

Ανίχνευση Ιογενούς ή Βακτηριακής Πνευμονίας σε Εικόνες Ακτινογραφίας Θώρακος

Κύρος Γεώργιος
Διεθνές Πανεπιστήμιο Ελλάδος
Θεσσαλονίκη, Ελλάδα
ait112020@ait.teithe.gr

Μανώλας Ιωάννης
Διεθνές Πανεπιστήμιο Ελλάδος
Θεσσαλονίκη, Ελλάδα
ait132020@ait.teithe.gr

Περίληψη— Η πνευμονία προκαλεί το θάνατο περίπου 700.000 παιδιών κάθε χρόνο και επηρεάζει το 7% του παγκόσμιου πληθυσμού. Οι ακτινογραφίες θώρακος χρησιμοποιούνται κυρίως για τη διάγνωση αυτής της ασθένειας. Ωστόσο, ακόμη και για έναν εκπαιδευμένο ακτινολόγο, είναι δύσκολο να εξεταστεί η ακτινογραφία θώρακος. Υπάρχει ανάγκη βελτίωσης της ακρίβειας της διάγνωσης. Σε αυτή την εργασία, παρουσιάζουμε δύο υλοποιήσεις μοντέλων βαθιάς μάθησης για την ανίχνευση της ιογενούς και βακτηριακής πνευμονίας σε ψηφιακές εικόνες ακτινογραφίας θώρακος, το οποίο θα μπορούσε να βοηθήσει τους ακτινολόγους στη διαδικασία λήψης αποφάσεων. Χρησιμοποιούνται τα βέλτιστα προεκπαιδευμένα μοντέλα βαθιάς μάθησης VGG16 και EfficientNetB0. Αυτή η προσέγγιση είναι μια εποπτευόμενη προσέγγιση μάθησης στην οποία το δίκτυο προβλέπει το αποτέλεσμα με βάση την ποιότητα των δεδομένων που χρησιμοποιούνται. Η μεταφορά μάθησης (transfer learning) χρησιμοποιείται για να βελτιώσει τα μοντέλα βαθιάς μάθησης για να αποκτήσει υψηλότερη εκπαίδευση και ακρίβεια. Χρησιμοποιούνται τεχνικές μερικής αύξησης δεδομένων (data augmentation) για την αύξηση του εκπαιδευτικού προγράμματος με ισορροπημένο τρόπο. Η εργασία αναφέρεται στον εκπαιδευτικό διαγωνισμό στην πλατφόρμα του kaggle με τίτλο “Detect pneumonia from bacteria or virus using chest X-ray images”.

Λέξεις κλειδιά— Πνευμονία; Μηχανική μάθηση; CNN; VGG16; EfficientNetB0

I. ΕΙΣΑΓΩΓΗ

Η πνευμονία είναι μια οξεία αναπνευστική λοίμωξη που προσβάλλει τους πνεύμονες. Είναι μια θανατηφόρα ασθένεια στην οποία οι πνεύμονες γεμίζουν με πύον ή και άλλο υγρό. Υπάρχουν κυρίως δύο τύποι πνευμονίας: βακτηριακή και ιογενής. Γενικά, παρατηρείται ότι η βακτηριακή πνευμονία προκαλεί οξύτερα συμπτώματα. Η πιο σημαντική διαφορά μεταξύ βακτηριακής και ιογενούς πνευμονίας είναι η θεραπεία. Η θεραπεία της βακτηριακής πνευμονίας γίνεται με αντιβιοτική θεραπεία, ενώ η ιογενής πνευμονία συνήθως μπορεί να αυτοϊαθεί. Η πνευμονία κατατάσσεται όγδοη στη λίστα των 10 κορυφαίων αιτιών θανάτου στις Ηνωμένες Πολιτείες. Η ασθένεια συχνά παραβλέπεται και δεν θεραπεύεται έως ότου φτάσει σε μοιραίο σημείο, ειδικά στην περίπτωση ηλικιωμένων ασθενών. Είναι η μοναδική μεγαλύτερη αιτία θανάτου σε παιδιά (ειδικά κάτω των πέντε ετών) παγκοσμίως [1]. Ωστόσο, η ταχύτερη διάγνωση της ιογενούς ή βακτηριακής πνευμονίας και η επακόλουθη εφαρμογή της σωστής θεραπείας μπορούν να βοηθήσουν

σημαντικά στην πρόληψη της επιδείνωσης της κατάστασης ενός ασθενούς, η οποία τελικά οδηγεί σε θάνατο. Οι ακτινογραφίες θώρακος είναι επί του παρόντος η καλύτερη μέθοδος για τη διάγνωση της πνευμονίας. Οι ακτινογραφίες της πνευμονίας δεν είναι πολύ σαφείς. Επιπλέον, οι ακτινογραφίες βακτηριακής και ιογενούς πνευμονίας μερικές φορές ταξινομούνται κατά λάθος από τους ειδικούς, γεγονός που οδηγεί σε λανθασμένη φαρμακευτική αγωγή στους ασθενείς και ως εκ τούτου επιδεινώνει την κατάσταση των ασθενών [2]. Επομένως, υπάρχει επείγουσα ανάγκη για έρευνα και ανάπτυξη σχετικά με τη διάγνωση με τη βοήθεια της τεχνολογίας και της μηχανικής μάθησης, ώστε να μειωθεί η θνησιμότητα που σχετίζεται με την πνευμονία, ειδικά στα παιδιά. Επί του παρόντος, πολλές βιοϊατρικές επιπλοκές (π.χ. ανίχνευση όγκου του εγκεφάλου, ανίχνευση καρκίνου του μαστού, κ.λπ.) χρησιμοποιούν λύσεις με βάση την Τεχνητή Νοημοσύνη (ΑΙ). Μεταξύ των τεχνικών βαθιάς μάθησης, τα συνελκτικά νευρωνικά δίκτυα (CNN) έχουν επιδείξει μεγάλες δυνατότητες στην ταξινόμηση των εικόνων και ως εκ τούτου υιοθετήθηκαν ευρέως από την ερευνητική κοινότητα. Οι τεχνικές Βαθιάς Μάθησης σε ακτινογραφίες θώρακος, κερδίζουν δημοτικότητα καθώς μπορούν εύκολα να χρησιμοποιηθούν με τεχνικές απεικόνισης χαμηλού κόστους και υπάρχει πληθώρα δεδομένων που διατίθενται για την εκπαίδευση διαφορετικών μοντέλων μηχανικής μάθησης. Αρκετές ερευνητικές ομάδες ανέφεραν τη χρήση αλγορίθμων βαθιάς μηχανικής μάθησης στην ανίχνευση πνευμονίας. Ωστόσο, μόνο ένα άρθρο ανέφερε την ταξινόμηση της βακτηριακής και ιογενούς πνευμονίας. Σε αυτή την εργασία προσπαθήσαμε να ταξινομήσουμε ακτινογραφίες θώρακος σε τρεις κατηγορίες με την χρήση μοντέλων βαθιάς μηχανικής μάθησης. Οι κατηγορίες αυτές ήταν οι εξής:

- 1) *Υγιής (Normal)*
- 2) *Βακτηριακή Πνευμονία (Bacterial Pneumonia)*
- 3) *Ιογενής Πνευμονία (Viral Pneumonia)*

Χρησιμοποιήσαμε 2 προ εκπαιδευμένα μοντέλα βαθιάς μηχανικής μάθησης τα οποία είναι το VGG16 και το EfficientNetB0. Παρακάτω θα αναλύσουμε τα βασικά μέρη και λειτουργίες των συνελκτικών νευρωνικών δικτύων (CNN) του τομέα της βαθιάς μηχανικής μάθησης καθώς και θα παρουσιάσουμε τα αποτελέσματα της έρευνάς μας.



Εικόνα 1. Δείγματα ακτινογραφιών θώρακος

II. ΒΑΣΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΤΗΝ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (ANN)

A. Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning) είναι ένα πεδίο στην επιστήμη των υπολογιστών όπου τα υπάρχοντα δεδομένα χρησιμοποιούνται για την πρόβλεψη σε μελλοντικά δεδομένα. Στη μηχανική μάθηση (ML), ένα υπολογιστικό σύστημα μαθαίνει πρώτα να εκτελεί μια εργασία μελετώντας ένα εκπαιδευτικό σύνολο παραδειγμάτων. Στη συνέχεια, ο υπολογιστής εκτελεί την ίδια εργασία με δεδομένα που δεν είχε συναντήσει στο παρελθόν με στόχο την αναγνώριση ή ταξινόμηση αυτών. Συνδέεται στενά με τους τομείς της αναγνώρισης προτύπων, των υπολογιστικών στατιστικών και της τεχνητής νοημοσύνης. Η μηχανική μάθηση είναι σημαντική σε τομείς όπως η αναγνώριση προσώπου, το φιλτράρισμα των ανεπιθύμητων μηνυμάτων, η ταξινόμηση φωτογραφιών και άλλους τομείς όπου δεν είναι εφικτό, ή ακόμη και δυνατό, η σύνταξη αλγορίθμων για την εκτέλεση μιας εργασίας. Τα σύγχρονα συστήματα χρησιμοποιούν τεχνικές μηχανικής μάθησης με πολύ μεγάλη επιτυχία. Η μηχανική μάθηση χωρίζεται σε τρεις κατηγορίες, αυτές είναι: a) Εποπτευόμενη μάθηση (Supervised Learning), b) Μη εποπτευόμενη Μάθηση (Unsupervised Learning), c) Ενισχυτική Μάθηση (Reinforcement Learning) [3].

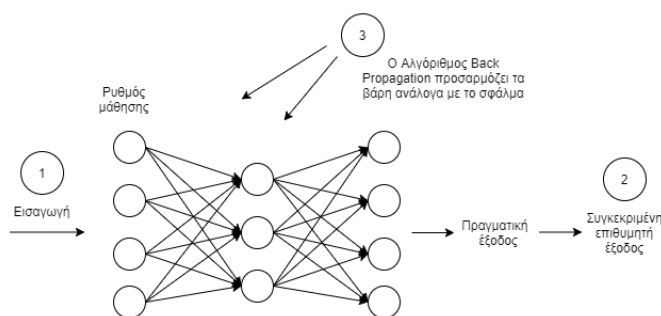
B. Τεχνητά Νευρωνικά Δίκτυα (ANNs)

Για την εκτέλεση εργασιών υπό επίβλεψη, χωρίς επίβλεψη, ημι-εποπτευόμενη ή μάθησης με ενίσχυση, έχουν αναπτυχθεί αρκετά πλαίσια. Μεταξύ αυτών των πλαισίων, τα τεχνητά νευρωνικά δίκτυα (ANNs) είναι αναμφισβήτητα τα πιο σημαντικά, καθώς μπορούν να μιμηθούν την ανθρώπινη νοημοσύνη [4].

Τα τεχνητά νευρωνικά δίκτυα (ANN) ή νευρωνικά δίκτυα (NN) αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Αυτά τα στοιχεία εμπνέονται από τα βιολογικά νευρικά συστήματα. Όπως στη φύση, η λειτουργία δικτύου καθορίζεται σε μεγάλο βαθμό από τις συνδέσεις μεταξύ των στοιχείων. Μπορούμε να εκπαιδεύσουμε ένα νευρωνικό δίκτυο για να εκτελέσει μια συγκεκριμένη λειτουργία προσαρμόζοντας τις τιμές των συνδέσεων (βάρη) μεταξύ των στοιχείων. Συνήθως τα νευρωνικά δίκτυα προσαρμόζονται ή εκπαιδεύονται, έτσι ώστε μια συγκεκριμένη είσοδος να οδηγεί σε συγκεκριμένη έξοδο στόχου. Εκεί, το δίκτυο προσαρμόζεται, με βάση τη σύγκριση της εξόδου και του στόχου, έως ότου η έξοδος δικτύου ταιριάζει με τον στόχο. Συνήθως χρησιμοποιούνται πολλά τέτοια ζεύγη εισόδου / στόχου, σε αυτήν την εποπτευόμενη μάθηση για την εκπαίδευση ενός δικτύου. Οι τεχνητοί νευρώνες είναι πολύ πιο απλούστεροι από έναν βιολογικό νευρώνα. Η Εικόνα 2 δείχνει

τα βασικά στοιχεία των τεχνητών νευρώνων. Η ικανότητα των νευρωνικών δικτύων να ανακαλύπτουν μη γραμμικές σχέσεις στα δεδομένα εισόδου τα καθιστά ιδανικά για μοντελοποίηση μη γραμμικών δυναμικών συστημάτων όπως το χρηματιστήριο κ.α [5]. Υπάρχουν διάφοροι τύποι ANN. Παρακάτω θα αναφερθούμε στους τύπους των ANNs που κάναμε χρήση στην εργασία μας :

- Deep Neural Networks (DNN): Όλα τα ANN που έχουν πολλαπλά επίπεδα είναι γνωστά ως DNN.
- Feedforward neural networks (FNN): Σε ένα νευρωνικό δίκτυο feedforward (FNN), κάθε νευρώνας έχει εισερχόμενες συνδέσεις μόνο από το προηγούμενο στρώμα και εξερχόμενες συνδέσεις μόνο στο επόμενο επίπεδο. Τα FNN μπορούν να χρησιμοποιηθούν για τον καθορισμό προηγμένων αρχιτεκτονικών όπως: α) μηχανές ακραίας μάθησης (ELM), β) συνελκτικά νευρωνικά δίκτυα (CNN), γ) νευρωνικά δίκτυα χρονικής καθυστέρησης (TDNNs), δ) αυτοκωδικοποιητές (autoencoders), ε) πιθανολογικά νευρωνικά δίκτυα (PNUs) και ε) λειτουργίες ακτινικής βάσης (RBF) [4].

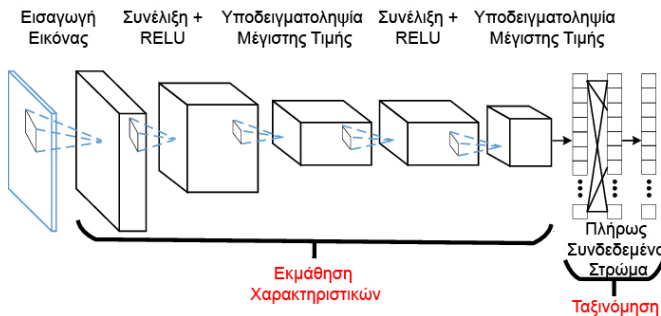


Εικόνα 2. Βασικό παράδειγμα Τεχνητού Νευρωνικού Δικτύου [5].

C. Συνελκτικά Νευρωνικά Δίκτυα (CNNs)

Το πρώτο βήμα στην ταξινόμηση είναι η επιλογή των κατάλληλων εικόνων. Η λειτουργία εξαγωγής χαρακτηριστικών από μια εικόνα είναι ένα θεμελιώδες βήμα στην επεξεργασία εικόνας. Τα εξαγόμενα χαρακτηριστικά παρουσιάζονται σε χώρο υψηλότερης διάστασης και παρέχουν δεδομένα εισόδου σε ένα τεχνητό νευρωνικό δίκτυο. Το αποτέλεσμα από το τεχνητό νευρωνικό δίκτυο συντάσσεται από τη μηχανή φορέα υποστήριξης (SVM) για το τελικό αποτέλεσμα ταξινόμησης. Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από τρία επίπεδα: το επίπεδο εισόδου, το κρυφό στρώμα και το επίπεδο εξόδου, όπου ο αριθμός των κόμβων εισόδου στο δίκτυο είναι μια μέτρηση του διανύσματος χαρακτηριστικών και ο αριθμός των κόμβων εξόδου είναι ίσος με τον αριθμό των κλάσεων. Για την επίλυση αυτών των προβλημάτων, για να βρεθούν σύνθετα χαρακτηριστικά από μια μεγάλη σειρά εικονοστοιχείων, χρησιμοποιείται ένα συνελκτικό νευρωνικό δίκτυο (CNN). Το συνελκτικό νευρωνικό δίκτυο είναι παρόμοιο με τον οπτικό φλοιό του εγκεφάλου. Στον οπτικό φλοιό υπάρχουν μικρά τμήματα κυττάρων νευρώνων που σχετίζονται με ορισμένα σημεία στο οπτικό πεδίο. Εάν εξετάσουμε το CNN με περισσότερες λεπτομέρειες, τότε αποτελείται από μια σειρά στρωμάτων. Οι δύο πρώτοι τύποι στρωμάτων (συνελκτικό,

υποδειγματοληψία), εναλλάσσονται μεταξύ τους και σχηματίζουν ένα φορέα χαρακτηριστικών εισόδου για ένα πολυεπίπεδο perceptron. Η τοπολογία του συνελκτικού νευρωνικού δικτύου φαίνεται στην εικόνα 3. Η εικόνα περνά μέσω εναλλαγής συνελκτικών, μη γραμμικών στρώματων και η έξοδος δημιουργείται χρησιμοποιώντας ένα πλήρως συνδεδεμένο στρώμα. Η έξοδος μπορεί να είναι η κλάση (0,1) ή η πιθανότητα της κλάσης που περιγράφει καλύτερα την εικόνα [6].



Εικόνα 3. Απλοποιημένο σχήμα ενός συνελκτικού νευρωνικού δικτύου (CNN) [10].

Όλα τα συστατικά και η δομή του απλού μοντέλου CNN που χρησιμοποιούνται περιγράφονται ως εξής [7]:

1) Συνελκτικό Στρώμα (Convolutional Layer)

Το Συνελκτικό Στρώμα ή Convolutional Layer αποτελεί το βασικό δομικό στοιχείο για ένα συνελκτικό νευρωνικό δίκτυο. Τα συνελκτικά στρώματα χρησιμοποιούν ένα φίλτρο σταθερού μεγέθους για να εξάγουν διάφορα χαρακτηριστικά. Η σάρωση της εικόνας γίνεται μετακινώντας το φίλτρο σύμφωνα με τα βήματα που καθορίστηκαν.

2) Κανονικοποίηση (Batch Normalization)

Η κανονικοποίηση ή Batch Normalization χρησιμοποιείται για τη βελτίωση του ρυθμού μάθησης του μοντέλου. Αυτό το επίπεδο ομαλοποιεί την είσοδο. Σε όλα τα μοντέλα, το Batch Normalization χρησιμοποιείται μετά από κάθε συνελκτικό στρώμα.

3) Επικαλυπτόμενη Υποδειγματοληψία (Pooling Layer)

Το pooling είναι η τεχνική που χρησιμοποιείται για την υποδειγματοληψία του χάρτη χαρακτηριστικών (feature map) που συλλέγεται από το συνελκτικό επίπεδο. Συνήθως χρησιμοποιούνται δύο τύποι συγκέντρωσης, δηλαδή η μέγιστη συγκέντρωση και η μέση συγκέντρωση. Στο απλό μοντέλο CNN, η μέγιστη συγκέντρωση χρησιμοποιείται μετά από κάθε συνελκτικό στρώμα με το φίλτρο συγκέντρωσης μεγέθους 2×2 .

4) Συνάρτηση Ενεργοποίησης (Activation)

Η συνάρτηση ενεργοποίησης χρησιμοποιείται στο τέλος ενός επιπέδου για τον μη γραμμικό μετασχηματισμό των εισόδων. Σήμερα, η συνήθης λειτουργία ενεργοποίησης σε ένα συνελκτικό νευρωνικό δίκτυο είναι η Rectified Linear Unit, δηλαδή το ReLU. Το ReLU χρησιμοποιείται στο τέλος κάθε στρώματος και η λειτουργία ενεργοποίησης σιγμοειδούς συνάρτησης (sigmoid) χρησιμοποιείται στο τελικό στρώμα με δύο κόμβους. Αν πρόκειται για πρόβλημα παραπάνω από 2 κλάσεων τότε χρησιμοποιείται η συνάρτηση ενεργοποίησης Softmax.

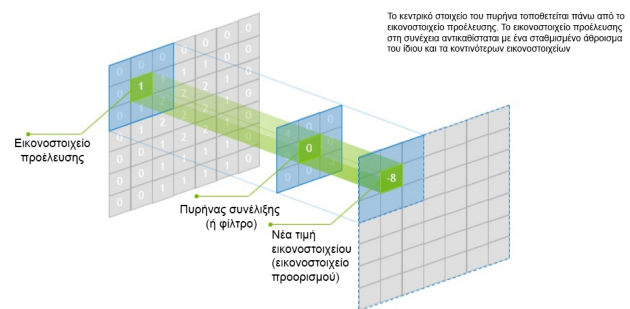
5) Αχρήστευση (Dropout)

Το Dropout είναι η τεχνική που χρησιμοποιείται για τη μείωση της υπερβολικής μάθησης του μοντέλου (overfitting). Χρησιμοποιώντας το dropout, ορισμένοι κόμβοι στο επίπεδο επιλέγονται τυχαία για να παραμείνουν ανενεργοί για ορισμένες περιπτώσεις. Με αυτόν τον τρόπο, εμποδίζει το μοντέλο να εξοικειωθεί πολύ με τα δεδομένα.

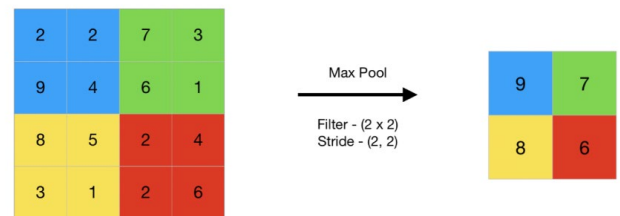
6) Πυκνό Στρώμα (Dense Layer)

Η έξοδος του συνελκτικού στρώματος ισοπεδώνεται περαιτέρω και δίνεται στο πυκνό στρώμα ως είσοδος. Ο ρόλος των συνελκτικών επιπέδων είναι η εξαγωγή των χαρακτηριστικών και η ταξινόμηση των εικόνων που γίνεται από τα πυκνά στρώματα.

Συνέλιξη (Convolution)



Εικόνα 4. Συνελκτικό Στρώμα (Convolutional Layer) [8].

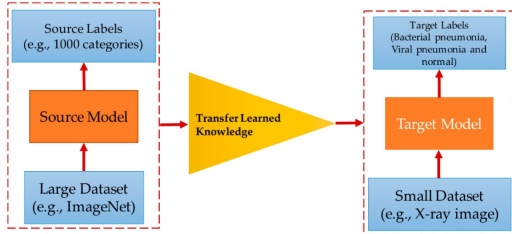


Εικόνα 5. Επικαλυπτόμενη Υποδειγματοληψία μέγιστης τιμής (Max Pooling) [8].

Τα συνελκτικά νευρωνικά δίκτυα (CNN's) αποτέλεσαν σημαντική ανακάλυψη για ανάλυση και ταξινόμηση βάσει εικόνας. Μπορούν να χρησιμοποιηθούν αποτελεσματικά για την ταξινόμηση εικόνας. Την τελευταία δεκαετία, υπήρξαν αξιοσημείωτα έργα στον τομέα της ιατρικής, ιδίως στην ιατρική διάγνωση. Η βαθιά μηχανική μάθηση και το CNN χρησιμοποιούνται ενεργά σε διάφορες ιατρικές έρευνες, ιδίως σε αυτές με δεδομένα εικόνας και ακτινογραφιών. Η μεταφορά μοντέλων μάθησης (transfer learning) όπως VGG16, InceptionV3, DenseNet κ.λπ., τα οποία εκπαιδεύονται χρησιμοποιώντας το σύνολο δεδομένων ImageNet, μπορούν να χρησιμοποιηθούν για τη δημιουργία ταχύτερων και αποτελεσματικότερων μοντέλων για την ταξινόμηση των εικόνων [7]. Παρακάτω θα παρουσιάσουμε κάποιες πληροφορίες για την μεταφορά γνώσης στην βαθιά μάθηση (transfer learning) καθώς και τα βασικά στοιχεία και την αρχιτεκτονική δύο συγκεκριμένων μοντέλων (CNN) που τα χρησιμοποιήσαμε και στην εργασία μας. Αυτά είναι το VGG16 και το EfficientNetB0.

III. ΜΕΤΑΦΟΡΑ ΓΝΩΣΗΣ ΣΤΗΝ ΒΑΘΙΑ ΜΑΘΗΣΗ (DEEP LEARNING TRANSFER)

Τα CNN συνήθως έχουν καλύτερη απόδοση σε ένα μεγαλύτερο σύνολο δεδομένων από ένα μικρότερο. Η μεταφορά μάθησης μπορεί να είναι χρήσιμη σε αυτές τις εφαρμογές του CNN όπου το σύνολο δεδομένων δεν είναι μεγάλο. Η έννοια της μεταφοράς μάθησης φαίνεται στην εικόνα 6, όπου το εκπαιδευμένο μοντέλο από μεγάλο σύνολο δεδομένων, όπως το ImageNet, μπορεί να χρησιμοποιηθεί για εφαρμογή με ένα συγκριτικά μικρότερο σύνολο δεδομένων [2].

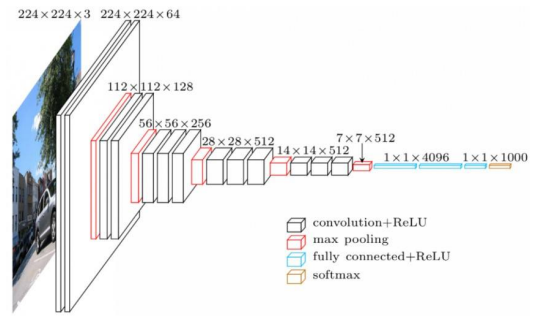


Εικόνα 6. Έννοια της μεταφοράς μάθησης [2].

Πρόσφατα, η μεταφορά μάθησης έχει χρησιμοποιηθεί με επιτυχία σε διάφορες εφαρμογές πολλών τομέων, όπως οι κατασκευές, η ιατρική και ο έλεγχος αποσκευών. Αυτό καταργεί την απαίτηση να διαθέτει κάποιος μεγάλο όγκο δεδομένων και επίσης μειώνει τη μακρά περίοδο εκπαίδευσης, όπως απαιτείται από τον αλγόριθμο βαθιάς μάθησης που αναπτύχθηκε από το μηδέν [2].

A. VGG16

Το VGG είναι ένα μοντέλο CNN που δημιουργήθηκε από το Πανεπιστήμιο της Οξφόρδης και αναπτύχθηκε από τους Simonyan και Zisserman στο πλαίσιο του διαγωνισμού ILSVRC 2014. Επιπλέον, η εκπαίδευση του VGG16 μοντέλου χρειάστηκε περίπου 2-3 εβδομάδες σε ένα υπολογιστικό σύστημα εξοπλισμένο με τέσσερις κάρτες γραφικών GPU NVIDIA Titan Black. Το VGG-16 σχεδιάστηκε αρχικά για ταξινόμηση φυσικής εικόνας μεγάλης κλίμακας. Το δίκτυο VGG-16 αποτελείται από 16 επίπεδα, όπου υπάρχουν ενσωματωμένα (13) συνελκτικά στρώματα, με φίλτρα 3×3 και 2×2 . Σε στρώματα max pooling στοιβάζονται μεταξύ αυτών των επιπέδων και εφαρμόζεται η λειτουργία ενεργοποίησης (activation) του relu. Στη συνέχεια, υπάρχουν τρία πλήρως συνδεδεμένα επίπεδα που περιέχουν τις περισσότερες από τις παραμέτρους του δικτύου. Τέλος χρησιμοποιείται, μια συνάρτηση softmax για να παράγει τις πιθανότητες για κάθε κλάση. Η εικόνα 7 απεικονίζει την ακριβή αρχιτεκτονική του VGG16. Σημειώστε ότι ένα σημαντικό μειονέκτημα του VGG-16 είναι η απαιτητική χρήση της μνήμης λόγω του τεράστιου αριθμού παραμέτρων (140M), η οποία οδηγεί σε επιβράδυνση της απόδοσης και της σπατάλης κατανάλωσης ενέργειας [11].



Εικόνα 7. Αρχιτεκτονική του VGG16 [11].

1) Αρχιτεκτονική ανά επίπεδο

a) *Πρώτο και δεύτερο στρώμα:* Η είσοδος για το AlexNet είναι μια εικόνα $224 \times 224 \times 3$ RGB που περνά μέσα από το πρώτο και το δεύτερο συνελκτικό επίπεδο με 64 χάρτες χαρακτηριστικών και φίλτρα μεγέθους 3×3 και το ίδιο επίπεδο υποδειγματοληψίας με βήμα 14. Οι διαστάσεις της εικόνας αλλάζουν σε $224 \times 224 \times 64$. Στη συνέχεια το VGG16 εφαρμόζει στρώμα υποδειγματοληψία μέγιστης τιμής ή απλή υποδειγματοληψία με φίλτρο 3×3 και βήμα δύο. Οι προκύπτουσες διαστάσεις εικόνας θα μειωθούν σε $112 \times 112 \times 64$.

b) *Τρίτο και τέταρτο στρώμα:* Στη συνέχεια, υπάρχουν δύο συνελκτικά στρώματα με 128 χάρτες χαρακτηριστικών που έχουν μέγεθος 3×3 και βήμα 1. Επίσης, υπάρχει και πάλι στρώμα υποδειγματοληψίας μέγιστης τιμής με μέγεθος φίλτρου 3×3 και βήμα 2. Αυτό το επίπεδο είναι ίδιο με το προηγούμενο στρώμα υποδειγματοληψίας εκτός από το ότι διαθέτει 128 χάρτες χαρακτηριστικών, οπότε η έξοδος θα μειωθεί σε $56 \times 56 \times 128$.

c) *Πέμπτο και έκτο στρώμα:* Το πέμπτο και το έκτο στρώμα αποτελούνται από συνελκτικά στρώματα με μέγεθος φίλτρου 3×3 και βήμα 1. Και τα δύο χρησιμοποίησαν 256 χάρτες χαρακτηριστικών. Τα δύο συνελκτικά στρώματα ακολουθούνται από ένα στρώμα υποδειγματοληψίας μέγιστης τιμής με μέγεθος φίλτρου 3×3 , βήμα 2 και έχουν 256 χάρτες χαρακτηριστικών.

d) *Εβδομο έως δωδέκατο στρώμα:* Στη συνέχεια υπάρχουν δύο σύνολα 3 συνελκτικών επιπέδων ακολουθούμενα από ένα στρώμα υποδειγματοληψίας μέγιστης τιμής. Όλα τα συνελκτικά στρώματα έχουν 512 φίλτρα μεγέθους 3×3 και βήμα 1. Το τελικό μέγεθος θα μειωθεί σε $7 \times 7 \times 512$.

e) *Δέκατο τρίτο στρώμα:* Σε αυτό το σημείο χρησιμοποιείται η τεχνική της ισοπέδωσης (flatten) στην έξοδο των συνελκτικών στρωμάτων μέσω ενός πλήρως συνδεδεμένου επιπέδου με 25088 χάρτες χαρακτηριστικών μεγέθους 1×1 .

f) *Δέκατο τέταρτο και δέκατο πέμπτο επίπεδο:* Στη συνέχεια υπάρχουν πάλι δύο πλήρως συνδεδεμένα στρώματα με 4096 κόμβους.

g) *Επίπεδο εξόδου:* Τέλος, υπάρχει ένα επίπεδο εξόδου softmax με 1000 πιθανές τιμές [9].

2) Παρατηρήσεις σχετικά με το μοντέλο VGGNet

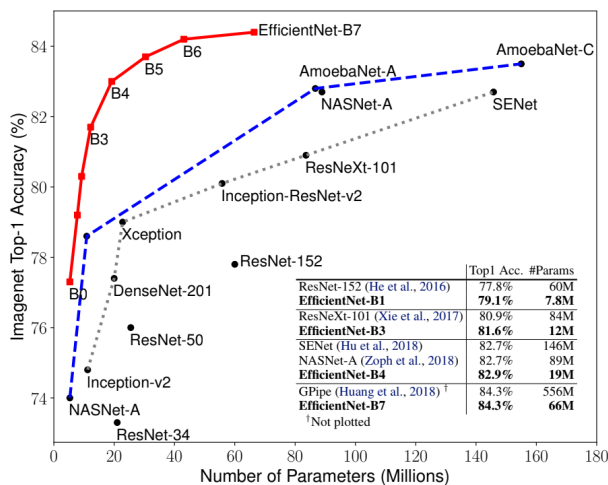
- Χρησιμοποιεί 3×3 φίλτρα στο αρχικό επίπεδο.
- Αποδίδει καλά στην ταξινόμηση των εικόνων.
- Αυτό το μοντέλο έχει δημιουργηθεί μαζί με την εργαλειοθήκη Caffe.
- Χρησιμοποιεί το jittering κλίμακα ως αύξηση δεδομένων, μέθοδος κατά τη στιγμή της εκπαίδευσης.
- Το στρώμα ReLU χρησιμοποιείται μετά από κάθε συνεκτικό στρώμα.
- Εκπαιδεύτηκε σε 4 κάρτες γραφικών Nvidia Titan Black GPUs για 2-3 εβδομάδες [12].

B. EfficientNetB0

Το EfficientNet, παρουσιάστηκε για πρώτη φορά από τους Tan και Le, το 2019 [13] είναι ένα από τα πιο αποτελεσματικά μοντέλα και κατέχει την μεγαλύτερη προηγμένη ακρίβεια τόσο στο σύνολο imagenet όσο και σε κοινές εργασίες μεταφοράς μάθησης για ταξινόμηση εικόνας.

Παρουσιάζοντας έναν ευρετικό τρόπο κλιμάκωσης του μοντέλου, το EfficientNet παρέχει μια οικογένεια μοντέλων (B0 έως B7) που αντιπροσωπεύει έναν καλό συνδυασμό αποτελεσματικότητας και ακρίβειας. Μια τέτοια ευρετική κλιμάκωση (σύνθετη κλιμάκωση, λεπτομέρειες βλέπε Tan and Le, 2019 [13]) επιτρέπει στο βασικό μοντέλο αποδοτικότητας (B0) να ξεπερνά τα μοντέλα σε κάθε κλίμακα, αποφεύγοντας παράλληλα την εκτεταμένη αναζήτηση πλέγματος υπερ παραμέτρων.

Στην εικόνα 8 μπορούμε να δούμε την απόδοση της οικογένειας EfficientNets στο σύνολο δεδομένων ImageNet.



Εικόνα 8. Συγκριτικό γράφημα απόδοσης μοντέλων στο imagenet [13].

1) Αρχιτεκτονική ανά επίπεδο

Θα εμβαθύνουμε στην αρχιτεκτονική του EfficientNet-B0 καθώς χρησιμοποιήθηκε και στην εργασία. Το B0 είναι μια αρχιτεκτονική μεγέθους με 11M εκπαιδευτικές παραμέτρους. Ανέπτυξαν αυτό το μοντέλο χρησιμοποιώντας μια πολυ-

αντικειμενική αναζήτηση νευρωνικής αρχιτεκτονικής που βελτιστοποιεί τόσο την ακρίβεια όσο και τις λειτουργίες κινητής υποδιαστολής (floating-point operations). Πριν προχωρήσουμε, ας δούμε πώς είναι αυτή η νέα αρχιτεκτονική στον πίνακα 1:

ΠΙΝΑΚΑΣ Ι. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ EFFICIENTNET-B0 [13].

Stage i	Operator \bar{F}	Resolution $\bar{H}_i \times \bar{W}_i$	#Channels \bar{C}_i	#Layers \bar{L}_i
1	Conv3X3	224×224	32	1
2	MBConv1, k3X3	112×12	16	1
3	MBConv6, k3X3	112×12	24	2
4	MBConv6, k5X5	56×56	40	2
5	MBConv6, k3X3	28×28	80	3
6	MBConv6, k5X5	14×14	112	3
7	MBConv6, k5X5	14×14	192	4
8	MBConv6, k3X3	7×7	320	1
9	Conv1X1 & Pooling & FC	7×7	1280	1

Η αρχιτεκτονική χρησιμοποιεί 7 Αντεστραμμένα Υπολοιπώμενα Μπλοκ (Inverted Residual Block), αλλά το καθένα έχει διαφορετικές ρυθμίσεις. Αυτά τα μπλοκ χρησιμοποιούν επίσης squeeze και excitation μπλόκ μαζί με ενεργοποίηση swish. Θα συζητήσουμε για αυτά παρακάτω.

a) Ενεργοποίηση Swish

Η ομάδα του Google Brain πρότεινε μια νεότερη συνάρτηση ενεργοποίησης (activation) που τείνει να λειτουργεί καλύτερα (για βαθύτερα δίκτυα) από το ReLU, το οποίο είναι μια Swish ενεργοποίηση. Αποδεικνύουν ότι εάν αντικαταστήσουμε το Swish με το ReLU στο InceptionResNetV2, μπορούμε να επιτύχουμε 0,6% περισσότερη ακρίβεια στο σύνολο δεδομένων ImageNet. Το Swish είναι ο πολλαπλασιασμός γραμμικής και σιγμοειδούς ενεργοποίησης.

b) Αντεστραμμένο Υπολοιπώμενο Μπλοκ (Inverted Residual Block)

Η ιδέα ενός υπολειμματικού μπλοκ εισήχθη στην αρχιτεκτονική MobileNet. Το MobileNet χρησιμοποιεί σε βάθος χωριστή συνέλιξη μέσα στο αντεστραμμένο υπολοιπώμενο μπλοκ το οποίο χρησιμοποιεί πρώτα τη βαθιά συνέλιξη και έπειτα τη δεξιόστροφη συνέλιξη. Αυτή η προσέγγιση μειώνει τις εκπαιδευτικές παραμέτρους κατά μεγάλο αριθμό. Σε ένα αρχικό υπολειμματικό μπλοκ (εισήχθη στο ResNet), οι συνδέσεις παράλειψης χρησιμοποιούνται για

τη σύνδεση μεγάλων επιπέδων (γνωστά και ως στρώματα με μεγάλο αριθμό καναλιών) και υπάρχουν λιγότεροι αριθμοί καναλιών μέσα σε ένα μπλοκ (γνωστό και ως στενά στρώματα). Το αντεστραμμένο υπολογιζόμενο μπλοκ κάνει το αντίθετο, οι συνδέσεις παράλειψης συνδέουν στενά στρώματα ενώ τα ευρύτερα στρώματα βρίσκονται μεταξύ των συνδέσεων παράλειψης.

c) Μπλοκ Συμπίεσης και Διέγερσης (Squeeze and Excitation Block)

Όταν το CNN δημιουργεί ένα χάρτη χαρακτηριστικών εξόδου από ένα συνελκτικό στρώμα, δίνει ίσα βάρη σε κάθε ένα από τα κανάλια. Το μπλοκ Squeeze και Excitation (SE) είναι μια μέθοδος για να δοθεί βάρος σε κάθε κανάλι ξεχωριστά αντί να τα εκπαιδεύονται μαζί. Το μπλοκ SE δίνει την έξοδο σχήματος ($1 \times 1 \times$ κανάλια) που καθορίζει το βάρος για κάθε κανάλι και το σημαντικό είναι ότι το νευρωνικό δίκτυο μπορεί να μάθει αυτό το βάρος από μόνο του, όπως και άλλες παραμέτρους.

d) EfficientNet MBConv Block

Το μπλοκ MBConv παίρνει δύο εισόδους, το πρώτο είναι δεδομένα και το άλλο είναι ορίσματα μπλοκ. Τα δεδομένα εξάγονται από το τελευταίο επίπεδο. Ένα όρισμα μπλοκ είναι μια συλλογή χαρακτηριστικών που θα χρησιμοποιηθούν σε ένα μπλοκ MBConv όπως φίλτρα εισόδου, φίλτρα εξόδου, αναλογία επέκτασης, αναλογία συμπίεσης κ.λπ.

Το EfficientNet χρησιμοποιεί 7 MBConv μπλόκς και παρακάτω είναι οι προδιαγραφές για καθένα από αυτά τα μπλοκ αντίστοιχα :

- Το μέγεθος πυρήνα (kernel size) για συνέλιξη π.χ. 3 x 3
- Το num_repeat καθορίζει πόσες φορές πρέπει να επαναληφθεί ένα συγκεκριμένο μπλοκ, πρέπει να είναι μεγαλύτερο από το μηδέν.
- Τα φίλτρα input_filters και output_filters είναι αριθμοί καθορισμένων φίλτρων.
- expand_ratio είναι ο λόγος επέκτασης φίλτρου εισόδου.
- Το id_skip προτείνει εάν θα χρησιμοποιηθεί σύνδεση παράλειψης ή όχι.
- Το se_ratio παρέχει αναλογία συμπίεσης για μπλοκ συμπίεσης και διέγερσης.

Το EfficientNet μπορεί να λάβει μικρότερες εικόνες ως είσοδο, αλλά θα είναι υπερβολικό για ένα σύνολο δεδομένων όπως το MNIST. Συνιστάται η χρήση EfficientNets για σύνθετα σύνολα δεδομένων [13].

IV. ΜΕΘΟΔΟΛΟΓΙΑ

A. Σύνολο δεδομένων (dataset)

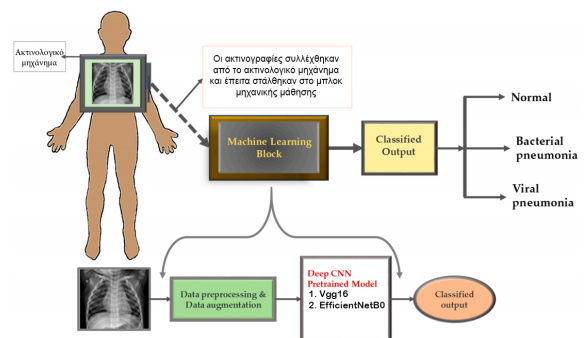
Σε αυτήν την εργασία, χρησιμοποιήθηκε η βάση δεδομένων πνευμονίας ακτινογραφίας θώρακος από τον εκπαιδευτικό διαγωνισμό Detect Pneumonia (Fall 2020) στο kaggle, η οποία αποτελείται από 4672 εικόνες ακτινογραφίας θώρακα με αναλύσεις που κυμαίνονται από 400p έως 2000p. Από τις 4672

εικόνες ακτινογραφίας θώρακος, 3445 εικόνες προέρχονται από διαφορετικά θέματα που επηρεάζονται από πνευμονία (2238 εικόνες για βακτηριακή πνευμονία και 1207 εικόνες για ιογενή πνευμονία) και 1227 εικόνες από φυσιολογικά άτομα (Πίνακας 1). Αυτό το σύνολο δεδομένων χωρίστηκε στο σετ εκπαίδευσης (training) και επαλήθευσης (validation).

ΠΙΝΑΚΑΣ II. ΣΤΟΙΧΕΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ .

Τύπος	Αριθμός x-ray εικόνων
Normal	1227
Bacterial Pneumonia	2238
Viral Pneumonia	1207
Σύνολο	4672

Επίσης, χρησιμοποιήθηκε η γλώσσα προγραμματισμού python μέσω της πλατφόρμας Colab της Google για την εκπαίδευση, την αξιολόγηση και τη δοκιμή διαφορετικών αλγορίθμων. Επιπλέον, χρησιμοποιήθηκαν οι βιβλιοθήκες του keras και tensorflow καθώς και όλες οι απαραίτητες βιβλιοθήκες της python για επεξεργασία των δεδομένων. Η εικόνα 9 απεικονίζει την επισκόπηση της μεθοδολογίας αυτής της εργασίας. Τα σύνολα εικόνων υποβάλλονται σε κάποια βήματα επεξεργασίας (preprocessing), αύξηση δεδομένων (data augmentation) και στη συνέχεια περνάνε στην εκπαίδευση χρησιμοποιώντας προ-εκπαιδευμένα μοντέλα βαθιάς μάθησης: VGG16 & EfficientNetB0 και όλους τους δοκιμασμένους αλγόριθμους στο σύνολο δεδομένων εκπαίδευσης (training set).



Εικόνα 9. Επισκόπηση της μεθοδολογίας [2].

Το σύνολο δεδομένων είναι χωρισμένο σε δύο φακέλους train & test. Στα train images τα οποία θα χρησιμοποιήσουμε για την εκπαίδευση του μοντέλου μας σε συνδυασμό με το αρχείο labels_train.csv στο οποίο μας δίνεται η πληροφορία για την κατηγορία κάθε εικόνας. Στα test images όπου βρίσκονται οι εικόνες στις οποίες θα εφαρμοστεί η πρόβλεψη του τελικού μοντέλου για την κατηγορία στην οποία ανήκουν, ενώ αυτές τις προβλέψεις αποθηκεύονται σε ένα νέο csv αρχείο στον φάκελο predicts το οποίο και θα κάνουμε υποβολή στο διαγωνισμό στο kaggle.

B. Προ-επεξεργασία (Pre-Processing)

Αρχικά για το διάβασμα των εικόνων χρησιμοποιήθηκε η βιβλιοθήκη ImageDataGenerator διαβάζοντας απο το αρχείο train_labels.csv και χωρίζοντας τα δεδομένα απο τον φάκελο train κατά 80% σε train set και κατά 20% σε δεδομένα του validation set. Ένα από τα πιο σημαντικά βήματα στην προ-επεξεργασία δεδομένων ήταν η αλλαγή του μεγέθους των εικόνων των ακτινογραφιών, καθώς η εικόνα εισόδου για διαφορετικούς αλγόριθμους ήταν διαφορετική. Το απαιτούμενο μέγεθος των εικόνων ήταν 224×224. Επιπλέον, θέσαμε το διάβασμα των εικόνων σε 3 κανάλια ως RGB και εφαρμόσαμε παρεμβολή (interpolation) τύπου nearest. Τέλος, πραγματοποιήσαμε μια τεχνική upsampling στις κλάσεις 1 και 0 ώστε να έχουμε ίδιο αριθμό δεδομένων σε όλες τις κλάσεις. Όλες οι εικόνες κανονικοποιήθηκαν σύμφωνα με τα προ-εκπαιδευμένα πρότυπα μοντέλων.

Όπως συζητήθηκε προηγουμένως, τα CNN λειτουργούν καλύτερα με ένα μεγάλο σύνολο δεδομένων. Ωστόσο, το μέγεθος του συγκεκριμένου συνόλου δεδομένων δεν είναι πολύ μεγάλο. Αναφέρεται ότι η αύξηση δεδομένων (data augmentation) μπορεί να βελτιώσει την ακρίβεια ταξινόμησης των αλγορίθμων βαθιάς μάθησης. Η απόδοση των μοντέλων βαθιάς μάθησης μπορεί να βελτιωθεί με την αύξηση των υπαρχόντων δεδομένων παρά τη συλλογή νέων δεδομένων [2].

Συνεπώς, εφαρμόσαμε την τεχνική αύξησης του δείγματος των δεδομένων (Data Augmentation) μέσω βιβλιοθηκών του Keras. Μερικές απο τις παραμέτρους που χρησιμοποιήσαμε ήταν οι παρακάτω:

- rescale
- rotation_range
- width_shift_range
- height_shift_range
- shear_range
- horizontal_flip

C. Τελειοποίηση (Fine Tuning).

Εκτός των μοντέλων που παρουσίασε εκτελέσαμε έναν μεγάλο αριθμό δοκιμών και πειραμάτων με μια ποικιλία διαφορετικών υπερ παραμέτρων και με βάση άλλα μοντέλα (π.χ DenseNet121), όπως οι μετρήσεις που παρακολουθούσαμε, οι εποχές που τρέξαμε σε όλους τους αλγόριθμους και τα μοντέλα καθώς και οι διάφοροι βελτιστοποιητές (optimizers). Παρακάτω θα δούμε την εφαρμογή των προ-εκπαιδευμένων μοντέλων.

Εφαρμόζοντας την μεταφορά μάθησης και αφού κατεβάσαμε τα προεκπαιδευμένα βάρη απο τα μοντέλα βαθιάς μάθησης προχωρήσαμε στην τελειοποίηση (Fine Tuning). Σε αυτό το στάδιο εφαρμόσαμε το “πάγωμα” (freeze) όλων των επιπέδων των προεκπαιδευμένων μοντέλων ώστε να μην εκπαιδευτούν ξανά. Στην συνέχεια αρχικοποιήσαμε και προετοιμάσαμε τα στοιχεία checkpoint και EarlyStopping τα οποία συμβάλλουν στην καλύτερη και πιο έγκαιρη εκπαίδευση του μοντέλου, αποθηκεύοντας τα καλύτερα βάρη

κάθε εποχής. Έπειτα γίνεται η εκτέλεση (compile) του μοντέλου θέτοντας κάποιες παραμέτρους όπως είναι οι παρακάτω:

- Optimizers : {Adam, RMSprop, Adagrad}
- Loss Functions: {categorical_crossentropy, binary_crossentropy}
- metrics : {accuracy}

Κάνοντας όλα τα παραπάνω στην συνέχεια βάζουμε σε λειτουργία την εκπαίδευση του μοντέλου συνθέτοντας όλες τις παραμέτρους. Στην συνέχεια αφού έχει τελειώσει η εκπαίδευση του μοντέλου εμφανίζουμε τις καμπύλες σφάλματος (loss) και ακρίβειας (accuracy) των μοντέλων. Τέλος, γίνεται αποθήκευση και επαλήθευση του μοντέλου πάνω στα δεδομένα των εικόνων του test_set παράγοντας ένα αρχείο υπολογιστικής μορφής (.csv) το οποίο διαθέτει 2 στήλες (file_name, class_id) και 1168 γραμμές με τις προβλέψεις του μοντέλου για κάθε εικόνα. Δείγμα του αρχείου φαίνεται στην εικόνα 12.

flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3211392
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 64)	4160
dense_3 (Dense)	(None, 32)	2080
dropout (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 16)	528
dropout_1 (Dropout)	(None, 16)	0
dense_5 (Dense)	(None, 3)	51
=====		
Total params: 17,941,155		
Trainable params: 3,226,467		
Non-trainable params: 14,714,688		

Εικόνα 10. Τα τελικά στρώματα (layers) και παράμετροι του VGG16.

top_activation (Activation)	(None, 7, 7, 1280)	0	top_bn[0][0]
flatten_3 (Flatten)	(None, 62720)	0	top_activation[0][0]
dense_18 (Dense)	(None, 128)	8028288	flatten_3[0][0]
dense_19 (Dense)	(None, 64)	8256	dense_18[0][0]
dense_20 (Dense)	(None, 64)	4160	dense_19[0][0]
dense_21 (Dense)	(None, 32)	2080	dense_20[0][0]
dropout_6 (Dropout)	(None, 32)	0	dense_21[0][0]
dense_22 (Dense)	(None, 16)	528	dropout_6[0][0]
dropout_7 (Dropout)	(None, 16)	0	dense_22[0][0]
dense_23 (Dense)	(None, 3)	51	dropout_7[0][0]
=====			
Total params: 12,092,927			
Trainable params: 8,043,363			
Non-trainable params: 4,049,564			

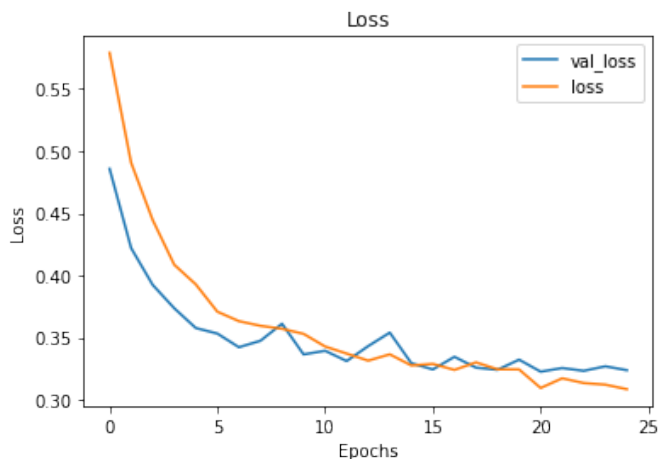
Εικόνα 11. Τα τελικά στρώματα (layers) και παράμετροι του EfficientNetB0.

	A	B
1	file_name	class_id
2	img_100555673279	0
3	img_101394799299	1
4	img_103737282429	0
5	img_103791004852	1
6	img_104618299654	0
7	img_105515553879	0
8	img_105649378982	1
9	img_106353932051	1
10	img_106762773318	1

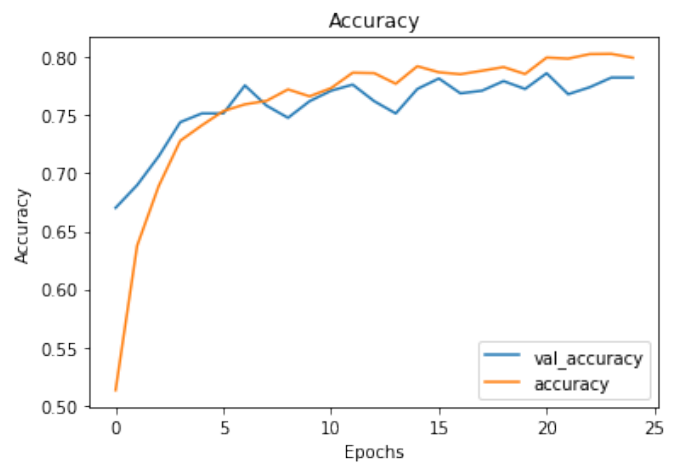
Εικόνα 12. Δείγμα του αρχείου .csv με τις προβλέψεις.

V. ΣΥΝΟΨΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

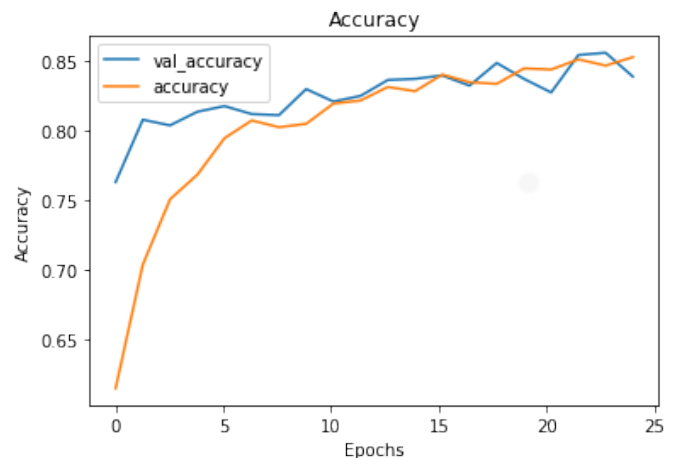
Παρακάτω βλέπουμε τις καμπύλες loss και accuracy των δύο μοντέλων που χρησιμοποιήσαμε για την ταξινόμηση των εικόνων. Όπως παρατηρούμε στην εικόνα 15 το μοντέλο του EfficientNetB0 ήταν το πιο αποδοτικό καταφέρνοντας να φτάσει σε ακρίβεια ακόμη και στο 85-86%, ενώ το μοντέλο του VGG16 κατάφερε να φτάσει σε ακρίβεια μέχρι και ~82%. Να σημειωθεί ως ακρίβεια των μοντέλων αναφέρεται πάντα η ακρίβειας επαλήθευσης (validation accuracy). Η τελική και πιο υψηλή βαθμολογία των μοντέλων πάνω στο σύνολο δεδομένων του test_set έφτασε το 84.6% στον διαγωνισμό του kaggle. Το συγκεκριμένο σκόρ ανήκε στο μοντέλο του EfficientNetB0.



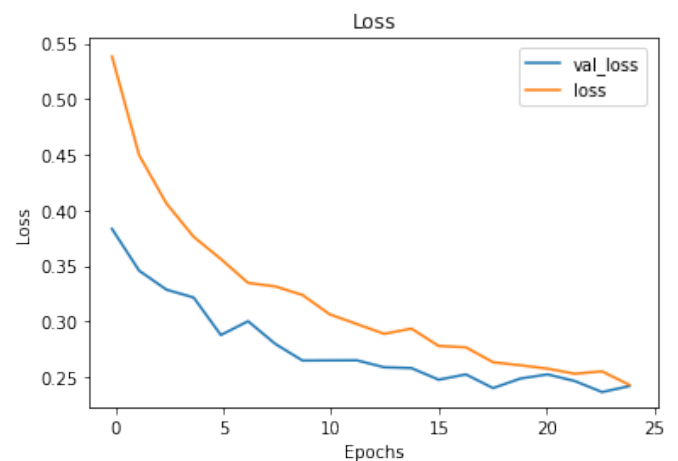
Εικόνα 13. Διάγραμμα καμπυλών loss του VGG16.



Εικόνα 14. Διάγραμμα καμπυλών accuracy του VGG16.



Εικόνα 15. Διάγραμμα καμπυλών accuracy του EfficientNetB0.



Εικόνα 16. Διάγραμμα καμπυλών accuracy του EfficientNetB0.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- [1] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning," *Diagnostics*, vol. 10, no. 6, p. 417, Jun. 2020.
- [2] T. Rahman et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, May 2020.
- [3] P. Louridas and C. Ebert, "Machine Learning," in *IEEE Software*, vol. 33, no. 5, pp. 110-115, Sept.-Oct. 2016.
- [4] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019.
- [5] S. K. Yewale and P. K. Bharné, "Hand gesture recognition using different algorithms based on artificial neural network," 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Udaipur, 2011, pp. 287-292.
- [6] I. I. Tsvetkovskaya, N. V. Tekutieva, E. N. Prokofeva and A. V. Vostrikov, "Methods of Obtaining Geospatial Data Using Satellite Communications and Their Processing Using Convolutional Neural Networks," 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), Moscow, Russia, 2020, pp. 1-5.
- [7] G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari and A. Shukla, "Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning," 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 2020, pp. 85-92.
- [8] M. Pandiya, S. Dassani and P. Mangalraj, "Analysis of Deep Learning Architectures for Object Detection - A Critical Review," 2020 IEEE-HYDCON, Hyderabad, India, 2020, pp. 1-6.
- [9] S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," *IJSRP*, vol. 9, no. 10, p. p9420, Oct. 2019.
- [10] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, "A New Method for Face Recognition Using Convolutional Neural Network," *AEEE*, vol. 15, no. 4, Nov. 2017.
- [11] M. Loukadakis, J. Cano, and M. OBoyle, "Accelerating deep neuralnetworks on low power heterogeneous architectures," 2018.
- [12] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Prog Artif Intell*, vol. 9, no. 2, pp. 85–112, Dec. 2019.
- [13] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficient-Det: Scalable and efficient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 4, 1300