

Εφαρμογές Βαθιάς Μάθησης για τον Εντοπισμό Καρκίνου του Πνεύμονα

Εμμανουήλ Διλιμπέρης
Τμήμα Ψηφιακών Συστημάτων
Πανεπιστήμιο Πειραιώς
Μεγάλα Δεδομένα και Αναλυτική
me2104@unipi.gr

Κοσμάς Κατσιούλας
Τμήμα Ψηφιακών Συστημάτων
Πανεπιστήμιο Πειραιώς
Μεγάλα Δεδομένα και Αναλυτική
me2109@unipi.gr

Άρμεν Μοβσεσιάν
Τμήμα Ψηφιακών Συστημάτων
Πανεπιστήμιο Πειραιώς
Μεγάλα Δεδομένα και Αναλυτική
me2117@unipi.gr

Περίληψη— Ο καρκίνος του πνεύμονα είναι μια από τις πιο επικίνδυνες ασθένειες παγκοσμίως και αποτελεί την συχνότερη μορφή καρκίνου στον άνθρωπο. Η έγκαιρη διάγνωση του στα αρχικά στάδια προσδίδει αναμφισβήτητο όφελος στην επιβίωση των ασθενών. Η διαδικασία της διάγνωσης περιλαμβάνει μια βιοψία, κατά την οποία αφαιρείται ένα μικρό κομμάτι ιστού και εξετάζεται για την ύπαρξη μη φυσιολογικών κυττάρων. Ένα από τα πιο συχνά λάθη στη βιοψία είναι η λανθασμένη ερμηνεία των αποτελεσμάτων, το οποίο μπορεί να οφείλεται σε πολλούς λόγους, όπως η απειρία ή η απροθυμία να μη ληφθούν υπόψιν παράγοντες που αντιτίθενται στην αρχική διάγνωση ενός παθολόγου, θέτοντας σε κίνδυνο με αυτόν τον τρόπο τη ζωή πολλών ασθενών. Αυτό το πρόβλημα θα μπορούσε να λυθεί με τη χρήση μεθόδων βαθιάς μάθησης, οι οποίες παρέχουν έναν αποτελεσματικό τρόπο κατασκευής ενός μοντέλου που θα μπορεί να ταξινομεί ιστοπαθολογικές εικόνες με μεγάλη ακρίβεια. Στην παρούσα εργασία θα υλοποιηθεί η δημιουργία ενός Συνελκτικού Νευρωνικού Δικτύου (CNN) από την αρχή, αλλά και η μετατροπή προεκπαιδευμένων CNN μοντέλων, με σκοπό τον εντοπισμό του καρκίνου του πνεύμονα σε ιστοπαθολογικές εικόνες μικροσκοπίου. Οι επιδόσεις των μοντέλων αξιολογούνται με βάση την ακρίβεια (accuracy).

Keywords— Lung Cancer Detection, Deep Learning, CNN, Pre-trained CNN, Computer-Aided Diagnosis,

I. ΕΙΣΑΓΩΓΗ

Ο καρκίνος είναι μια απειλητική για τη ζωή ασθένεια, που χαρακτηρίζεται από τη μη φυσιολογική κυτταρική ανάπτυξη και μπορεί να εμφανιστεί σε οποιοδήποτε όργανο και ιστό του ανθρώπινου σώματος αλλά και στο αίμα. Στη λίστα με τις κύριες αιτίες θανάτου καταλαμβάνει τη δεύτερη θέση σε παγκόσμιο επίπεδο. Η Αμερικάνικη Αντικαρκινική Εταιρία (American Cancer Society) εκτιμά ότι για το έτος 2022 θα διαγνωστούν 1.9 εκατομμύρια νέες περιπτώσεις καρκίνου και 609,360 θάνατοι από καρκίνο στις Ηνωμένες Πολιτείες. Ο καρκίνος του πνεύμονα θα αντιπροσωπεύει τις 236,740 νέες περιπτώσεις και τους 130,180 θανάτους.

Οι καρκίνοι του πνεύμονα ταξινομούνται σύμφωνα με τον ιστολογικό τύπο. Η ταξινόμηση αυτή έχει σημαντικές συνέπειες στην κλινική διαχείριση και την πρόγνωση της νόσου. Η συντριπτική πλειονότητα των καρκίνων του πνεύμονα είναι καρκινώματα-κακοήθειες που προκύπτουν από επιθηλιακά κύτταρα. Τα καρκινώματα του πνεύμονα κατηγοριοποιούνται από το μέγεθος και την εμφάνιση των κακοηθών κυττάρων που μπορούν να παρατηρηθούν από παθολογοανατόμο με μικροσκόπηση. Οι δύο ευρείες κατηγορίες είναι το μη μικροκυτταρικό καρκίνωμα (NSCLC) και το μικροκυτταρικό καρκίνωμα του πνεύμονα (SCLC). Το μη μικροκυτταρικό καρκίνωμα αφορά το 82% των περιπτώσεων και οι τρεις κύριες υποκατηγορίες του είναι το αδενοκαρκίνωμα, το καρκίνωμα των πλακωδών κυττάρων του πνεύμονα και των μεγαλοκυττάρων του πνεύμονα.

Η διάγνωση με τη βοήθεια υπολογιστή (CAD) αποτελεί σημαντικό πεδίο έρευνας τις τελευταίες δεκαετίες. Το CAD

χρησιμοποιεί μεθόδους μηχανικής μάθησης για να αναλύσει απεικονιστικά και μη απεικονιστικά δεδομένα ασθενών και να αξιολογήσει την κατάσταση του ασθενούς, οι οποίες στη συνέχεια μπορούν να χρησιμοποιηθούν για να βοηθήσουν τους γιατρούς στη διαδικασία λήψης αποφάσεων. Η επιτυχία της τεχνολογίας της βαθιάς μάθησης στη μηχανική μάθηση, ωθεί συνεχόμενα σε προσπάθειες ανάπτυξης και βελτίωσης της απόδοσης του CAD.

Το επιστημονικό άρθρο [1] με τίτλο ‘CNN-Based CAD for Breast Cancer Classification in Digital Breast Tomosynthesis’ των Jinn-Yi Yeh και Siwa Chan αφορά μια σχετική μελέτη με θέμα την δημιουργία ενός Συνελκτικού Νευρωνικού Δικτύου (CNN) το οποίο αναγνωρίζει τον καρκίνο του μαστού από τομοσυνθέσεις, δηλαδή τρισδιάστατες απεικονίσεις του μαστού. Ένα ακόμα σχετικό άρθρο [2] είναι αυτό των Sun Young Park και Dusty Sargent με τίτλο ‘Colonoscopic polyp detection using convolutional neural networks’, όπου αναπτύσσεται ένας CAD αλγόριθμος βασισμένος σε CNN για την ανίχνευση πολυπόδων σε βίντεο κολονοσκόπησης.

Σκοπός της παρούσας εργασίας είναι η αυτοματοποίηση της δουλειάς του παθολογοανατόμου, ο οποίος είναι ο γιατρός που εξετάζει στο μικροσκόπιο τον καρκινικό ιστό και τα κύτταρα του, βγάζει τη διάγνωση του καρκίνου και συντάσσει την έκθεση ιστολογικής-παθολογοανατομικής εξέτασης. Με τη βοήθεια των Συνελκτικών Νευρωνικών Δικτύων (CNN), τα οποία ανήκουν στην κατηγορία των αλγορίθμων της βαθιάς μάθησης, θα δημιουργηθούν μοντέλα ικανά να αναγνωρίσουν και να εντοπίσουν τον καρκίνο του πνεύμονα σε ιστοπαθολογικές εικόνες.

II. ΜΕΘΟΔΟΛΟΓΙΑ

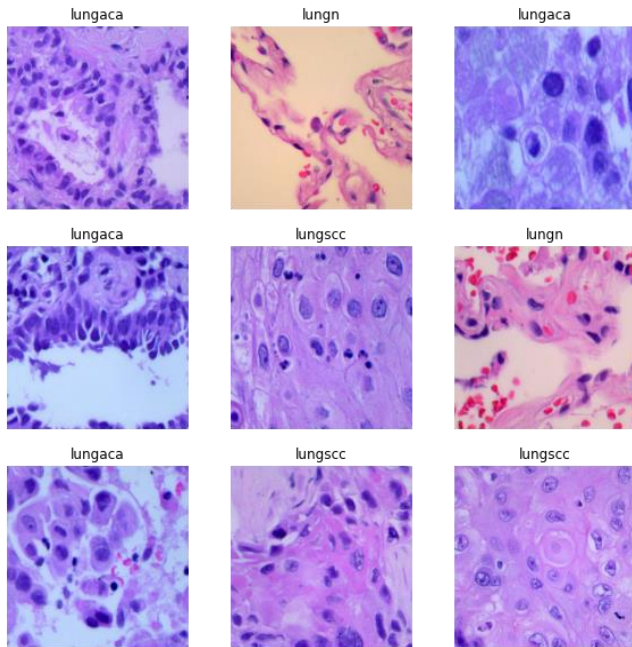
A. Σύνολο Δεδομένων και Προεπεξεργασία

Το σύνολο δεδομένων περιέχει 7,500 έγχρωμες ιστοπαθολογικές εικόνες πνευμόνων. Όλες οι εικόνες έχουν μέγεθος 768×768 pixels και η μορφή των αρχείων είναι jpeg. Υπάρχουν τρεις κλάσεις στο σύνολο δεδομένων και σε κάθε μια αναλογούν 2500 εικόνες. Οι κλάσεις αυτές είναι :

- Καλοήθης (lungn)
- Αδενοκαρκίνωμα (lungaca)
- Καρκίνωμα πλακωδών κυττάρων (lungsc)

Η κάθε εικόνα είναι μια διάταξη 768×768×3, όπου κάθε έγχρωμο pixel συσχετίζεται με τρεις τιμές που αντιστοιχούν στο κόκκινο, στο μπλε και στο πράσινο χρώμα της εικόνας σε μια καθορισμένη χωρική θέση. Οι τιμές που μπορεί να πάρει ένα pixel σε κάθε διάσταση χρώματος είναι από 0 μέχρι 255, όπου το 0 αντιστοιχεί στο μαύρο και το 255 στο λευκό. Στην διαδικασία της προεπεξεργασίας γίνεται η

κανονικοποίηση αυτών των τιμών στο πεδίο $[0, 1]$. Αυτό το βήμα είναι απαραίτητο, διότι τα νευρωνικά δίκτυα επεξεργάζονται τις εισόδους χρησιμοποιώντας μικρές τιμές στα βάρη τους και οι εισόδοι με μεγάλες ακέραιες τιμές μπορούν να διαταράξουν ή να επιβραδύνουν τη διαδικασία εκμάθησης. Επίσης για όλες τις εικόνες γίνεται αλλαγή μεγέθους από 768×768 που ήταν το αρχικό τους μέγεθος σε 224×224 καθώς τα μοντέλα βαθιάς μάθησης εκπαιδεύονται πιο γρήγορα σε μικρές εικόνες. Μια μεγαλύτερη εικόνα εισόδου απαιτεί πολύ περισσότερο χρόνο εκπαίδευσης.



Εικ. 1. Δείγμα εικόνων από το σύνολο δεδομένων

Για τη δημιουργία αξιόπιστων μοντέλων είναι ανάγκη να διαχωριστεί το αρχικό σύνολο δεδομένων σε τρία μικρότερα, το σύνολο εκπαίδευσης (train set) το οποίο περιέχει το 60% των δεδομένων του αρχικού, και τα σύνολα επικύρωσης (validation set) και ελέγχου (test set), που περιέχουν το καθένα το 20%. Το σύνολο εκπαίδευσης είναι το δείγμα το οποίο χρησιμοποιείται για την διαδικασία της εκπαίδευσης του μοντέλου κατά την οποία διαμορφώνονται κατάλληλα τα βάρη του, με σκοπό την ελαχιστοποίηση του σφάλματος. Μία από τις μεγαλύτερες και πιο συχνές προκλήσεις στα νευρωνικά δίκτυα είναι το φαινόμενο της υπερπροσαρμογής (overfitting). Η υπερπροσαρμογή του μοντέλου συνήθως συμβαίνει όταν η διαδικασία της εκπαίδευσης επαναλαμβάνεται για πολλές εποχές. Έτσι σχηματίζονται βάρη, τα οποία αποδίδουν πολύ καλά για το σύνολο εκπαίδευσης, όμως αν εφαρμοστούν σε άγνωστα δεδομένα, το μοντέλο έχει πολύ χαμηλή ακρίβεια. Για τον εντοπισμό αυτού του φαινομένου δημιουργείται το σύνολο επικύρωσης, πάνω στο οποίο γίνονται προβλέψεις μετά από κάθε εποχή, πριν ξεκινήσει πάλι η διαδικασία της εκπαίδευσης και της αναπροσαρμογής των βαρών. Με αυτόν τον τρόπο παρακολουθείται η επίδοση του μοντέλου και στα δυο σύνολα. Σταματώντας την διαδικασία πριν το σημείο όπου η απόδοση στο σύνολο εκπαίδευσης συνεχίζει να αυξάνεται, ενώ στο σύνολο επικύρωσης μένει στάσιμη ή αρχίζει να μειώνεται, αποφεύγεται η υπερπροσαρμογή. Το σύνολο ελέγχου παρέχει μια αμερόληπτη αξιολόγηση του τελικού μοντέλου. Μία από τις καλύτερες τεχνικές για την μείωση

του overfitting, είναι η επαύξηση [3] (augmentation) των δεδομένων του συνόλου εκπαίδευσης. Οι τεχνικές που χρησιμοποιήθηκαν στο παρόν σύνολο, είναι δεξιά και αριστερή περιστροφή μέχρι 10 μοίρες και οριζόντια και κάθετη αντιστροφή.

B. Περιγραφή Συνελικτικών Νευρωνικών Δικτύων

Για τους σκοπούς της εργασίας θα χρησιμοποιηθούν τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) [4], τα οποία είναι ένας τύπος βαθιάς μάθησης που έχει αποδειχθεί ότι αποδίδει καλά σε προβλήματα υπολογιστικής όρασης όπως ταξινόμηση εικόνας, ανίχνευση αντικειμένων.

Κύριο χαρακτηριστικό τους είναι η πράξη της συνέλιξης. Η συνέλιξη στα μαθηματικά εφαρμόζεται μεταξύ δυο συναρτήσεων και παράγει μια τρίτη συνάρτηση, η οποία εκφράζει, πως το μέγεθος της μιας συνάρτησης τροποποιείται από την άλλη. Αποτελούνται από ένα επίπεδο εισόδου, κρυφά επίπεδα και ένα επίπεδο εξόδου. Τα επίπεδα που μπορεί να εμπεριέχονται στην αρχιτεκτονική ενός συνελικτικού δικτύου είναι:

1) *Συνελικτικό Επίπεδο (Convolution Layer)*: Ένα επίπεδο συνέλιξης μετασχηματίζει την εικόνα εισόδου για να εξαγάγει χαρακτηριστικά από αυτήν. Σε αυτόν τον μετασχηματισμό, η εικόνα συμπλέκεται με έναν φίλτρο (kernel). Ο kernel είναι ένας πίνακας μικρότερων διαστάσεων από την εικόνα εισόδου, ο οποίος σύρεται κατά ύψος και πλάτος της εικόνας, υπολογίζοντας έτσι το εσωτερικό γινόμενο Frobenius μεταξύ αυτού και της εικόνας σε κάθε χωρική θέση της. Η εικόνα εξόδου αναφέρεται και ως χαρακτηριστικό συνέλιξης. Το πρώτο συνελικτικό επίπεδο είναι υπεύθυνο για την αναγνώριση χαμηλού επιπέδου χαρακτηριστικών όπως για παράδειγμα ακμές και χρώματα. Προσθέτοντας και άλλα τέτοια επίπεδα γίνεται εφικτή η αναγνώριση χαρακτηριστικών υψηλού επιπέδου, δημιουργώντας με αυτόν τον τρόπο ένα δίκτυο που μπορεί να κατανοήσει πλήρως το νόημα των εικόνων.

2) *Επίπεδα Συγκέντρωσης (Pooling Layers)*: Τα επίπεδα συγκέντρωσης [5] είναι υπεύθυνα για την μείωση διαστάσεων των συνελικτικών χαρακτηριστικών. Υπάρχουν δύο τύποι, το Max Pooling και το Average Pooling. Το πρώτο επιστρέφει τη μέγιστη τιμή από το κομμάτι της εικόνας που θα καλυφθεί από έναν προκαθορισμένο kernel, ενώ το δεύτερο τη μέση τιμή αυτών. Επιπλέον τα επίπεδα αυτά είναι χρήσιμα για την εξαγωγή κυρίαρχων χαρακτηριστικών και μειώνουν παράλληλα το υπολογιστικό κόστος.

3) *Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)*: Ένα πλήρως συνδεδεμένο επίπεδο [6] βρίσκεται στο τέλος ενός CNN. Τα χαρακτηριστικά που έχουν παραχθεί προηγουμένως μετατρέπονται σε ένα διάνυσμα, το οποίο τροφοδοτεί ένα πλήρως συνδεδεμένο επίπεδο, με σκοπό την εύρεση σύνθετων συσχετίσεων μεταξύ των χαρακτηριστικών υψηλού επιπέδου. Η έξοδος του επιπέδου είναι ένα διάνυσμα χαρακτηριστικών μίας διάστασης.

C. Transfer Learning

Κατά τη διαδικασία του πειραματισμού, πέρα από την κατασκευή ενός CNN από το μηδέν, μια ακόμα προσέγγιση είναι η χρήση του Transfer Learning [7], για τη δημιουργία μοντέλων. Το Transfer Learning είναι μια μέθοδος, στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης, που στοχεύει τη μεταφορά της γνώσης που αποκτάται από μία εργασία, σε μια διαφορετική αλλά παρόμοια εργασία.

Επίσης βοηθά στη μείωση της υπολογιστικής δύναμης και επιταχύνει την απόδοση του συστήματος, που επιφέρει την εμφάνιση ενδείξεων ταχύτερων και βελτιωμένων αποτελεσμάτων. Για την αξιοποίηση της μεθόδου θα γίνει η χρήση προεκπαιδευμένων CNN μοντέλων, τα οποία έχουν εκπαιδευτεί σε μεγάλα σύνολα δεδομένων, καθώς επίσης και η βελτιστοποίηση τους (fine-tuning) για την επίτευξη υψηλής ακρίβειας. Σε ένα νευρωνικό δίκτυο, τα κατώτατα και το μεσαία επίπεδα αντιπροσωπεύουν συνήθως τα γενικά χαρακτηριστικά και τα ανώτατα επίπεδα αντιπροσωπεύουν τα χαρακτηριστικά του συγκεκριμένου προβλήματος. Δεδομένου ότι το νέο πρόβλημα είναι διαφορετικό από το αρχικό θα πετάξουμε τα ανώτερα επίπεδα. Στη θέση τους θα προστεθούν νέα, με σκοπό την επιθυμητή έξοδο, η οποία θα είναι σχετική με το νέο πρόβλημα. Πριν την διαδικασία της εκπαίδευσης, γίνεται πάγωμα των κατώτατων και μεσαίων επιπέδων, έτσι ώστε να μην μεταβληθούν τα βάρη τους.

D. Μετάφραση των Μοντέλων (Compilation)

Στη διαδικασία του compilation του μοντέλου, ορίζεται ο αλγόριθμος βελτιστοποίησης, ο οποίος είναι υπεύθυνος για την τροποποίηση των χαρακτηριστικών του νευρωνικού δικτύου, όπως για παράδειγμα τα βάρη και τον ρυθμό εκμάθησης (learning rate), με σκοπό τη μείωση του συνολικού σφάλματος και την αύξηση της απόδοσης. Υπάρχουν διάφοροι αλγόριθμοι βελτιστοποίησης όπως ο Adaptive Learning Rate (Adam), ο Stochastic Gradient Descent (SGD), ο Adadelata, ο RM-SProp, ο AdaMax και πολλοί άλλοι. Επιλέχθηκε ο αλγόριθμος Adam [8] καθώς λειτουργεί αρκετά καλά για κάθε μοντέλο βαθιάς μάθησης. Οι υπερπαραμέτροι που ορίστηκαν είναι, ο ρυθμός εκμάθησης (learning rate) στην τιμή 0.001, οι εποχές εκπαίδευσης (epochs) σε 10 και ο αριθμός των παραδειγμάτων εκπαίδευσης που χρησιμοποιούνται σε μία επανάληψη (batch size) σε 32.

E. Μετρικές Αξιολόγησης και Συνάρτηση Σφάλματος

Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκε η μετρική της ακρίβειας (accuracy) και η συνάρτηση σφάλματος.

Η ακρίβεια υπολογίζεται από τον τύπο:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

Όπου TP (True Positive) τα αληθώς θετικά αποτελέσματα, TN (True Negative) τα αληθώς αρνητικά αποτελέσματα, FP (False Positive) τα ψευδώς θετικά αποτελέσματα και FN (False Negative) τα ψευδώς αρνητικά αποτελέσματα.

Όταν η ταξινόμηση έχει παραπάνω από δυο κλάσεις μια συνάρτηση σφάλματος που μπορεί να χρησιμοποιηθεί είναι η categorical cross-entropy. Για τη συγκεκριμένη συνάρτηση πρέπει ο αριθμός των κόμβων εξόδου να είναι όσες και οι κλάσεις, δηλαδή τρεις. Επίσης η έξοδος του τελικού στρώματος θα πρέπει να περάσει μέσω μιας συνάρτησης ενεργοποίησης, τη softmax, έτσι ώστε ο κάθε κόμβος να βγάξει μια τιμή πιθανότητας μεταξύ [0,1].

F. Περιγραφή Υλοποιημένων Μοντέλων

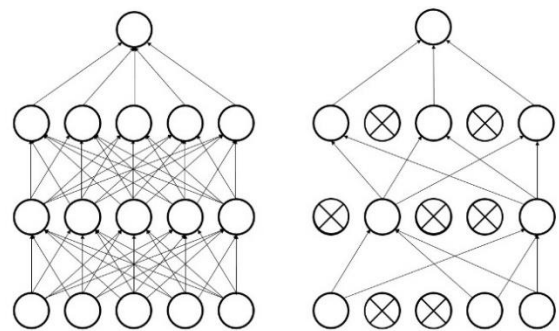
Στα πλαίσια της εργασίας δημιουργήθηκε ένα Συνελικτικό Νευρωνικό Δίκτυο από το μηδέν και τέσσερα διαφορετικά Pre-Trained δίκτυα για την κατηγοριοποίηση των ιστοπαθολογικών εικόνων του πνεύμονα. Στην διαδικασία της εκπαίδευσης των μοντέλων προστέθηκε η μέθοδος ανάκλησης (callback) ModelCheckpoint. Η ModelCheckpoint παρακολουθεί στη συγκεκριμένη περίπτωση, τη συνάρτηση σφάλματος και αποθηκεύει τα

βάρη του διαστήματος, όπου το σφάλμα, στο σύνολο επικύρωσης, είχε τη χαμηλότερη τιμή. Στη συνέχεια τα βάρη αυτά, μετά το πέρας της διαδικασίας της εκπαίδευσης, μπορούν να φορτωθούν στο τελικό μοντέλο.

1) *Απλό Συνελικτικό Νευρωνικό Δίκτυο*: Το συνελικτικό νευρωνικό δίκτυο που δημιουργήθηκε αποτελείται από εννιά επίπεδα. Στα έξι πρώτα βρίσκονται τρία συνελικτικά επίπεδα με 32, 64, 128 κόμβους αντίστοιχα και ύστερα από το κάθε ένα τοποθετείται ένα επίπεδο MaxPooling που μειώνει τη διάσταση στο μισό. Στα συνελικτικά επίπεδα το μέγεθος του kernel έχει οριστεί ως ένας πίνακας 3×3, για κάθε διάσταση χρώματος. Επίσης στα συνελικτικά επίπεδα, εφαρμόζεται η τεχνική padding, η οποία προσθέτει συμμετρικά στην εικόνα μηδενικές τιμές, έτσι ώστε η έξοδος να έχει το ίδιο μέγεθος με την είσοδο. Η συνάρτηση ενεργοποίησης που εφαρμόστηκε είναι η Rectified Linear Unit (ReLU) [9], μια τμηματικά γραμμική συνάρτηση που θα εξάγει απευθείας την είσοδο εάν είναι θετική, διαφορετικά θα επιστρέψει μηδέν. Στη συνέχεια η έξοδος από τα συνελικτικά δίκτυα μετατρέπεται σε ένα διάνυσμα χαρακτηριστικών (feature vector) με τη χρήση ενός Flatten επιπέδου, το οποίο συνδέεται με τα πλήρως συνδεδεμένα επίπεδα (Fully Connected Layers). Υπάρχουν δυο πλήρως συνδεδεμένα επίπεδα με 128 και 64 κόμβους, με συνάρτηση ενεργοποίησης ομοίως τη ReLU. Το τελευταίο επίπεδο (output layer) που παράγει την τελική πρόβλεψη αποτελείται από 3 κόμβους, όσες και οι κλάσεις του συνόλου δεδομένων, και σαν συνάρτηση ενεργοποίησης, διαθέτει την softmax.

Κατά τη διάρκεια της εκπαίδευσης παρατηρήθηκε το φαινόμενο του overfitting (Εικ. 4). Το σύνολο επικύρωσης εμφάνισε στο σφάλμα την χαμηλότερη τιμή την 7^η εποχή και από εκεί και ύστερα δεν εμφάνισε βελτίωση. Αντίθετα, το σύνολο εκπαίδευσης συνέχισε με το πέρας των εποχών να βελτιώνεται, γεγονός που αποδεικνύει ότι το μοντέλο υπερπροσαρμόζεται σε αυτό.

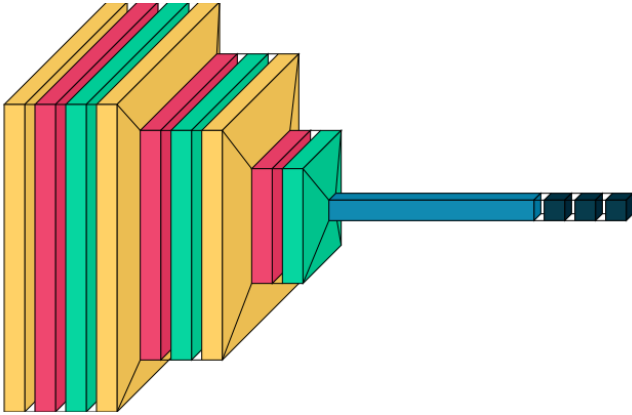
Για την αντιμετώπιση του overfitting στο αρχικό μοντέλο, έγιναν δοκιμές με την εφαρμογή τεχνικών κανονικοποίησης (regularization techniques). Μια τέτοια τεχνική είναι η Dropout [10]. Το Dropout είναι μια μέθοδος στην οποία σε κάθε στάδιο εκπαίδευσης επιλέγονται τυχαία μεμονωμένοι κόμβοι, οι οποίοι απορρίπτονται από το δίκτυο με πιθανότητα 1-p είτε διατηρούνται με πιθανότητα p, με αποτέλεσμα κάθε φορά η δομή του δικτύου να διαφέρει. Οι είσοδοι και εξοδοί σε αυτούς του κόμβους επίσης αφαιρούνται.



Εικ. 2. Μέθοδος Dropout

Στο αρχικό μοντέλο ύστερα από κάθε επίπεδο MaxPooling προστέθηκε και ένα επίπεδο Dropout το οποίο θα αγνοεί το 25% των κόμβων και παράλληλα τις ακμές του. Είναι φανερό

πως το φαινόμενο του overfitting αντιμετωπίστηκε αποτελεσματικά (Εικ. 4) με τη χρήση της παραπάνω μεθόδου. Η απόδοση του μοντέλου έχει ανοδική τάση με το πέρασμα των κύκλων εκπαίδευσης και στα δύο σύνολα δεδομένων, χωρίς μεγάλες αποκλίσεις.



Εικ. 3. Αρχιτεκτονική Συνελκτικού Μοντέλου από το Μηδέν (κίτρινο: Convolutional, κόκκινο: Max Pooling, πράσινο: Dropout, γαλάζιο: Flatten, μπλε: Dense)

2) *Pre-Trained Μοντέλο MobileNet-v2*: Η υλοποίηση του MobileNet-v2 [11] βασίζεται στο δομικό σχήμα του μπλοκ συμφόρησης με κατάλοιπο (bottleneck residual block), που περιλαμβάνει ένα επίπεδο επέκτασης, το οποίο χρησιμοποιεί συνέλιξη ανά σημείο για να επεκτείνει τα χαρακτηριστικά εισόδου σε μεγαλύτερο βάθος ανάλυσης, μια διαχωρίσιμη ανά κανάλι συνέλιξη και ένα επίπεδο γραμμικής συμφόρησης. Κάθε επιμέρους επίπεδο ακολουθείται από ένα επίπεδο κανονικοποίησης και, πλην του επιπέδου γραμμικής συμφόρησης, από ένα επίπεδο ενεργοποίησης με Ανορθωμένη Γραμμική Μονάδα. Το μέγεθος των εικόνων εισόδου που δέχεται είναι 224×224 pixels.

3) *Pre-Trained Μοντέλο Inceptionv3 (GoogleNet)*: Το Inceptionv3 [12] βασίζεται σε μια αρχιτεκτονική εντοπισμού αντικειμένων σε εικόνες χρησιμοποιώντας συνελίξεις kernel διαφορετικών μεγεθών, το Inception Module. Τα kernel μεγεθών 1×1 ή 3×3 εντοπίζουν πιο εύκολα την πληροφορία όταν είναι συγκεντρωμένη σε μια περιοχή ενώ τα μεγαλύτερα είναι προτιμότερα για χρήση σε εικόνες με πληροφορία περισσότερη διάχυτη σε ολόκληρη την απεικόνιση. Έτσι οι συγγραφείς σχεδίασαν το Inception Module με τη λογική ότι θα εκτελούνται στο ίδιο επίπεδο συνελίξεις με φίλτρα 1×1 , 3×3 και 5×5 παράλληλα με Max Pooling και οι έξοδοί τους θα συνενώνονται, ενώ η τελική έξοδος θα στέλνεται στο επόμενο Module. Το τελικό μοντέλο που παρουσίασαν οι συγγραφείς αποτελείται από 9 στοιβαγμένα Inception Modules με συνολικό βάθος 27 επίπεδα και το ονόμασαν GoogLeNet.

4) *Pre-Trained Μοντέλο ResNet50 (Residual Network)*: Η αρχιτεκτονική του ResNet [13] επιχειρεί να διευθύνει το πρόβλημα της εκπαίδευσης συνελκτικών νευρωνικών δικτύων μεγαλύτερου βάθους, αξιοποιώντας μια δομή υπολειπόμενης (residual) μάθησης. Ο σχεδιασμός τους είναι απλός, απλά πολύ βαθύς. Ακολουθεί ένα βασικό σχέδιο, όπου σχεδόν όλα τα επίπεδα συνέλιξης εφαρμόζουν φίλτρα (kernels) 3×3 , και ανά τακτά σημεία υποδιπλασιάζεται το

χωρικό μέγεθος, ενώ διπλασιάζεται το πλήθος των φίλτρων (σε πλήθος καναλιών), κρατώντας περίπου την ίδια πολυπλοκότητα σε κάθε επίπεδο. Υπάρχει, ωστόσο, μια διαφοροποίηση στις εκδόσεις της αρχιτεκτονικής με μεγαλύτερο βάθος, όπως του ResNet50 που χρησιμοποιείται στην παρούσα εργασία. Χρησιμοποιείται ένα διαφορετικό δομικό μπλοκ, που ονομάζεται bottleneck, όμοιο με αυτό του MobileNet-v2, με σκοπό να βελτιωθεί ο χρόνος εκπαίδευσης.

5) *Pre-Trained Μοντέλο EfficientNetB7*: Το EfficientNet [14] χρησιμοποιεί μια τεχνική που ονομάζεται compound coefficient για να κλιμακώνει τα μοντέλα με απλό αλλά αποτελεσματικό τρόπο. Αντί να κλιμακώνονται τυχαία το πλάτος, το βάθος ή η ανάλυση, και τα τρία μαζί κλιμακώνονται ομοιόμορφα με ένα συγκεκριμένο σταθερό σύνολο συντελεστών κλιμάκωσης. Χρησιμοποιώντας τη μέθοδο κλιμάκωσης και AutoML (αυτοματοποιημένη μηχανική μάθηση), οι συγγραφείς του ανέπτυξαν επτά μοντέλα διαφόρων διαστάσεων, τα οποία ξεπέρασαν την ακρίβεια των περισσότερων συνελκτικών νευρωνικών δικτύων και με πολύ καλύτερη απόδοση. Στην παρούσα εργασία θα χρησιμοποιηθεί το μοντέλο B7.

III. ΑΠΟΤΕΛΕΣΜΑΤΑ

Σε αυτήν την εργασία όπως προαναφέρθηκε αναπτύχθηκε ένα CNN εξ' ολοκλήρου από την αρχή, το οποίο θα αναφέρεται παρακάτω ως Simple CNN και τέσσερα Pre-Trained CNN μοντέλα που προσαρμόστηκαν στο παρόν πρόβλημα ταξινόμησης ιστοπαθολογικών εικόνων, για το αν εμφανίζεται στον πνεύμονα κάποιο είδος καρκίνου όπως αδενοκαρκίνωμα ή καρκίνωμα πλακωδών κυττάρων ή υπάρχει καλοήθεια. Στα Pre-Trained μοντέλα το ανώτατο επίπεδο τους αντικαταστάθηκε από ένα Dense επίπεδο με 3 κόμβους και συνάρτηση ενεργοποίησης τη softmax, κατάλληλο για την έξοδο του προβλήματος. Επίσης απενεργοποιήθηκε η εκπαίδευση των υπολοίπων επιπέδων τους, με σκοπό να διατηρηθούν τα βάρη τους ως έχουν.

Στην εκπαίδευση χρησιμοποιήθηκαν τα σύνολα εκπαίδευσης και επικύρωσης και η αξιολόγηση (evaluation) των μοντέλων έγινε στο σύνολο ελέγχου. Το σφάλμα (loss) και η ακρίβεια (accuracy) ήταν τα κριτήρια για την επιλογή του βέλτιστου μοντέλου. Στους παρακάτω πίνακες εμφανίζονται οι τιμές τους και στα τρία σύνολα, για κάθε μοντέλο.

ΠΙΝΑΚΑΣ Ι. ΤΙΜΕΣ ΣΦΑΛΜΑΤΟΣ ΜΟΝΤΕΛΩΝ

Μοντέλα	Train Loss	Validation Loss	Test Loss
Simple Cnn	0.145	0.242	0.244
MobileNet-v2	0.029	0.048	0.057
Inceptionv3	0.047	0.062	0.068
ResNet50	0.034	0.054	0.052
EfficientNetB7	0.043	0.068	0.087

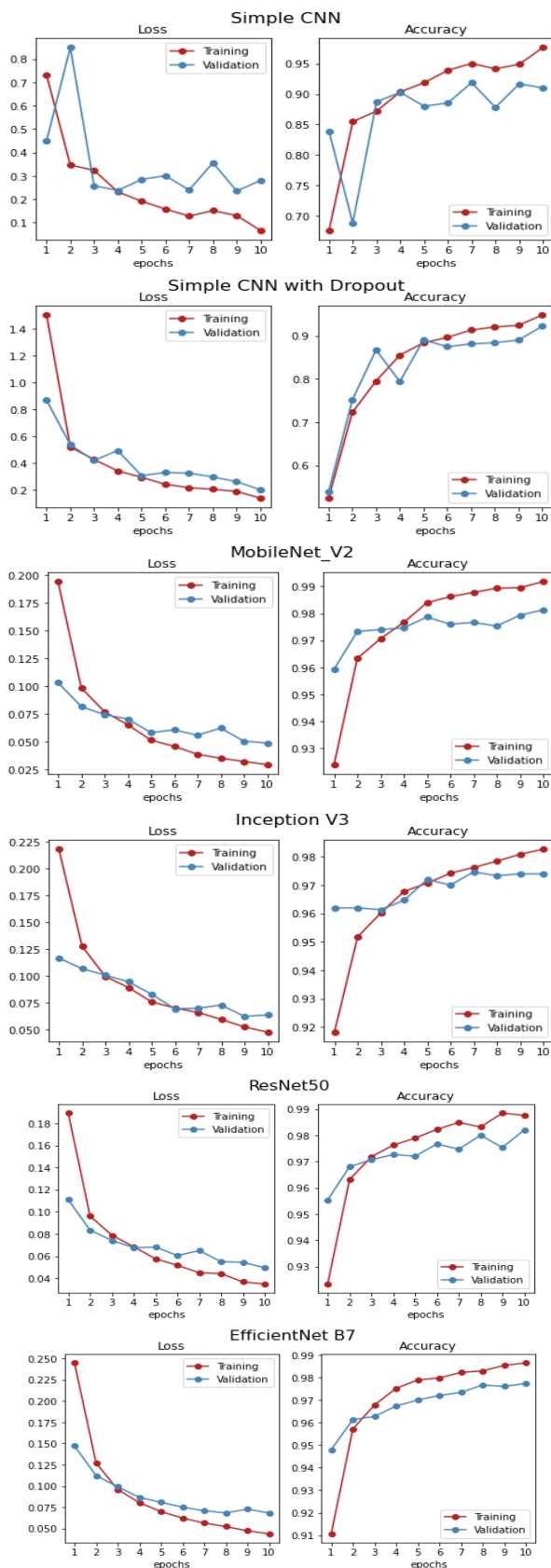
ΠΙΝΑΚΑΣ ΙΙ. ΤΙΜΕΣ ΑΚΡΙΒΕΙΑΣ ΜΟΝΤΕΛΩΝ

Μοντέλα	Train Accuracy	Validation Accuracy	Test Accuracy
Simple Cnn	94.39	92.20	91.26
MobileNet-v2	99.18	98.13	98.00
Inceptionv3	98.27	97.40	97.41
ResNet50	98.76	98.20	98.10
EfficientNetB7	98.64	97.73	96.46

Σύμφωνα με τα αποτελέσματα τα Pre-Trained μοντέλα είχαν τις καλύτερες επιδόσεις, αφού η ακρίβεια τους ξεπερνούσε το 97% και το σφάλμα είχε πολύ χαμηλές τιμές.

Ομοίως και το Simple CNN έδειξε θετικά στοιχεία, όμως η απόδοση του είναι κατώτερη των υπολοίπων.

Δημιουργήθηκαν γραφικές παραστάσεις με την ακρίβεια στο σύνολο εκπαίδευσης και επικύρωσης και το σφάλμα αντίστοιχα, σε σχέση με τις εποχές, με σκοπό την οπτική ανάλυση των επιδόσεων των μοντέλων.



Εικ. 4. Επιδόσεις Μοντέλων κατά τη διάρκεια της εκπαίδευσης

Αρχικά είναι εμφανής η αντιμετώπιση του overfitting στο Simple CNN μοντέλο με την προσθήκη της μεθόδου Dropout. Επίσης σε όλα τα υπόλοιπα μοντέλα εμφανίζεται πρόωγη σύγκλιση στη μέγιστη ακρίβεια που μπορούν να επιτύχουν.

ΠΙΝΑΚΑΣ ΙΙΙ ΧΡΟΝΟΙ ΕΚΠΑΙΔΕΥΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΣΕ ΔΕΥΤΕΡΟΛΕΙΤΑ

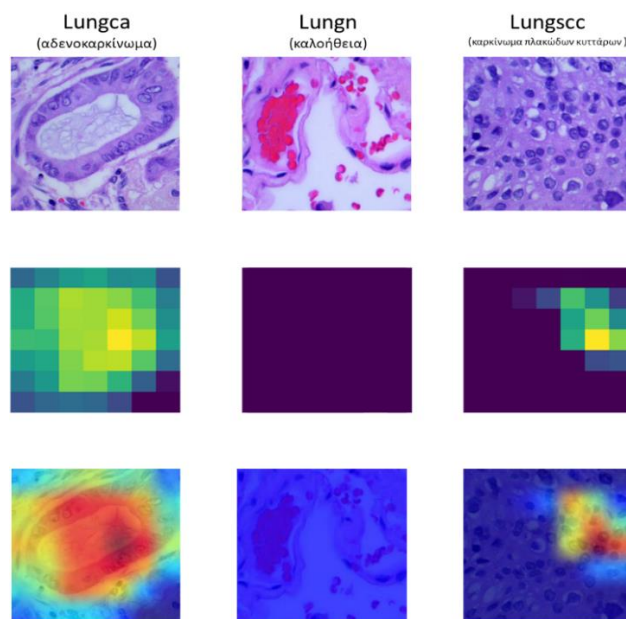
Μοντέλα	Χρόνος Εκπαίδευσης	Χρόνος Πρόβλεψης	Πλήθος Χαρακτηριστικών
Simple CNN	194	5	12.946.883
MobileNet-v2	186	4	2.261.827
Inceptionv3	306	8	21.808.931
ResNet50	252	6	23.567.299
EfficientNetB7	806	19	64.105.363

Όσο αναφορά τους χρόνους εκπαίδευσης και πρόβλεψης των μοντέλων, παρατηρείται ότι όσο αυξάνεται το πλήθος το χαρακτηριστικών που δημιουργούνται στο καθένα, ανάλογα αυξάνονται και οι χρόνοι.

Η επιλογή του προτεινόμενου μοντέλου κρίθηκε στις λεπτομέρειες καθώς οι αποδόσεις τους δεν είχαν μεγάλες διαφορές. Το μοντέλο που επιλέχθηκε είναι το ResNet50 με ακρίβεια 98.10% και τιμή σφάλματος 0.052 στο σύνολο ελέγχου.

Πέρα από την απόδοση, ένα πολύ σημαντικό κομμάτι είναι η ερμηνεία [15] του συνελκτικού νευρωνικού δικτύου για το πώς διαμορφώνει μια απόφαση. Αυτό είναι εφικτό με την παραγωγή οπτικών, επεξηγηματικών χαρτών θερμότητας (heat map), με τον αλγόριθμο Grad-CAM. Το Grad-CAM [16] χρησιμοποιεί τις διαβαθμίσεις κάθε έννοιας του στόχου που ρέουν στο τελευταίο συνελκτικό επίπεδο, για να παράγει ένα χάρτη εντοπισμού που τονίζει τις σημαντικές περιοχές μιας εικόνας, στις οποίες οφείλεται η πρόβλεψη. Βλέποντας που δίνει έμφαση το μοντέλο, πέρα από το πώς λειτουργεί το δίκτυο, μπορεί κάποιος να καταλάβει τις αδυναμίες του μοντέλου.

Στην παρακάτω εικόνα βλέπουμε τα αποτελέσματα που παίρνουμε από τον αλγόριθμο Grad-CAM. Στην πρώτη σειρά βρίσκονται οι πρωτότυπες εικόνες, στη δεύτερη οι χάρτες θερμότητας που δημιουργήθηκαν στο τελευταίο συνελκτικό επίπεδο και τέλος τοποθετείται ο χάρτης θερμότητας στην αρχική εικόνα.



Εικ. 5. Grad-CAM

Για τις κλάσεις, που αφορούν κάποιο τύπο καρκίνου, το μοντέλο εστιάζει στα χαρακτηριστικά του κάθε καρκίνου που απεικονίζεται. Συγκεκριμένα το μοντέλο καταφέρει να αναγνωρίσει τους κακοήθεις όγκους, στην κλάση των πλακωδών καρκινικών κυττάρων. Τους ξεχωρίζει από το αδενοκαρκίνωμα, το οποίο εμφανίζεται στις βλεννώδεις μεμβράνες και φαίνεται ως ένας παχύρευστος λευκός βλεννογόνο που μοιάζει με πλάκα. Τέλος σύμφωνα με τον αλγόριθμο Grad-CAM, αν το μοντέλο δεν εντοπίσει στην εικόνα κάποιο από τα παραπάνω χαρακτηριστικά, την κατατάσσει σε καλοήθεια.

Confusion Matrix ResNet50

Προαγματικές Τιμές	Adenocarcinoma	503	0	19
	Benign	0	473	0
	Squamous Cell Cancer	26	0	479
		Adenocarcinoma	Benign	Squamous Cell Cancer
		Προβλέψεις		

Εικ. 6. Πίνακας Σύγχυσης για τις Προβλέψεις του Μοντέλου

Τα δεδομένα του προβλήματος αφορούν ιατρικά δεδομένα. Αυτό σημαίνει πως δεν υπάρχει περιθώριο λάθους, διότι μία λάθος πρόβλεψη, θα μπορούσε να αποβεί μοιραία για τη ζωή ενός ασθενούς. Από τον παραπάνω πίνακα σύγχυσης, παρατηρείται, ότι το μοντέλο ταξινομεί τους καλοήθεις όγκους με ακρίβεια 100%. Όμως για τους δύο τύπους καρκίνου, υπάρχουν δεδομένα που προβλέφθηκαν με τον λάθος τύπο. Αυτό το σενάριο είναι ευνοϊκό, διότι ο ασθενής έχει θεωρηθεί ορθώς ως καρκινοπαθής, αλλά θα πρέπει να υπάρξουν επαναληπτικές εξετάσεις, με σκοπό την επαλήθευση του τύπου καρκίνου. Το χειρότερο σενάριο, δηλαδή ο ασθενής να θεωρηθεί υγιής, ενώ στην πραγματικότητα δεν είναι, δεν αποτελεί πρόβλημα για το συγκεκριμένο μοντέλο.

IV. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία υλοποιήθηκε ένας αλγόριθμος CAD, για την ανίχνευση αδενοκαρκινώματος, καρκίνου πλακωδών κυττάρων ή καλοήθειας στους πνεύμονες. Για την δημιουργία του, δοκιμάστηκαν μοντέλα που βασίζονται στα συνελκτικά νευρωνικά δίκτυα, τα οποία είναι ικανά να ταξινομούν εικόνες. Στα δεδομένα έγινε η προεπεξεργασία που απαιτείται, για τη βελτιστοποίηση της απόδοσης του μοντέλου, όπως αλλαγή μεγέθους της εικόνας και κανονικοποίηση των τιμών των πινάκων τους, καθώς δοκιμάστηκαν και τεχνικές επαύξησης, στο σύνολο εκπαίδευσης. Στην πειραματική διαδικασία αναπτύχθηκε ένα μοντέλο από το μηδέν και δοκιμάστηκαν προ εκπατευμένα μοντέλα, καταλήγοντας στο συμπέρασμα πως τα Pre-Trained μοντέλα είναι τα κατάλληλα για τη δημιουργία του CAD αλγορίθμου. Χρησιμοποιώντας την προ υπάρχουσα γνώση των προεκπαιδευμένων μοντέλων και προσαρμόζοντάς τη στο παρόν πρόβλημα ταξινόμησης, επιτεύχθηκαν πολύ υψηλές αποδόσεις στην μετρική της ακρίβειας και πολύ χαμηλές τιμές στην συνάρτηση σφάλματος, βάσει των

οποίων έγινε η επιλογή του προτεινόμενου μοντέλου. Σύμφωνα με αυτά τα κριτήρια, το προτεινόμενο μοντέλο είναι το ResNet50 με ακρίβεια 98.10% και τιμή σφάλματος 0.052 στο σύνολο ελέγχου. Με τη βοήθεια του αλγορίθμου Grad-CAM, ήταν εφικτή η ερμηνεία του μοντέλου, δείχνοντας πως λειτουργεί σωστά, εστιάζοντας ως επί το πλείστον στο ground truth της εικόνας. Τέλος, σε γενικότερο πλαίσιο, καταφέρει να κατατάσσει με 100% επιτυχία, αν ένας ασθενής είναι καρκινοπαθής ή υγιής.

Σαν μελλοντική εργασία, θα δοκιμαστούν περισσότερες αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων, προ εκπαιδευμένα και μη, στο παρόν πρόβλημα, για να ελεγχθεί αν επιδέχεται ακόμη μεγαλύτερη βελτίωση στις προβλέψεις. Επίσης το μοντέλο θα μπορούσε να εκπαιδευτεί πάνω σε περισσότερες εικόνες που αφορούν διαφορετικούς τύπους καρκίνου σε περισσότερα όργανα, δίχως να περιορίζεται μόνο στον πνεύμονα.

V. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Jinn-Yi Yeh, Siwa Chan, "CNN-Based CAD for Breast Cancer Classification in Digital Breast Tomosynthesis", ICGSP'18: Proceedings of the 2nd International Conference on Graphics and Signal Processing, October 2018.
- [2] Sun Young Park, Dusty Sargent, "Colonoscopic polyp detection using convolutional neural networks", Conference: SPIE Medical Imaging, March 2016.
- [3] Shorten, C., Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", J Big Data 6, 60 (2019).
- [4] Keiron O'Shea, Ryan Nash, "An Introduction to Convolutional Neural Networks", December 2015.
- [5] Nirthika, R., Manivannan, S., Ramanan, A. et al. "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study". Neural Comput & Applic 34, 5321–5347 (2022).
- [6] Basha S.S., et al., "Impact of fully connected layers on performance of convolutional neural networks for image classification, Neurocomputing", 378 (2020).
- [7] M. Shaha and M. Pawar, "Transfer Learning for Image Classification", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 656-660, doi: 10.1109/ICECA.2018.8474802.
- [8] Kingma, Diederik P. and Jimmy Ba. "Adam: A Method for Stochastic Optimization." CoRR abs/1412.6980 (2015).
- [9] Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee, "Understanding deep neural networks with rectified linear units", arXiv preprint arXiv:1611.01491, 2016.
- [10] Shaofeng Cai, Yao Shu, Gang Chen, Beng Chin Ooi, Wei Wang, Meihui Zhang, "Effective and Efficient Dropout for Deep Convolutional Neural Networks", July 2020.
- [11] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", March 2019.
- [12] C. Szegedy et al., "Going deeper with convolutions", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [14] Mingxing Tan, Quoc Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6105-6114, 2019.
- [15] Q. Zhang, Y. N. Wu and S. -C. Zhu, "Interpretable Convolutional Neural Networks", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8827-8836, doi: 10.1109/CVPR.2018.00920.
- [16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.", 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.