	87.255%(0.0055) (100x)	41.163%(0.028) (100x)	91.965%(0.0060) (100x)	91.345%(0.0209) (100x)	93.129%(0.0101) (100x)
<u>DenseNet121</u> (Πίνακας 5.8)	94.833%(0.0036) (40x)	14.676%(0.0175) (40x)	95.812%(0.0063) (40x)	97.471%(0.0087) (40x)	95.954%(0.0110) (100x)

Πίνακας 5.25: Τα οκτώ δίκτυα με τις καλλίτερες επιδόσεις τους, ανεξάρτητα από την εστίαση των εικόνων. Σε παρένθεση δίνεται η εστίαση στην οποία αναφέρεται το αποτέλεσμα.

Παρατηρώντας τον Πίνακα 5.25, διαπιστώνουμε ότι το δίκτυο **DenseNet121** παρουσιάζει την καλλίτερη απόδοση, με βάση τις περισσότερες (4 από τις 5) μετρικές (για 40x):

- *Accuracy*=**94.833%(0.0036)**, *Loss*=**14.676%(0.0175)**, *F1*=**95.812%(0.0063)**, *Precision*=**97.471%(0.0087)** και *Recall*=**95.954%(0.0110)** (κατά 0.806% λιγότερο από την καλλίτερη απόδοση που έχει το δίκτυο **MobileNet**)

Τέλος, συγκεντρώνοντας τα καλλίτερα δίκτυα, ανά μετρική κατηγορία, για κάθε κατηγορία εστίασης, δημιουργούμε τον Πίνακα 5.26. Ο πίνακας αυτός μας δίνει τη δυνατότητα μιας διαφορετικής ανάγνωσης από τον Πίνακα 5.25.

Καλύτερα Δίκτυα	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
<u>40x</u>	Densenet121 94.833%(0.0036)	Densenet121 14.676%(0.0175)	Densenet121 95.812%(0.0063)	Densenet121 97.471%(0.0087)	Densenet121 94.985%(0.0075)
<u>100x</u>	Densenet121 93.845%(0.0089)	Densenet121 19.770%(0.0175)	Densenet121 95.566%(0.0092)	Densenet121 95.811%(0.0091)	MobileNet 96.760%(0.0081)
<u>200x</u>	Densenet121 91.056%(0.0068)	Densenet121 22.449%(0.0097)	MobileNet 92.867%(0.0076)	Resnet101 93.728%(0.0086)	MobileNet 95.1810%(0.0077)
<u>400x</u>	Resnet152 90.412%(0.0076)	Resnet50 26.370%(0.0158)	Resnet152 92.684%(0.0085)	Resnet152 93.333%(0.0124)	Resnet152 93.505%(0.0132)

Πίνακας 5.26: Τα δίκτυα με τις καλλίτερες επιδόσεις, σε κάθε κατηγορία εστίασης, για κάθε μετρική κατηγορία.

Στον Πίνακα 5.26, βλέπουμε ότι για εστίαση 40x και 100x, το δίκτυο *DenseNet121* έχει την καλλίτερη απόδοση, για όλες τις μετρικές κατηγορίες. Για εστίαση 200x, σε Accuracy και Loss, υπερέχει το *DenseNet121*, ενώ σε F1-score και Recall το *Mobilenet*, και σε Precision το *Resnet152*. Τέλος, για εστίαση 400x, υπερέχει *Resnet152*, για όλες τις μετρικές. Τέλος διαπιστώνεται ότι, εκτός της μετρικής Recall, η μικρότερη εστίαση (40x) δίνει τα καλύτερα αποτελέσματα και αυξανόμενης της εστίασης τα αποτελέσματα γίνονται λιγότερο καλά.

Συμπερασματικά, θα μπορούσαμε, με ρακέτα ρίσκα, να γενικεύσουμε και να ισχυριστούμε ότι αυξάνοντας την εστίαση δεν βελτιώνεται κατ' ανάγκη και η απόδοση.

5.5.4.2 Συγκεντρωτικοί Πίνακες, δίκτυα με χρήση ZCA

Από τους Πίνακες 5.9-5.16, συγκεντρώσαμε τα καλλίτερα δίκτυα, ανά μετρική κατηγορία (*Accuracy*, *Precision*, *Recall*, *F1*, *Loss*), για κάθε μία από τις τέσσερις κατηγορία εστίασης (40x, 100x, 200x, 400x) και δημιουργήσαμε τον Πίνακα 5.27.

Καλύτερα Δίκτυα, με ZCA	Accuracy mean(std)	Loss mean(std)	F1-score mean(std)	Precision mean(std)	Recall mean(std)
<u>40x</u>	Densenet121 93.456%(0.0049)	Densenet121 15.449%(0.0084)	Densenet121 94.077%(0.0233)	Resnet101 95.443%(0.0093)	VGG16 95.611%(0.0332)
<u>100x</u>	Densenet121 94.654%(0.0086)	Densenet121 18.476%(0.0106)	VGG16 95.011%(0.0132)	Densenet121 93.223%(0.0087)	VGG16 94.742%(0.0139)
<u>200x</u>	Densenet121 92.435%(0.0062)	Densenet121 22.449%(0.0097)	Mobilenet 93.776%(0.0112)	Resnet101 93.345%(0.0143)	VGG16 95.675%(0.0046)
<u>400x</u>	Densenet121 92.447%(0.0062)	Densenet121 23.223%(0.0153)	Resnet152 93.756%(0.0143)	Resnet101 93.519%(0.0087)	Resnet152 93.654%(0.0091)

Πίνακας 5.27: Τα δίκτυα με τις καλλίτερες επιδόσεις, σε κάθε κατηγορία εστίασης, για κάθε μετρική κατηγορία.

Παρατηρώντας τον Πίνακα 5.27, διαπιστώνουμε ότι το δίκτυο DenseNet121 έχει και πάλι την καλύτερη επίδοση, στην μετρική Accuracy και Loss και για τις τέσσερις εστιάσεις. Τα δίκτυα Resnet δείχνουν επίσης έχει καλή συμπεριφορά, ενώ και τα δίκτυο παρουσιάζει (παραδόξως) Vgg19 καλά αποτελέσματα με τη μετρική Recall.

5.5.4.3 Συγκεντρωτικοί Πίνακες, δίκτυα με χρήση PCA

Από τους Πίνακες 5.17-5.24, συγκεντρώσαμε τα καλλίτερα δίκτυα, ανά μετρική κατηγορία (*Accuracy*, *Precision*, *Recall*, *F1*, *Loss*), για κάθε μία από τις τέσσερις κατηγορία εστίασης (40x, 100x, 200x, 400x) και δημιουργήσαμε τον Πίνακα 5.27. \

Τέλος, συγκεντρώνοντας τα καλύτερα δίκτυα, ανά μετρική κατηγορία, για κάθε κατηγορία εστίασης, δημιουργούμε τον Πίνακα 5.30. Ο πίνακας αυτός μας δίνει τη δυνατότητα μιας διαφορετικής ανάγνωσης από τον προηγούμενο πίνακα.

Καλύτερα Δίκτυα, με PCA	<Accuracy> (std)	<Loss> (std)	<F1-score> (std)	<Precision> (std)	<Recall> (std)
40x	Densenet121 93.773%(0.0033)	Densenet121 18.002%(0.0202)	Densenet121 94.334 %(0.008)	Resnet152 92.545%(0.0078)	Mobilenet 95.763%(0.0043)
100x	Resnet152 93.457%(0.0156)	Densenet121 20.445%(0.0079)	VGG16 92.786%(0.0067)	Resnet152 94.445%(0.0352)	Resnet101 94.223%(0.0087)
200x	Resnet152 92.779%(0.0056)	Densenet121 20.332%(0.0105)	Resnet152 93.465%(0.0092)	Resnet101 94.354%(0.0146)	Densenet121 93.612%(0.0196)
400x	Resnet152 92.354%(0.0048)	Resnet152 23.221%(0.0134)	Densenet121 92.564%(0.0082)	Resnet50 93.221%(0.0123)	Resnet101 93.765%(0.0112)

Πίνακας 5.28: Τα δίκτυα με τις καλλίτερες επιδόσεις, σε κάθε κατηγορία εστίασης, για κάθε μετρική κατηγορία.

Παρατηρώντας τον Πίνακα 5.28, διαπιστώνουμε ότι το δίκτυο DenseNet121 με τα δίκτυα Resnet, μονοπωλούν τα καλύτερα αποτελέσματα.

Συμπεράσματα

Φτάνοντας στο τελευταίο κεφάλαιο της παρούσας διπλωματικής εργασίας θα μπορούσαμε να ισχυριστούμε πως οι στόχοι που αρχικά είχαν τεθεί, σε ικανοποιητικό βαθμό έχουν επιτευχθεί.

Οι στόχοι αυτοί μπορούν να συνοψιστούν στα ακόλουθα:

Η βαθιά κατανόηση των Συνελικτικών Νευρωνικών Δικτύων και της μεγάλης ποικιλίας των διαφορετικών μοντέλων αυτών, που τα τελευταία δέκα χρόνια σχεδόν μονοπωλούν τα συστήματα ταξινόμησης προτύπων και ιδιαίτερα αυτών που αναφέρονται στην αναγνώριση εικόνας.

Η πειραματική διαδικασία: επιλέχθηκαν οκτώ διακεκριμένα δίκτυα, τα οποία υλοποιήθηκαν, εκπαιδεύτηκαν και εκλέχθηκαν, χρησιμοποιώντας την Ιστοπαθολογική Βάση Δεδομένων καρκίνου του μαστού (BreCaKHis). Να σημειωθεί ότι όλα τα δίκτυα που χρησιμοποιήσαμε ήταν προεκπαιδευμένα, στο μεγάλο όγκο των δεδομένων της ImageNet. Τέλος η σύγκριση μεταξύ των οκτώ δικτύων με βάση την απόδοσή τους σε χαρακτηριστικές μετρικές (*Accuracy*, *Precision*, *Recall*, *F1*, *Loss*), και τα αποτελέσματα σχολιάστηκαν. Ακόμα, θα ήθελα να συμπληρώσω εδώ την πολύ θετική εμπειρία που αποκόμισα από την εφαρμοσμένη πληροφορική, κάτι που στο βαθμό αυτό δεν το είχα μέχρι τώρα αντιληφθεί. Στη συνεισφορά της τεχνολογίας στην αντιμετώπιση της μάστιγας του καρκίνου του μαστού αλλά και γενικότερα σε κρίσιμες επιστημονικές διαγνώσεις.

Κατά την πειραματική διαδικασία χρησιμοποιήσαμε οκτώ γνωστά Συνελικτικά Νευρωνικά Δίκτυα: *Vgg16*, *Vgg19*, *ResNet50*, *ResNet101*, *ResNet152*, *MobileNet*, *MobileNetV2*, *DenseNet121*. Τα οκτώ (8) αυτά δίκτυα, εκπαιδεύτηκαν και ελέγχτηκαν σε τρεις (3) πειραματικές διαδικασίες (χωρίς προεπεξεργασία, με προεπεξεργασία ZCA, και με προεπεξεργασία PCA). Όλα τα πειράματα επαναλήφθηκαν τέσσερις (4) φορές, για τις διαφορετικές κατηγορίες των δεδομένων εστίασης (40x, 100x, 200x, 400x). Έτσι, το σύνολο των πειραμάτων που διεξήχθησαν ήταν **96 πειράματα** (=8(δίκτυα) x 3(προεπεξεργασία) x 4(εστιάσεις)). Από τα πειράματα αυτά προέκυψαν 96 πίνακες, συν 12 συγκεντρωτικοί πίνακες και 96 σχήματα. Ο μεγάλος αριθμός των πινάκων και των σχημάτων, μας οδήγησε, για πρακτικούς λόγους, στην συμπίεση αυτών, και τελικά περιοριστήκαμε σε 28 πίνακες (Πίνακας 5.1-5.28) και 8 διπλά σχήματα (Σχήμα 5.1-5.8).

Συγκρίνοντας τώρα τα αποτελέσματα που περιγράφονται στους πίνακες και ιδιαίτερα στους τρεις συγκεντρωτικούς πίνακες, δηλαδή τον Πίνακα 5.26, που αναφέρεται σε δίκτυα χωρίς την διαδικασία προεπεξεργασίας, τον Πίνακα 5.27 που αναφέρεται σε δίκτυα με προεπεξεργασία ZCA, και τον Πίνακα 5.28 που αναφέρεται σε δίκτυα με προεπεξεργασία PCA, μπορούμε να οδηγηθούμε στα ακόλουθα συμπεράσματα:

Σε ότι αφορά τα δίκτυα χωρίς προεπεξεργασία (Πίνακες 5.1-5.8 και 5.25, 5.26), τα αποτελέσματα ήταν μάλλον αναμενόμενα. Σαφή υπεροχή του δικτύου *Densenet121*, που αποτελεί εξέλιξη του *Resnet152*, ένα από τα πιο βαθιά νευρωνικά δίκτυα. Επίσης εξαιρετική (συγκριτικά) απόδοση είχε και το *Resnet152*. Συγκεκριμένα, το δίκτυο **DenseNet121** παρουσιάζει την καλλίτερη απόδοση, με βάση τις περισσότερες (4 από τις 5) μετρικές (για 40x): **Accuracy=94.833%(0.0036)**, **Loss=14.676%(0.0175)**, **F1=95.812%(0.0063)**, **Precision=97.471%(0.0087)**. Η μετρική **Recall=95.954%(0.0110)** και δίνει, κατά 0.806%, λιγότερο καλό αποτέλεσμα, (ο δίκτυο *MobileNet* δίνει **Recall=96.760%(0.0081)**).

Ενδιαφέρον ακόμα είχε το γεγονός ότι εικόνες με χαμηλή εστίαση έδωσαν καλύτερα αποτελέσματα από αυτές με υψηλή. Ενδιαφέρον ακόμα παρουσιάζει το γεγονός δεδομένα από εικόνες χαμηλής εστίασης (40x) έδωσαν καλύτερα αποτελέσματα από αυτές με υψηλή.

Στα πειράματα, όπου στα δίκτυα είχε προηγηθεί προεπεξεργασία με τις δύο προσεγγίσεις (ZCA και PCA), τα αποτελέσματα (Πίνακες 5.9-5.24 και 5.27, 5.28), δεν θα μπορούμε να τα χαρακτηρίσουμε αναμενόμενα. Διαπιστώσαμε βεβαίως βελτιώσεις σχεδόν παντού, ωστόσο αυτές οι βελτιώσεις δεν ακολούθησαν μια συγκεκριμένη λογική. Η εξήγησή που μπορούμε να δώσουμε, για τα αποτελέσματα αυτά, είναι ο περιορισμένος όγκος των δεδομένων εκπαίδευσης και ελέγχου, που δεν επέτρεψαν να αναδειχθεί η σημαντική συνεισφορά του σταδίου της προεπεξεργασίας.

Βιβλιογραφία

- [1] Anand, P., et al., Cancer is a preventable disease that requires major lifestyle changes. Pharm Res, 2008. 25(9): p. 2097-116.
- [2] American Cancer Society, A.C., Global Cancer Facts & Figures 2nd Edition. 2011.
- [3] Organization, W.H., GLOBOCAN(IARC) Section of Cancer Information. 2008.
- [4] Organization, W.H., Cancer. Fact sheet N°297, 2012.
- [5] Ν. Δημητρόπουλος, Α. Κεραμόπουλος. «Η Μαστογραφία στην κλινική πράξη». Επιστημονικές Εκδόσεις «Γρ. Παρισιάνος», Αθήνα, 2000.
- [6] Χατζημπούγιας Ι. Ο Μαστός. Στο: Χατζημπούγιας Ι (εκδ). Στοιχεία ανατομικής του Ανθρώπου. Θεσσαλονίκη: GM Designs; 2000:267-276.
- [7] <https://doi.org/10.1016/j.dib.2019.104863>
- [8] Stamatakis E.A., Ricketts I.W., Cairns A.Y. & Walker C. (March 27, 1996). “Detecting abnormalities on mammograms by bilateral comparison”, IEE Colloquium on Digital Mammography, London.
- [9] Kowal M., Filipczuk P., Obuchowicz A., Korbicz J., Monczak R. (October 1, 2013). “Computer aided diagnosis of breast cancer based on fine needle biopsy microscopic images”, Computers in Biology and Medicine, Volume: 43, Issue: 10, Pages: 1563-1572.
- [10] K.N. Plataniotis, A.N.V., Color Image Processing and Applications. Springer-Verlag, 2000.
- [11] Russ, J.C., The image-processing handbook. 5th ed. 2007, Boca Raton: CRC/Taylor and Francis.
- [12] <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/> (<http://www.prevencaoedigaose.com.br>).
- [13] Turing, A.: 1950, Computing machinery and intelligence, Mind LIX (59) (236), 433–460.
- [14] Rich E. & K. Knight “Artificial Intelligence”, 2nd Edition, 1991
- [15] Russell, S. and Norvig, P.: 2002, Artificial Intelligence: A Modern Approach, Prentice Hall, Upper Saddle River, NJ.
- [16] J. G. Nicholls, A. R. Martin, P. A. Fuchs, D. A. Brown, M. E. Diamond, and D. A. Weisblat, From Neuron to Brain, Fifth Edition.
- [17] Geoffrey E. Hinton et al. A Fast Learning Algorithm for Deep Belief Nets,
- [18] I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. Adaptive computation and machine learning, MIT Press, 2016.
- [19] D. Svozil, V. Kvasnicka, and J. Pospichal, “Introduction to multi-layer feed-forward neural networks,” Chemometrics and Intelligent Laboratory Systems, vol. 39, no. 1, pp. 43–62, Nov. 1997.
- [20] Srivastava, Nitish, et al.”Dropout: a simple way to prevent neural networks from overfitting”, JMLR 2014
- [21] S. Ioffe and C. Szegedy,”Batch normalization: Accelerating deep network training by reducing internal covariate shift”, in Int. Conf. on Machin. Learning (ICML), 2015.
- [22] Adit Deshpande. The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3). The-9-Deep-Learning-Papers-You-Need-To-Know-About.html. <https://adeshpande3.github.io/adeshpande3.github.io/>
- [23] Yann LeCun, Leon Bottou, Yoshua Bengio and Patrick Haffner, “Gradient-Based Learning Applied to Document Recognition”, In Proceedings of the IEEE, Volume: 86, Issue 11, Nov 1998, Pages: 2278 -2324
- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks”, In Proceedings of the 25th International

-
- Conference on Neural Information Processing Systems -Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105. 2012.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv: 1409.1556, 2014.
 - [26] W. Yu et al., "Visualizing and Comparing AlexNet and VGG using Deconvolutional Layers," *Icml*, vol. 48, pp. 1–7, 2016.
 - [27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions", In *Cvpr*, 2015
 - [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In *CVPR*, pages 770–778, 2016.
 - [29] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks", arXiv preprint arXiv: 1709.01507, 2017.
 - [30] D. M. W. D. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," *J. Mach. Learn. Technol.*, vol. 2, no. December, p. 24, 2007.
 - [31] Y. Sasaki, "The truth of the F-measure." pp. 1–5, 2007.
 - [32] V. Garcia, R. A. Mollineda, and J. S. Sanchez, "Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions," pp. 73–87, 2015.
 - [33] <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>
 - [34] Spanhol, F., Oliveira, L. S., Petitjean, C., Heutte, L., A Dataset for Breast Cancer Histopathological Image Classification, *IEEE Transactions on Biomedical Engineering (TBME)*, 63(7):1455-1462, 2016. [pdf]
 - [35] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," arXiv, vol. 1702.05747, no. 1995, pp. 1–34, 2017.
 - [36] <https://link.springer.com/content/pdf/bbm%3A978-1-4842-3453-2%2F1.pdf>
 - [37] Large Scale Visual Recognition Challenge 2015 (ILSVRC2015). <http://imagenet.org/challenges/LSVRC/2015/>.
 - [38] Tu, E., Kasabov, N., Othman, M., Li, Y., Worner, S., Yang, J., Jia, Z.: Neucube(st) for spatio-temporal data predictive modelling with a case study on ecological data. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 638–645 (July 2014)
 - [39] Hirvensalo, Mika, Quantum Computing. Springer, 2003.
 - [40] Kaburlasos VG, Papadakis SE, and Papakostas GA, "A lattice computing extension of the FAM neural classifier for human facial expression recognition", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1526-1538, 2013.
 - [41] Vikhar, P. A. (2016). "Evolutionary algorithms: A critical review and its future prospects". *Proceedings of the 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*. Jalgaon. pp. 261–265. doi:10.1109/ICGTSPICC.2016.7955308.
 - [42] Ashlock, D. (2006), *Evolutionary Computation for Modeling and Optimization*, Springer, ISBN0-387-22196-4.
 - [43] Arthur, Samuel (1959-03-03). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. 3 (3): 210–229. CiteSeerX 10.1.1.368.2254. doi:10.1147/rd.33.0210.
 - [44] Livia Faes, Siegfried K Wagner, Dun Jack Fu, Xiaoxuan Liu, Edward Korot, Joseph R Ledsam, et al. Automated deep learning design for medical image classification by healthcare professionals with no coding experience: a feasibility study, «[The Lancet Digital Health](#)». VOLUME 1, ISSUE 5, E232-E242, SEPTEMBER 01, 2019
 - [45] [assessed by the Nottingham Grading System].

-
- [46] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto και Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. CoRR, abs/1704.04861, 2017.
 - [47] J. Shlens. (2005, December) A tutorial on principal component analysis.
 - [48] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” Apr. 2009.
 - [49] G. Huang et al, “Densely Connected Convolutional Networks”, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.