



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Αξιολόγηση της απόδοσης Συνελκτικών Νευρωνικών Δικτύων
κατά την ανάγνωση και κατηγοριοποίηση αντικειμένων**

Μαρκαντώνης Δημήτριος

A.M. ice19390135

Εισηγητής: Ηλίας Δρίτσας

Αθήνα 2024



**UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF INFORMATICS AND COMPUTER ENGINEERING**

DIPLOMA THESIS

Performance evaluation of Convolutional Neural Networks in object detection and classification.

MARKANTONIS DIMITRIOS

A.M. ice19390135

SUPERVISOR:

Elias Dritsas, Appointed Lecturer

Examination Committee:

Maria Trigka, Appointed Lecturer
Foivos Mylonas, Associate Professor

Examination Date: 10/10/2024

ATHENS 2024

**Αξιολόγηση της απόδοσης Συνελκτικών Νευρωνικών Δικτύων
κατά την ανίχνευση και κατηγοριοποίηση
αντικειμένων**

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η πτυχιακή/διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

A/a	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Ηλίας Δρίτσας	Εντεταλμένος Διδάσκων	
2	Μαρία Τρίγκα	Εντεταλμένη Διδάσκουσα	
3	Φοίβος Μυλωνάς	Αναπληρωτής Καθηγητής	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΡΟΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Μαρκαντώνης Δημήτριος του Αντωνίου, με αριθμό μητρώου 19390135 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Μαρκαντώνης Δημήτριος



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη σε όσους με στήριξαν καθ' όλη τη διάρκεια αυτού του ταξιδιού:

Τον Ηλία Δρίτσα, Εντεταλμένο Διδάσκοντα του Τμήματος Μηχανικών Πληροφορικής & Υπολογιστών, για την ανεκτίμητη καθοδήγηση, τις ιδέες και την ενθάρρυνσή του καθ' όλη τη διάρκεια της ερευνητικής διαδικασίας. Η τεχνογνωσία και οι συμβουλές του βοήθησαν στη διαμόρφωση αυτής της διπλωματικής εργασίας.

Τη Μαρία Τρίγκα, Εντεταλμένη Διδάσκουσα του Τμήματος Μηχανικών Πληροφορικής & Υπολογιστών, για τη συνεχή υποστήριξή της και τη χρήσιμη ανατροφοδότηση. Η συμβολή της ήταν ουσιαστική για την τελειοποίηση της κατεύθυνσης και της δομής αυτής της εργασίας.

Τον Φοίβο Μυλωνά, Αναπληρωτή Καθηγητή του Τμήματος Μηχανικών Πληροφορικής & Υπολογιστών, για την επίβλεψη της διαδικασίας και ως μέλος της τελικής επιτροπής αξιολόγησης. Οι γνώσεις και η αξιολόγησή σας βοήθησαν στην ολοκλήρωση αυτού του έργου.

Τέλος θα ήθελα να ευχαριστήσω το Πανεπιστήμιο Δυτικής Αττικής, για την παροχή των πόρων και του περιβάλλοντος που μου επέτρεψαν να συνεχίσω και να ολοκληρώσω αυτό το έργο.

Σας ευχαριστώ όλους για την υποστήριξη και την πίστη σας στο έργο μου.

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία διερευνά την εφαρμογή μοντέλων βαθιάς μάθησης για την ανίχνευση καρκίνου σε εικόνες αξονικής τομογραφίας, εστιάζοντας σε δύο εξέχουσες αρχιτεκτονικές νευρωνικών δικτύων: VGG και EfficientNetB3. Η μελέτη αποσκοπεί στην αξιολόγηση της απόδοσης αυτών των μοντέλων υπό διάφορες συνθήκες, συμπεριλαμβανομένης της εισαγωγής θορύβου και της χρήσης τεχνικών αποθορυβοποίησης. Το σύνολο δεδομένων που χρησιμοποιήθηκε περιλαμβάνει 15.000 εικόνες αξονικής τομογραφίας που κατηγοριοποιούνται σε τρεις κατηγορίες: καλοήθης ιστός, αδενοκαρκίνωμα και πλακώδες καρκίνωμα. Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν βασικές μετρικές επιδόσεων, όπως η ακρίβεια, η ανάκληση, η τομή πάνω από την ένωση (IoU) και η απώλεια κατηγορικής διασταυρούμενης εντροπίας.

Η έρευνα έδειξε ότι τόσο το μοντέλο VGG όσο και το μοντέλο EfficientNetB3 είναι αποτελεσματικά στην ταξινόμηση εικόνων καρκίνου του πνεύμονα, αν και η απόδοσή τους ποικίλλει ανάλογα με την παρουσία θορύβου και τις τεχνικές φιλτραρίσματος που εφαρμόζονται. Το EfficientNetB3 υπερέιχε σταθερά έναντι του VGG όσον αφορά την ακρίβεια. Ο γκαουσιανός θόρυβος αποδείχθηκε πιο επιβλαβής στην απόδοση του μοντέλου από τον κοκκώδη. Το φίλτρο κινητού μέσου όρου αποδείχθηκε πιο αποτελεσματικό στη μείωση του γκαουσιανού θορύβου από το μέσο φίλτρο, ενώ το φίλτρο διαμέσου ήταν αποτελεσματικότερο στον κοκκώδη θόρυβο. Τα ευρήματα αυτά υποδηλώνουν ότι η απόδοση των μοντέλων αυξάνεται όταν βρίσκονται σε συνδυασμό με κατάλληλες μεθόδους χειρισμού του θορύβου. Αυτή η διπλωματική εργασία αποτελεί μια πολλά υποσχόμενη προσέγγιση για τη βελτίωση της ακρίβειας της ταξινόμησης εικόνων καρκίνου του πνεύμονα.

ABSTRACT

This thesis investigates the application of deep learning models for the detection of cancer in CT scan images, focusing on two prominent neural network architectures: VGG and EfficientNetB3. The study aims to evaluate the performance of these models under various conditions, including the introduction of noise and the use of denoising techniques. The dataset utilized comprises 15,000 CT scan images categorized into three classes: benign tissue, adenocarcinoma, and squamous cell carcinoma. Key performance metrics such as accuracy, recall, Intersection over Union (IoU), and categorical cross-entropy loss were employed to assess the models.

The study showed that both the VGG and EfficientNetB3 models are effective in classifying lung cancer images, although their performance varies depending on the presence of noise and the filtering techniques applied. EfficientNetB3 consistently outperformed VGG in terms of accuracy. Gaussian noise proved to be more detrimental to model performance than granular noise. The moving average filter proved more effective in reducing Gaussian noise than the median filter, while the median filter was more effective on granular noise. These findings suggest that the performance of the models increases when they are coupled with appropriate noise handling methods. This thesis is a promising approach to improve the accuracy of lung cancer image classification.

Πίνακας περιεχομένων

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή	12
1.1 Καρκίνος του Πνεύμονα	12
1.1.1 Σύντομη επισκόπηση του καρκίνου του πνεύμονα.....	12
1.1.2 Τύποι καρκίνου του πνεύμονα.....	12
1.1.3 Παράγοντες που επηρεάζουν τον καρκίνο του πνεύμονα.....	13
1.1.4 Προκλήσεις στην έγκαιρη ανίχνευση και διάγνωση	15
1.1.5 Ρόλος της ιατρικής απεικόνισης στη διάγνωση του καρκίνου.....	16
1.2 Αναγνώριση Εικόνας (Image Recognition)	17
1.2.1 Σύντομη επισκόπηση στην αναγνώριση εικόνας.....	17
1.2.2 Περιπτώσεις χρήσης αναγνώρισης εικόνας	17
1.2.3 Μέθοδοι και τεχνικές αναγνώρισης εικόνας.....	19
1.3 Βαθιά Μάθηση (Deep Learning)	23
1.3.1 Εισαγωγή στην βαθιά μάθηση	23
1.3.2 Η άνοδος της βαθιάς μάθησης	23
ΚΕΦΑΛΑΙΟ 2: Μεθοδολογίες	25
2.1 Μοντέλα CNN και Μεθοδολογίες	25
2.1.1 Ανασκόπηση μοντέλων CNN.....	25
2.1.2 Βασικά στοιχεία της αρχιτεκτονικής μοντέλων CNN.....	25
2.1.3 Δεδομένα και διαχωρισμός (Εκπαίδευση, Επικύρωση, Δοκιμή)	28
2.1.4 Πλεονεκτήματα των CNN από παραδοσιακές μεθόδους αναγνώρισης εικόνας.....	29
2.2 Προ-εκπαιδευμένες αρχιτεκτονικές CNN	30
2.2.1 Εισαγωγή	30
2.2.2 Οι Δημοφιλέστερες αρχιτεκτονικές CNN	32
2.2.3 Μάθηση Μεταφοράς.....	38
2.2.4 Λεπτομερής ρύθμιση (Fine-Tuning) μέσω Transfer Learning.....	38
2.3 Δεδομένα και επεξεργασία	39
2.3.1 Προεπεξεργασία και Ενίσχυση (Augmentation) Δεδομένων	39
2.3.2 Τεχνικές Προεπεξεργασίας (Preprocessing Techniques).....	40
2.3.3 Τεχνικές Ενίσχυσης (Augmentation Techniques)	41
2.4 Αποδοτικότητα μοντέλου (Model Performance)	44
2.4.1 Αξιολόγηση μοντέλου (Model Evaluation)	44
2.4.2 Μετρικές Αξιολόγησης (Evaluation Metrics).....	45
2.4.3 Συναρτήσεις Απώλειας (Loss Functions).....	48
2.4.4 Αλγόριθμοι βελτιστοποίησης (Optimization Algorithms).....	50

ΚΕΦΑΛΑΙΟ 3: Προτεινόμενο Πλαίσιο Διάγνωσης Ασθένειας με χρήση CNN Δικτύων σε Ιατρικές Εικόνες.....	52
3.1 Επισκόπηση Πλαισίου	52
3.1.1 Υπόβαθρο και Κίνητρο	52
3.1.2 Σημασία απαλοιφής θορύβου.....	52
3.1.3 Σενάρια Χρήσης του Πλαισίου Διάγνωσης Ασθένειας	53
3.2.1 Δεδομένα και Προεπεξεργασία	54
3.2.2 Μοντέλα και Αρχιτεκτονικές	55
3.3 Μετρικές αξιολόγησης.....	59
3.4 Βιβλιοθήκες και Χαρακτηριστικά υπολογιστικού συστήματος.....	60
Κεφάλαιο 4: Αποτελέσματα Αξιολόγησης Πλαισίου	62
4.1 Σύγκριση VGG και EfficientNetB3 στα αρχικά δεδομένα (Raw Unprocessed Data): ...	62
4.2 Αποτελέσματα με προσθήκη γκαουσιανού (Gaussian) και κοκκώδη (Granular) θορύβου	65
4.3 Αποτελέσματα με απαλοιφή Gaussian θορύβου μέσω Median και Moving Average φίλτρου.....	68
4.4 Αποτελέσματα με απαλοιφή Granular θορύβου μέσω Median και Moving Average φίλτρου.....	71
4.5 Αποτελέσματα με διαδοχική απαλοιφή Gaussian θορύβου μέσω Moving Average φίλτρου και Granular θορύβου μέσω Median Φίλτρου	74
4.6 Σύνοψη και Αποτίμηση Αποτελεσμάτων	77
4.7 Περιορισμοί	82
ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα και Μελλοντικές κατευθύνσεις.....	84
5.1 Συμπεράσματα και μελλοντικές εφαρμογές.....	84
5.2 Μελλοντικές Κατευθύνσεις	85
ΒΙΒΛΙΟΓΡΑΦΙΑ	86
ΠΑΡΑΡΤΗΜΑ	91

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 2.1.2.1: Βασικά Χαρακτηριστικά Αρχιτεκτονικής CNN [25].....	26
Εικόνα 2.2.2.1: Οπτικοποίηση Αρχιτεκτονικής VGG16 [29]	33
Εικόνα 2.2.2.2: Οπτικοποίηση Αρχιτεκτονικής VGG19 [29]	35
Εικόνα 2.2.2.3: Οπτικοποίηση Αρχιτεκτονικής Resnet [29]	36
Εικόνα 2.2.2.4: Οπτικοποίηση LRN [30]	38
Εικόνα 2.3.3.1: Προσθήκη 2D Θορύβου σε εικόνα που ακολουθεί την γκαουσιανή κατανομή [40] ..	43
Εικόνα 2.3.3.2: Διαδικασία Median Filtering σε πίνακα [41]	44
Εικόνα 2.3.3.3: Εφαρμογή Median Φίλτρου σε MRI ανθρώπινου εγκεφάλου με σκοπό την αφαίρεση του τύπου θορύβου αλατιού και πιπεριού (salt – pepper noise) [42].....	44
Εικόνα 2.3.3.4: Διαδικασία Mean Filtering σε πίνακα [44].....	45
Εικόνα 2.4.2.1: AUC Περιοχή κάτω από την καμπύλη ROC (Receiver Operating Characteristic Curve) [48].....	47
Εικόνα 2.4.2.2: Οπτικοποίηση πίνακα σύγχυσης (Confusion Matrix) [50].....	49
Εικόνα 2.4.2.3: Οπτικοποίηση Intersection over Union – IoU [52]	49
Εικόνα 2.4.3.1: Οπτικοποίηση SoftMax Loss.....	50
Εικόνα 2.4.4.1: Οπτικοποίηση της Gradient Descent, ξεκινά από ένα σημείο υψηλής απώλειας και μέσω πολλαπλών επαναλήψεων, κάνει βήματα προς την κατεύθυνση της χαμηλότερης απώλειας, με στόχο να βρει τη βέλτιστη διαμόρφωση βάρους [57].	52
Εικόνα 3.2.1: Διάγραμμα Ροής Εργασιών	55
Εικόνα 3.2.2.1: [63] a) Αρχιτεκτονική EfficientNetB3 b) Compound Scaling.	59
Εικόνα 4.1.1: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell	63
Εικόνα 4.1.2: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell	63
Εικόνα 4.1.3: Πίνακας Σύγχυσης VGG με Original Dataset	64
Εικόνα 4.1.4: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell	64
Εικόνα 4.1.5: Πίνακας Σύγχυσης EfficientNetB3 με Original Dataset	65
Εικόνα 4.1.6: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές.....	65
Εικόνα 4.2.1: Gaussian Noised Samples: Adenocarcinoma - Benign Tissue - Squamous Cell	66
Εικόνα 4.2.2: Granular Noised Samples: Adenocarcinoma - Benign Tissue - Squamous Cell	66
Εικόνα 4.2.3: Πίνακας Σύγχυσης Gaussian – Granular θορύβου	66
Εικόνα 4.2.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές για Gaussian noise.....	67
Εικόνα 4.2.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές για Granular noise	67
Εικόνα 4.3.1: Median Filtered Gaussian Noised Samples	69
Εικόνα 4.3.2: Moving Average Filtered Gaussian Noised Samples	69
Εικόνα 4.3.3: Πίνακας Σύγχυσης απαλοιφής Gaussian θορύβου Median - Moving Average Filtere ..	69
Εικόνα 4.3.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Median Filter	70
Εικόνα 4.3.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter	70
Εικόνα 4.4.1: Median Filtered Granular Noised Samples Adenocarcinoma - Benign Tissue - Squamous Cell	72
Εικόνα 4.4.2: Moving Average Filtered Granular Noised Samples.....	72
Εικόνα 4.4.3: Πίνακας Σύγχυσης απαλοιφής Granular θορύβου Median - Moving Average Filtered.72	72
Εικόνα 4.4.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Median Filter	73
Εικόνα 4.4.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter	73
Εικόνα 4.5.1: Διαδοχική απαλοιφή γκαουσιανού θορύβου με φίλτρο κινητού μέσου όρου - απαλοιφή κοκκώδους θορύβου με φίλτρο διαμέσου.....	75
Εικόνα 4.5.2: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter	75
Εικόνα 4.5.3: Πίνακας Σύγχυσης απαλοιφής Granular θορύβου Median - Moving Average Filtered.76	76

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 4.1.1: Πίνακας απόδοσης VGG για 20 εποχές.	63
Πίνακας 4.1.2: Πίνακας απόδοσης EfficientNetB3 για 20 εποχές.....	65
Πίνακας 4.2.1: Αποτελέσματα απόδοσης για 20 εποχές Gaussian Noise.....	67
Πίνακας 4.2.2: Αποτελέσματα απόδοσης για 20 εποχές Granular Noise.	67
Πίνακας 4.3.1: Αποτελέσματα απόδοσης για 20 εποχές Median Filter σε Gaussian noise	70
Πίνακας 4.3.2: Αποτελέσματα απόδοσης για 20 εποχές Moving Average Filter σε Gaussian noise ...	70
Πίνακας 4.4.1: Αποτελέσματα για 20 εποχές Median Filter σε Granular noise:	73
Πίνακας 4.4.2: Αποτελέσματα για 20 εποχές Moving Average Filter σε Granular noise:	73
Πίνακας 4.5.1: Αποτελέσματα για 20 εποχές Double Filter:	76
Πίνακας 4.6.1: Σύγκριση μοντέλων VGG και EfficientNetB3	78
Πίνακας 4.6.2: Σύγκριση δεδομένων με θόρυβο με αποθορυβοποιημένα δεδομένα.....	80

ΑΚΡΩΝΥΜΙΑ – ΟΡΙΣΜΟΙ

CT	Computed Tomography	Υπολογιστική Τομογραφία
CNN	Convolutional Neural Network	Συνελικτικό Νευρωνικό Δίκτυο
LRN	Local Response Normalization	Τοπική Κανονικοποίηση Απόκρισης
MRI	Magnetic Resonance Imaging	Απεικόνιση Μαγνητικού Συντονισμού
ReLU	Rectified Linear Unit	Διορθωμένη Γραμμική Μονάδα
NSCLC	Non – Small Cell Lung Cancer	Μη Μικροκυτταρικός Καρκίνος του Πνεύμονα
SCLC	Small Cell Lung Cancer	Μικροκυτταρικός Καρκίνος του Πνεύμονα
LDCT	Low Dose - Computed Tomography	Υπολογιστική Τομογραφία Χαμηλής Δόσης
CMS	Centers for Medicare and Medicaid Services	Κέντρα για Υπηρεσίες Medicare και Medicaid
CTLs	Cytotoxic T - Lymphocytes	Κυτταροτοξικά Τ Λεμφοκύτταρα
CDC	Centers for Disease Control and Prevention	Κέντρα Ελέγχου και Πρόληψης Νοσημάτων
EMR	Electronic Medical Record	Ηλεκτρονικός Ιατρικός Φάκελος
NLST	National Lung Screening Trial	Εθνική Δοκιμή Ανίχνευσης Καρκίνου του Πνεύμονα
COVID 19	Corona Virus Disease 2019	Νόσος του Κορωνοϊού 2019
CAD	Computer Aided Design	Σχεδίαση με Βοήθεια Υπολογιστή
MV	Machine Vision	Μηχανική Όραση

Αξιολόγηση της απόδοσης CNN κατά την ανίχνευση και κατηγοριοποίηση αντικειμένων

AI	Artificial Intelligence	Τεχνητή Νοημοσύνη
ML	Machine Learning	Μηχανική Μάθηση
CV	Computer Vision	Υπολογιστική Όραση
ANN	Artificial Neural Network	Τεχνητό Νευρωνικό Δίκτυο
SVM	Support Vector Machine	Μηχανή Υποστήριξης Διανυσμάτων
CPU	Central Processing Unit	Κεντρική Μονάδα Επεξεργασίας
GPU	Graphical Processing Unit	Μονάδα Επεξεργασίας Γραφικών
FC	Fully Connected Layer	Πλήρως Συνδεδεμένο Επίπεδο
RGB	Red Green Blue	Κόκκινο Πράσινο Μπλε
TL	Transfer Learning	Μεταφορά Μάθησης
DNN	Deep Neural Network	Βαθύ Νευρωνικό Δίκτυο
GANS	Generative Adversarial Networks	Γενετικά Ανταγωνιστικά Δίκτυα
AUC	Area Under Curve	Εμβαδόν Υπό Καμπύλης
TN	True Negative	Αληθώς Αρνητικό
TP	True Positive	Αληθώς Θετικό
FP	False Positive	Ψευδώς Θετικό
FN	False Negative	Ψευδώς Αρνητικό
TPR	True Positive Rate	Ρυθμός Αληθώς Θετικών
TNR	True Negative Rate	Ρυθμός Αληθώς Αρνητικών
FPR	False Positive Rate	Ρυθμός Ψευδώς Θετικών
ROC	Receiver - Operating Characteristic Curve	Καμπύλη Δέκτη - Λειτουργού (ROC)
CE	Cross - Entropy	Διασταυρούμενη Εντροπία
SGD	Stochastic Gradient Descent	Στοχαστική Κατάβαση Κλίσης
IoU	Intersection over Union	Διατομή επί Ένωσης
HIPAA	Health Insurance Portability and Accountability Act	Νόμος για τη Φορητότητα και Λογοδοσία Ασφάλισης Υγείας
RMSPROP	Root Mean Squared Propagation	Προώθηση Τετραγωνικής Μέσης Ρίζας
NS	Noise Scale	Κλίμακα Θορύβου
RAM	Random Access Memory	Μνήμη Τυχαίας Προσπέλασης
PET	Positron Emission Tomography	Τομογραφία Εκπομπής Ποζιτρονίων

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

1.1 Καρκίνος του Πνεύμονα

1.1.1 Σύντομη επισκόπηση του καρκίνου του πνεύμονα

Ο καρκίνος του πνεύμονα είναι ένας τύπος καρκίνου που προέρχεται από τους πνεύμονες και χαρακτηρίζεται από την ανεξέλεγκτη ανάπτυξη κυττάρων στον πνευμονικό ιστό. Ο καρκίνος του πνεύμονα είναι ένας από τους πιο κοινούς και σοβαρούς καρκίνους παγκοσμίως. Υπάρχουν δύο κύριοι τύποι: ο μη μικροκυτταρικός καρκίνος του πνεύμονα (NSCLC) και ο μικροκυτταρικός καρκίνος του πνεύμονα (SCLC), ο καθένας με διαφορετικά χαρακτηριστικά και θεραπευτικές επιλογές. Η κύρια αιτία του καρκίνου του πνεύμονα είναι το κάπνισμα, αυτό όμως δεν σημαίνει ότι και οι μη καπνιστές δεν μπορούν να αναπτύξουν τη συγκεκριμένη νόσο λόγω διαφόρων παραγόντων, όπως το αέριο ραδόνιο, ο αμίαντος και η ατμοσφαιρική ρύπανση. Παρόλο που η έγκαιρη ανίχνευση και η πρόοδος στη θεραπεία έχουν βελτιώσει τα αποτελέσματα, ο καρκίνος του πνεύμονα παραμένει μια πολύ σημαντική πρόκληση για την υγεία παγκοσμίως.

Ο καρκίνος του πνεύμονα θεωρείται συχνά ως μια διάγνωση με ένα πιθανό αποτέλεσμα. Ωστόσο, υπάρχουν πολλοί διαφορετικοί τύποι καρκίνου του πνεύμονα και οι αιτίες τους ποικίλλουν σε μεγάλο βαθμό. Υπάρχουν όμως διαφορές στον τρόπο με τον οποίο αναπτύσσεται ο κάθε καρκίνος του πνεύμονα στο σώμα και στον τρόπο αντιμετώπισής του.

Οι περισσότεροι καρκίνοι του πνεύμονα μπορούν να ταξινομηθούν σε διάφορους τύπους και υποτύπους. Ωστόσο, ακόμη και εντός αυτής της νόσου, όλοι οι άνθρωποι που πάσχουν από καρκίνο του πνεύμονα εμφανίζουν διαφορετικά συμπτώματα.

Παρακάτω υπάρχουν περισσότερες πληροφορίες σχετικά με αυτούς τους τύπους καρκίνου του πνεύμονα και πώς διαφέρουν ανάλογα με την ηλικία, το φύλο, την κατάσταση καπνίσματος και άλλους παράγοντες.

1.1.2 Τύποι καρκίνου του πνεύμονα

Κύριοι τύποι καρκίνου του πνεύμονα

Πρωτοπαθής καρκίνος του πνεύμονα είναι εκείνος που ξεκινά στους πνεύμονες και δεν εξαπλώνεται (κάνει μετάσταση) στους πνεύμονες από άλλο σημείο του σώματος.

Αναγνωρίζονται δύο κύριοι τύποι πρωτοπαθών καρκίνων του πνεύμονα:

1. Ο μη μικροκυτταρικός καρκίνος του πνεύμονα (NSCLC) θεωρείται ο πιο κοινός τύπος καρκίνου του πνεύμονα συνολικά, αντιπροσωπεύοντας το 80% έως 85% των καρκίνων του πνεύμονα στις Ηνωμένες Πολιτείες.
2. Ο μικροκυτταρικός καρκίνος του πνεύμονα (SCLC) διαγιγνώσκεται στο 10% έως 15% των περιπτώσεων καρκίνου του πνεύμονα.
3. Ο τρίτος τύπος, που ονομάζεται καρκινοειδής όγκος, είναι λιγότερο συχνός στους πνεύμονες, αντιπροσωπεύοντας μόλις το 1% έως 2% των καρκίνων του πνεύμονα.

Ο NSCLC και ο SCLC αναλύονται περαιτέρω σε συγκεκριμένους υποτύπους, με ονομασίες που βασίζονται στα είδη των κυττάρων που συνθέτουν τους όγκους όταν παρατηρούνται στο μικροσκόπιο [1].

Μη μικροκυτταρικός καρκίνος του πνεύμονα (NSCLC)

Ο NSCLC συνήθως αναπτύσσεται και εξαπλώνεται πιο αργά από τον SCLC. Και οι δύο συνδέονται με το κάπνισμα, αλλά ο NSCLC είναι όμως ο πιο συχνός τύπος καρκίνου του πνεύμονα σε νεότερα άτομα και σε άτομα που δεν έχουν καπνίσει ποτέ.

Οι κίνδυνοι και οι αιτίες μπορεί να διαφέρουν ανάλογα με το ποιος υποτύπος του NSCLC διαγιγνώσκεται. Αναγνωρίζονται τρεις βασικοί υπότυποι:

- Το αδenoκαρκίνωμα του πνεύμονα (Lung adenocarcinoma) αντιπροσωπεύει το 40% έως 50% των περιπτώσεων NSCLC.
- Το καρκίνωμα πλακωδών κυττάρων (Squamous cell carcinoma) εμφανίζεται σε ποσοστό περίπου 20% έως 30% των περιπτώσεων NSCLC.
- Ο ευμεγέθης καρκίνος του πνεύμονα (Large cell carcinoma) ευθύνεται για περίπου 3% έως 10% των περιπτώσεων NSCLC.

Μικροκυτταρικός καρκίνος του πνεύμονα (SCLC)

Ο SCLC θεωρείται συνήθως ένας επιθετικός, ταχέως αναπτυσσόμενος καρκίνος. Συνδέεται στενά με το κάπνισμα, αν και μπορεί να εμπλέκονται και άλλοι παράγοντες, συμπεριλαμβανομένης της έκθεσης σε ραδόνιο. Το ραδόνιο είναι ένα άοσμο, άχρωμο αέριο που διαρρέει στα σπίτια από το περιβάλλον έδαφος.

Αναγνωρίζονται δύο υπότυποι του SCLC:

- Καρκίνωμα μικρών κυττάρων (Small cell carcinoma).
- Συνδυασμένο μικροκυτταρικό καρκίνωμα (Combined Small cell carcinoma): κύτταρα αναμειγμένα με άλλο τύπο [2].

1.1.3 Παράγοντες που επηρεάζουν τον καρκίνο του πνεύμονα

Διαφορές με βάση το φύλο

Υψηλότερα ποσοστά επίπτωσης του καρκίνου του πνεύμονα εξακολουθούν να παρατηρούνται στους άνδρες σε σύγκριση με τις γυναίκες, αν και το χάσμα κλείνει λόγω των αλλαγών στην πολιτική υγείας και της διακοπής του καπνίσματος.

Ο NSCLC διαγιγνώσκεται συνήθως τόσο στους άνδρες όσο και στις γυναίκες, με το αδenoκαρκίνωμα να είναι ο πιο διαδεδομένος υποτύπος. Ωστόσο, τα αδenoκαρκινώματα του πνεύμονα περιγράφονται ως διαφορετική ασθένεια στις γυναίκες και στους άνδρες.

Μεταξύ των γυναικών, τα αδenoκαρκινώματα του πνεύμονα αντιπροσωπεύουν το 40% έως 60% των περιπτώσεων καρκίνου του πνεύμονα. Περίπου οι μισοί από αυτούς τους καρκίνους στις γυναίκες σχετίζονται με το κάπνισμα, σε σύγκριση με το 85% των περιπτώσεων στους άνδρες που είναι καπνιστές. Οι γυναίκες τείνουν να είναι νεότερες κατά τη διάγνωση και ανταποκρίνονται διαφορετικά

στις θεραπείες. Στους άνδρες με NSCLC, είναι πιο πιθανό να αναπτυχθεί καρκίνος του πνεύμονα από πλακώδη κύτταρα, ο οποίος συνδέεται με το κάπνισμα. Ωστόσο, οι άνδρες είναι λιγότερο πιθανό να διαγνωστούν με SCLC από ό,τι οι γυναίκες.

Διαφορές με βάση την ηλικία

Ο καρκίνος του πνεύμονα είναι γενικά πιο συχνός σε άτομα μεγαλύτερης ηλικίας, με τη μέση ηλικία διάγνωσης να είναι τα 70 έτη. Μόνο το 5% όλων των καρκίνων διαγιγνώσκεται σε νέους ενήλικες ηλικίας 18-39 ετών. Το αδενοκαρκίνωμα του πνεύμονα είναι ο συχνότερος τύπος καρκίνου του πνεύμονα στους νέους ενήλικες, αντιπροσωπεύοντας σχεδόν το 50% των περιπτώσεων. Οι καρκινοειδείς όγκοι του πνεύμονα εντοπίζονται συχνότερα σε άτομα ηλικίας 45-55 ετών, αλλά μπορούν να διαγνωστούν σε οποιαδήποτε ηλικία, συμπεριλαμβανομένων των παιδιών και των εφήβων. Ο SCLC είναι σχετικά σπάνιος σε νέους ανθρώπους, παρατηρείται σε λιγότερο από το 5% των περιπτώσεων καρκίνου του πνεύμονα.

Οι νέοι ενήλικες με καρκίνο του πνεύμονα είναι πολύ πιο πιθανό να έχουν γενετικό παράγοντα που συμβάλλει στον καρκίνο τους, γεγονός που οδηγεί σε πιθανή θεραπεία με νεότερες θεραπείες που στοχεύουν σε συγκεκριμένες μεταλλάξεις.

Διαφορές παρατηρούνται μεταξύ καπνιστών και μη καπνιστών, όχι μόνο στον τύπο καρκίνου που παρουσιάζουν αλλά και στις ηλικίες. Ο καρκίνος του πνεύμονα σε μη καπνιστές διαφέρει σημαντικά από τον καρκίνο του πνεύμονα σε άτομα που καπνίζουν, ακόμη και εντός των πιο κοινών τύπων της νόσου.

Τόσο ο NSCLC όσο και ο SCLC συνδέονται με το ιστορικό καπνίσματος, αν και η συσχέτιση είναι πολύ ισχυρότερη για τον SCLC. Από τους τρεις υποτύπους του NSCLC, το αδενοκαρκίνωμα του πνεύμονα είναι το πιο πιθανό να βρεθεί σε άτομα που δεν κάπνιζαν ποτέ (50% έως 60%). Περίπου 10% έως 20% των ατόμων που δεν κάπνιζαν ποτέ διαγιγνώσκονται με πλακώδες καρκίνωμα. Μόνο 6% έως 8% των ατόμων που δεν κάπνιζαν ποτέ διαγιγνώσκονται με SCLC. Οι καρκινοειδείς όγκοι δεν φαίνεται να σχετίζονται με το κάπνισμα και ανευρίσκονται τόσο σε καπνιστές όσο και σε μη καπνιστές σε αριθμούς παρόμοιους με αυτούς που παρατηρούνται στο γενικό πληθυσμό [3].

Συνοψίζοντας, οι σημερινοί και πρώην καπνιστές αναπτύσσουν SCLC και τον πλακώδες κυτταρικό υποτύπο του NSCLC συχνότερα από τους άλλους. Ο υποτύπος αδενοκαρκινώματος του NSCLC στον πνεύμονα παρατηρείται συχνότερα σε γυναίκες, νεότερα άτομα και σε όσους δεν έχουν καπνίσει ποτέ.

Ποσοστά και ο ρόλος της γενετικής

Τα ποσοστά του καρκίνου του πνεύμονα έχουν μετατοπιστεί με την πάροδο του χρόνου, εν μέρει λόγω της διακοπής του καπνίσματος. Ωστόσο, τα αυξανόμενα ποσοστά σε νεότερα άτομα και μη καπνιστές εγείρουν ερωτήματα.

Περιβαλλοντικοί παράγοντες

Περιβαλλοντικοί παράγοντες όπως η ατμοσφαιρική ρύπανση μπορεί να συμβάλλουν, αλλά οι περιπτώσεις παραμένουν ανεξήγητες. Οι έρευνες υπό την ηγεσία του Εθνικού Ινστιτούτου Καρκίνου και των Εθνικών Ινστιτούτων Υγείας υποδεικνύουν τη γενετική. Μια μελέτη του Σεπτεμβρίου του 2021

εξέτασε γονιδιακές αλλαγές σε 232 άτομα με NSCLC που δεν είχαν καπνίσει ποτέ, συγκρίνοντας τον φυσιολογικό ιστό τους με δείγματα όγκων.

Από την ομάδα μελέτης, 189 είχαν αδενοκαρκινώματα πνεύμονα, ενώ οι υπόλοιποι είχαν άλλους τύπους. Οι γονιδιακές μεταλλάξεις που προέρχονταν από φυσικές διεργασίες στο εσωτερικό του σώματος συσχετίστηκαν με τον καρκίνο του πνεύμονα.

Οι ερευνητές εντόπισαν τρεις νέους γενετικούς υποτύπους καρκίνου του πνεύμονα σε αυτούς τους μη καπνιστές, υποδεικνύοντας μελλοντικούς θεραπευτικούς στόχους. Χρειάζονται περισσότερες έρευνες για να επιβεβαιωθούν αυτά τα ευρήματα.

1.1.4 Προκλήσεις στην έγκαιρη ανίχνευση και διάγνωση

Εμπόδια στον έλεγχο

Σε μία έρευνα που πραγματοποιήθηκε στην Αμερική το έτος 2015, το ποσοστό των επιλέξιμων καπνιστών που πέρασαν την εξέταση υπολογιστικής τομογραφίας χαμηλής δόσης (LDCT) αυξήθηκε από 3,3% το 2010 σε 3,9%, σύμφωνα με την Εθνική Έρευνα Συνεντεύξεων Υγείας των ΗΠΑ . Εκείνη την χρονιά, το CMS ενέκρινε την ασφαλιστική κάλυψη της LDCT για ασθενείς υψηλού κινδύνου. Ωστόσο, το 2016, μόνο το 1,9% των ασθενών που πληρούσε τις προϋποθέσεις υποβλήθηκε σε έλεγχο . Μια ανάλυση του Κέντρου Ελέγχου Ασθενειών (CDC) του 2017 σε 10 πολιτείες διαπίστωσε ότι το 12,5% των επιλέξιμων υποψηφίων έλαβαν LDCT κατά το προηγούμενο έτος [4].

Η παραπομπή των ιατρών αποτελεί σημαντικό εμπόδιο λόγω παραγόντων όπως οι πολυάσχολες πρακτικές, η έλλειψη ενημέρωσης για τις κατευθυντήριες γραμμές διαλογής και τα ειδικά για τον ασθενή ζητήματα. Υπάρχουν, όμως και άλλα εμπόδια που περιλαμβάνουν την έλλειψη ειδοποιήσεων από τους ηλεκτρονικούς ιατρικούς φακέλους (EMR), την άρνηση των ασθενών, τους χρονικούς περιορισμούς και τις διαφορετικές γνώσεις των παρόχων σχετικά με την ασφάλιση. Σημαντικούς λόγους αδυναμίας για πρόληψη αποτελούν οι ανησυχίες για το κόστος, ο φόβος των ψευδώς θετικών αποτελεσμάτων και της υπερθεραπείας, η έλλειψη γνώσης των ασθενών, τα ζητήματα πρόσβασης, οι ανησυχίες για την έκθεση στην ακτινοβολία, η ανεπαρκής στελέχωση και οι ασυνεπείς συστάσεις για τον έλεγχο .

Παρά τις βελτιώσεις στις τεχνικές προληπτικού ελέγχου, οι κίνδυνοι που σχετίζονται με την έκθεση στην ακτινοβολία από τις τομογραφίες για την πρόβλεψη καρκίνου του πνεύμονα (CTLS) είναι ελάχιστοι και σχεδόν μη υπαρκτοί. Σε μια μελέτη εκτιμήθηκε ότι ο κίνδυνος εμφάνισης μείζονος καρκίνου κατά τη διάρκεια της ζωής ανά 1000 άτομα που ελέγχονται μετά από 10 ετήσιες LDCT υπολογίζεται σε 0,26 έως 0,81 . Τα ποσοστά ψευδώς θετικών αποτελεσμάτων και υπερδιάγνωσης είναι όμως χαμηλά και δεν πρέπει να αποτρέπουν τον έλεγχο . Η μελέτη NLST διαπίστωσε ότι 779 από τα 1000 άτομα είχαν φυσιολογικές εξετάσεις, ενώ 180 από τα 1000 είχαν ψευδώς θετικά αποτελέσματα, με 13 να χρειάζονται επεμβατικές διαδικασίες και 1 ανά 2500 να εμφανίζει μείζονες επιπλοκές . Τα τυποποιημένα συστήματα ελέγχου θα μπορούσαν να συμβάλουν στη μείωση της υπερδιάγνωσης και των επιπλοκών [5].

Ο προσυμπτωματικός έλεγχος του καρκίνου του πνεύμονα σε μη καπνιστές παραμένει πρόκληση χωρίς τυποποιημένες κατευθυντήριες γραμμές. Αν και ο καρκίνος του πνεύμονα συνδέεται στενά με το κάπνισμα, τα ποσοστά στους μη καπνιστές αυξάνονται. Οι παράγοντες περιλαμβάνουν την ηλικία, το παθητικό κάπνισμα, την περιβαλλοντική έκθεση, το ραδόνιο, γενετικούς παράγοντες, υποκείμενη πνευμονοπάθεια και ογκογόνους ιούς [6].

Έλεγχος στην εποχή του COVID-19

Κατά τη διάρκεια της πανδημίας του COVID-19, το Αμερικανικό Κολλέγιο Ιατρών Θώρακος συνέστησε την καθυστέρηση της εξέτασης για τον καρκίνο του πνεύμονα ώστε να αποφευχθεί η έκθεση. Αυτό οδήγησε σε αύξηση των ασθενών με ύποπτους πνευμονικούς όζους μόλις συνεχίστηκε ο έλεγχος, από 8% σε 29%, προκαλώντας αύξηση των χειρουργικών παραπομπών. Η πανδημία αύξησε τη χρήση της τηλεϊατρικής, βοηθώντας την από κοινού λήψη αποφάσεων για τον έλεγχο του καρκίνου του πνεύμονα [7].

Εξελίξεις στην LDCT

Η ενσωμάτωση της τεχνητής νοημοσύνης στην LDCT εξελίσσεται. Τα πρώιμα μοντέλα τεχνητής νοημοσύνης για την ανίχνευση καρκίνου του πνεύμονα ήταν ισοδύναμα με άλλους ανθρώπους, αλλά είχαν υψηλότερα ποσοστά ψευδώς θετικών αποτελεσμάτων. Τα νεότερα μοντέλα, όπως τα συστήματα ελέγχου με την βοήθεια υπολογιστών (CAD), παρουσιάζουν βελτιωμένη απόδοση, διακρίνοντας μεταξύ διαφόρων τύπων όζων και επιτυγχάνοντας υψηλή ακρίβεια. Η τεχνητή νοημοσύνη μπορεί να εντοπίσει τις ασθένειες που σχετίζονται με το κάπνισμα στην LDCT, διευρύνοντας τον αντίκτυπο του ελέγχου για τον καρκίνο του πνεύμονα [8].

1.1.5 Ρόλος της ιατρικής απεικόνισης στη διάγνωση του καρκίνου

Απεικονιστική εμφάνιση του καρκίνου του πνεύμονα

Η πλειονότητα των ασθενών με καρκίνο του πνεύμονα είναι κλινικά συμπτωματικοί και παρουσιάζουν βήχα, αιμόπτυση, δύσπνοια, θωρακικό άλγος και μη επιλύσιμη πνευμονία. Λιγότερο από το 10% των ασθενών είναι ασυμπτωματικοί όταν ο καρκίνος ανιχνεύεται ως τυχαίο εύρημα.

Ο καρκίνος του πνεύμονα ταξινομείται είτε ως μη μικροκυτταρικός καρκίνος του πνεύμονα (NSCLC) είτε ως μικροκυτταρικός καρκίνος του πνεύμονα, με τον NSCLC να αντιπροσωπεύει τη συντριπτική πλειοψηφία. Η ακτινογραφία θώρακος είναι η πρώτη εξέταση που διενεργείται κατά τη διερεύνηση μιας ύποπτης περίπτωσης καρκίνου του πνεύμονα. Αν και αποτελεί ένα πολύ καλό εργαλείο για την παροχή προκαταρκτικών πληροφοριών σχετικά με τη νόσο, είναι ανεπαρκής για τον βέλτιστο χαρακτηρισμό και τη σταδιοποίηση.

Η αξονική τομογραφία (CT) του θώρακα αποτελεί τον ακρογωνιαίο λίθο της απεικόνισης του καρκίνου του πνεύμονα με βάση την οποία αποφασίζεται η περαιτέρω διαχείριση. Ο πρωτοπαθής όγκος παρουσιάζει ένα ευρύ φάσμα απεικονιστικών εμφανίσεων. Οι μη μικροκυτταρικοί καρκίνοι του πνεύμονα μπορεί να είναι κεντρικά τοποθετημένες μάζες, που εισβάλλουν στις δομές του μεσοθωρακίου ή περιφερικά τοποθετημένες βλάβες που εισβάλλουν στο θωρακικό τοίχωμα. Οι όγκοι

μπορεί να έχουν περιθώρια που είναι ομαλά, λοβωτά ή ακανόνιστα και ακανόνιστα. Μπορεί να είναι ομοιόμορφα συμπαγείς ή να έχουν κεντρική νέκρωση και σπηλαιώση [9].

Οι κεντρικά τοποθετημένοι και οι σπηλαιώδεις όγκοι είναι πιο πιθανό να είναι πλακώδους ιστολογίας. Από την άλλη, ο υποτύπος αδενοκαρκίνωμα παρουσιάζει διάφορες εκδηλώσεις στην αξονική τομογραφία (CT), κάποιες από τις οποίες μιμούνται μη κακοήθεις διεργασίες και συχνά επικαλύπτονται με άλλες νόσους [10].

1.2 Αναγνώριση Εικόνας (Image Recognition)

1.2.1 Σύντομη επισκόπηση στην αναγνώριση εικόνας

Η αναγνώριση εικόνας, στο πλαίσιο της μηχανικής όρασης (Machine Vision – MV), είναι η ικανότητα του λογισμικού να αναγνωρίζει αντικείμενα, μέρη, ανθρώπους, γραφή και ενέργειες σε ψηφιακές εικόνες. Οι υπολογιστές μπορούν να χρησιμοποιήσουν τεχνολογίες μηχανικής όρασης σε συνδυασμό με μια κάμερα και λογισμικό τεχνητής νοημοσύνης (AI) για να επιτύχουν την αναγνώριση εικόνας. Ενώ οι εγκέφαλοι των ζώων και των ανθρώπων αναγνωρίζουν τα αντικείμενα με ευκολία, οι υπολογιστές χρειάζονται καθοδήγηση και προγραμματισμό από τους ανθρώπους για να μπορούν να εκτελέσουν με την ίδια και καλύτερη ακρίβεια την ίδια διαδικασία. Υπάρχουν πολυάριθμοι τρόποι για την πραγματοποίηση επεξεργασίας εικόνας (Image Processing), συμπεριλαμβανομένων των μοντέλων βαθιάς μάθησης (Deep Neural Networks) και μηχανικής μάθησης (Machine Learning – ML). Ωστόσο, η χρησιμοποιούμενη προσέγγιση καθορίζεται από την περίπτωση χρήσης. Για παράδειγμα, οι τεχνικές βαθιάς μάθησης χρησιμοποιούνται συνήθως για την επίλυση πιο σύνθετων προβλημάτων από ό,τι τα μοντέλα μηχανικής μάθησης, όπως η ασφάλεια των εργαζομένων στον βιομηχανικό αυτοματισμό και η ανίχνευση του καρκίνου μέσω της ιατρικής έρευνας.

Τυπικά, η αναγνώριση εικόνων συνεπάγεται τη δημιουργία βαθιών νευρωνικών δικτύων που αναλύουν κάθε εικονοστοιχείο (pixel) εικόνας. Αυτά τα δίκτυα τροφοδοτούνται με όσο το δυνατόν περισσότερες επισημειωμένες (labeled) εικόνες για να εκπαιδευτούν στην αναγνώριση σχετικών εικόνων [11].

1.2.2 Περιπτώσεις χρήσης αναγνώρισης εικόνας

Η αναγνώριση εικόνας χρησιμοποιείται για την εκτέλεση πολλών μηχανικών οπτικών εργασιών (CV Tasks), όπως η επισήμανση του περιεχομένου των εικόνων με μετα-ετικέτες (meta-tags), η αναζήτηση περιεχομένου εικόνας και η καθοδήγηση αυτόνομων ρομπότ, αυτοκινούμενων αυτοκινήτων και συστημάτων αποφυγής ατυχημάτων.

Ακολουθούν ορισμένες από τις βασικότερες περιπτώσεις χρήσης της αναγνώρισης εικόνας στον πραγματικό κόσμο:

- **Αναγνώριση προσώπου.** Η αναγνώριση προσώπου χρησιμοποιείται συχνά στα μέσα κοινωνικής δικτύωσης, στα συστήματα ασφαλείας και στην ψυχαγωγία για την αναγνώριση προσώπων σε φωτογραφίες και βίντεο. Οι αλγόριθμοι βαθιάς μάθησης χρησιμοποιούνται για την αναγνώριση προσώπου για την αξιολόγηση μιας φωτογραφίας ενός ατόμου και τη δημιουργία μιας ακριβούς

ταυτότητας του ατόμου στην εικόνα. Αυτός ο αλγόριθμος μπορεί να επεκταθεί για να εξαγάγει σημαντικά χαρακτηριστικά όπως η ηλικία, το φύλο και η έκφραση του προσώπου ενός ατόμου μέσω εικόνων. Οι δυνατότητες αναγνώρισης προσώπου σε smartphones και η ηλεκτρονική επαλήθευση της αναγνώρισης εικόνας σε σημεία ελέγχου ασφαλείας όπως αεροδρόμια και εισόδους κτιρίων είναι από τις πιο κοινές εφαρμογές αναγνώρισης εικόνας.

- **Οπτική αναζήτηση.** Η αναζήτηση εικόνας με λέξεις-κλειδιά ή οπτικά χαρακτηριστικά χρησιμοποιεί τεχνολογία αναγνώρισης εικόνας. Για παράδειγμα, το Google Lens, μία από τις πρόσφατες καινοτομίες της Google, επιτρέπει στους χρήστες να πραγματοποιούν αναζητήσεις βάσει εικόνας και η εφαρμογή Translate της Google προσφέρει μετάφραση σε πραγματικό χρόνο με σάρωση κειμένου από φωτογραφίες. Αυτές οι τεχνολογικές εξελίξεις επιτρέπουν στους καταναλωτές να πραγματοποιούν αναζητήσεις σε πραγματικό χρόνο. Για παράδειγμα, αν κάποιος βρει ένα λουλούδι σε ένα πικνίκ και ενδιαφέρεται να μάθει περισσότερα γι' αυτό, μπορεί απλά να τραβήξει μια φωτογραφία του λουλουδιού και να χρησιμοποιήσει το διαδίκτυο για να αναζητήσει πληροφορίες γι' αυτό αμέσως.
- **Ιατρική διάγνωση.** Χρησιμοποιώντας την τεχνολογία αναγνώρισης εικόνας, οι επαγγελματίες υγείας και οι κλινικοί γιατροί εξετάζουν ιατρικές εικόνες για να διαγνώσουν ασθένειες και καταστάσεις. Για παράδειγμα, το λογισμικό αναγνώρισης εικόνας μπορεί να εκπαιδευτεί ώστε να αναλύει και να εντοπίζει μοτίβα σε δεδομένα από συσκευές μαγνητικής τομογραφίας ή ακτινών X (CT Scans ή X-Rays). Αυτό επιτρέπει στους κλινικούς γιατρούς όχι μόνο να βρίσκουν και να ανιχνεύουν, αλλά και να αναφέρουν ιατρικές ανωμαλίες σε πρώιμο στάδιο. Τρεις τομείς που χρησιμοποιούν συχνά την αναγνώριση εικόνας για ιατρική διάγνωση είναι η ακτινολογία, η οφθαλμολογία και η παθολογία.
- **Ποιοτικός έλεγχος.** Ο παραδοσιακός χειροκίνητος έλεγχος ποιότητας είναι επίπονη διαδικασία για τους εργαζόμενους, χρονοβόρος και επιρρεπής σε σφάλματα. Παρόλα αυτά, χρησιμοποιώντας ένα σύνολο φωτογραφιών με ετικέτες (ονομασμένες – labeled) του προϊόντος ενδιαφέροντος, ένα μοντέλο τεχνητής νοημοσύνης ή ένα νευρωνικό δίκτυο μπορεί να εκπαιδευτεί ώστε να εντοπίζει αυτόματα μοτίβα δυσλειτουργικού εξοπλισμού. Με την χρησιμοποίηση της Τεχνητής Νοημοσύνης ή του νευρωνικού δικτύου καθίσταται δυνατός ο εντοπισμός και η απομόνωση στοιχείων που δεν πληρούν τα πρότυπα, βελτιώνοντας αμέσως τη συνολική ποιότητα του προϊόντος.
- **Ανίχνευση απάτης.** Η διαδικασία ανίχνευσης απάτης μπορεί να αυτοματοποιηθεί και να ενισχυθεί με τη χρήση εργαλείων αναγνώρισης φωτογραφιών τεχνητής νοημοσύνης. Για παράδειγμα, μια μέθοδος ανίχνευσης απάτης είναι η χρήση ενός εργαλείου αναγνώρισης εικόνας AI για την επεξεργασία επιταγών ή άλλων εγγράφων που υποβάλλονται στις τράπεζες. Για να αξιολογήσει τη γνησιότητα και τη νομιμότητα μιας επιταγής, ο υπολογιστής αναλύει σαρωμένες εικόνες για να εξάγει κρίσιμα δεδομένα, όπως ο αριθμός λογαριασμού, ο αριθμός της επιταγής, το ποσό της επιταγής και η υπογραφή του κατόχου του λογαριασμού.

- **Ταυτοποίηση προσώπων.** Οι κυβερνητικές υπηρεσίες, οι υπηρεσίες επιβολής του νόμου και οι υπηρεσίες ασφαλείας χρησιμοποιούν την αναγνώριση εικόνας για την αναγνώριση και τη συλλογή πληροφοριών σχετικά με άτομα σε φωτογραφίες και βίντεο.

Οι τρέχουσες και μελλοντικές εφαρμογές της αναγνώρισης εικόνας περιλαμβάνουν έξυπνες βιβλιοθήκες φωτογραφιών, στοχευμένη διαφήμιση, διαδραστικά μέσα, προσβασιμότητα για άτομα με προβλήματα όρασης και βελτιωμένες ερευνητικές δυνατότητες [12].

1.2.3 Μέθοδοι και τεχνικές αναγνώρισης εικόνας

Η ταξινόμηση μεταξύ αντικειμένων είναι ένα πολύπλοκο έργο και, ως εκ τούτου, η ταξινόμηση εικόνων αποτελεί σημαντικό έργο στον τομέα της όρασης υπολογιστών (CV). Η ταξινόμηση εικόνων αναφέρεται στην επισήμανση εικόνων σε μία από ορισμένες προκαθορισμένες κατηγορίες (Classes). Υπάρχει δυνητικά n αριθμός κλάσεων στις οποίες μπορεί να ταξινομηθεί μια δεδομένη εικόνα. Ο χειροκίνητος έλεγχος και η ταξινόμηση των εικόνων θα μπορούσε να είναι μια επίπονη και χρονοβόρα εργασία, ιδίως όταν ο αριθμός τους είναι τεράστιος, και ως εκ τούτου θα ήταν πολύ χρήσιμο αν μπορούσαμε να αυτοματοποιήσουμε ολόκληρη αυτή τη διαδικασία με τη χρήση της όρασης υπολογιστών. Οι εξελίξεις στον τομέα της αυτόνομης οδήγησης αποτελούν ένα καλό παράδειγμα της χρήσης της ταξινόμησης εικόνων στον πραγματικό κόσμο. Οι εφαρμογές περιλαμβάνουν την αυτοματοποιημένη οργάνωση εικόνων, ιστοσελίδες αποθεμάτων (stock) φωτογραφιών και βίντεο, οπτική αναζήτηση για βελτιωμένη δυνατότητα ανακάλυψης προϊόντων, μεγάλες οπτικές βάσεις δεδομένων, αναγνώριση εικόνων και προσώπων σε κοινωνικά δίκτυα και πολλά άλλα- γι' αυτό, χρειαζόμαστε ταξινομητές για να επιτύχουμε τη μέγιστη δυνατή ακρίβεια [12].

Δομή για την ταξινόμηση εικόνων

- **Προεπεξεργασία εικόνας (Image Preprocessing)**

Στόχος αυτής της διαδικασίας είναι η βελτίωση των δεδομένων εικόνας (χαρακτηριστικά) με την καταστολή ανεπιθύμητων παραμορφώσεων και την ενίσχυση ορισμένων σημαντικών χαρακτηριστικών της εικόνας, έτσι ώστε τα μοντέλα υπολογιστικής όρασης να μπορούν να επωφεληθούν από αυτά τα βελτιωμένα δεδομένα για να εργαστούν. Τα βήματα για την προεπεξεργασία εικόνας περιλαμβάνουν την ανάγνωση της εικόνας, την αλλαγή μεγέθους της εικόνας και την ενίσχυση των δεδομένων (κλιμάκωση του γκρι της εικόνας, αντανάκλαση, θόλωση Gauss, ιστόγραμμα, εξισορρόπηση, περιστροφή και μετάφραση).

- **Ανίχνευση ενός αντικειμένου (Object Detection and Localization)**

Η ανίχνευση αναφέρεται στον εντοπισμό του αντικειμένου που σημαίνει την κατάτμηση της εικόνας και τον προσδιορισμό της θέσης του αντικειμένου ενδιαφέροντος.

- **Εξαγωγή και εκπαίδευση χαρακτηριστικών (Feature Extraction and Training)**

Πρόκειται για ένα κρίσιμο βήμα κατά το οποίο χρησιμοποιούνται στατιστικές τεχνικές ή μέθοδοι βαθιάς μάθησης για τον εντοπισμό των πιο σημαντικών μοτίβων της εικόνας, χαρακτηριστικών που μπορεί να είναι μοναδικά για μια συγκεκριμένη κλάση και τα οποία θα βοηθήσουν, αργότερα, το

μοντέλο να διακρίνει μεταξύ διαφορετικών κλάσεων. Αυτή η διαδικασία κατά την οποία το μοντέλο μαθαίνει τα χαρακτηριστικά από το σύνολο δεδομένων ονομάζεται εκπαίδευση μοντέλου.

- **Ταξινόμηση αντικειμένου (Object Classification)**

Το βήμα αυτό κατηγοριοποιεί τα ανιχνευμένα αντικείμενα στις προκαθορισμένες κλάσεις με τη χρήση μιας κατάλληλης τεχνικής ταξινόμησης που συγκρίνει τα πρότυπα της εικόνας με τα πρότυπα-στόχους.

- **Επιβλεπόμενη ταξινόμηση (Supervised Classification)**

Η ταξινόμηση με επίβλεψη βασίζεται στην ιδέα ότι ο χρήστης μπορεί να επιλέξει δείγματα εικονοστοιχείων (sample pixels) σε μια εικόνα που είναι αντιπροσωπευτικά συγκεκριμένων κατηγοριών και στη συνέχεια να καθοδηγήσει το λογισμικό επεξεργασίας εικόνας να χρησιμοποιήσει αυτές τις περιοχές εκπαίδευσης ως αναφορά για την ταξινόμηση όλων των άλλων εικονοστοιχείων στην εικόνα. Οι περιοχές εκπαίδευσης (γνωστές ως σύνολα δοκιμών ή κλάσεις εισόδου) επιλέγονται με βάση τις γνώσεις του χρήστη. Ο χρήστης θέτει κάποια τα όρια για το πόσο παρόμοια πρέπει να είναι άλλα εικονοστοιχεία για να τα ομαδοποιήσει. Τα όρια αυτά συχνά τίθενται με βάση τα φασματικά χαρακτηριστικά της περιοχής εκπαίδευσης. Ο χρήστης ορίζει τον αριθμό των κλάσεων στις οποίες ταξινομείται η εικόνα. Μόλις επιτευχθεί ο στατιστικός χαρακτηρισμός για κάθε κλάση πληροφορίας, η εικόνα ταξινομείται στη συνέχεια εξετάζοντας την ανάκλαση για κάθε εικονοστοιχείο και λαμβάνοντας μια απόφαση σχετικά με το ποια από τις υπογραφές μοιάζει περισσότερο. Η ταξινόμηση με επίβλεψη χρησιμοποιεί αλγόριθμους ταξινόμησης και τεχνικές παλινδρόμησης για την ανάπτυξη προγνωστικών μοντέλων. Οι αλγόριθμοι περιλαμβάνουν τη γραμμική παλινδρόμηση (linear regression), τη λογιστική παλινδρόμηση (logistic regression), τα νευρωνικά δίκτυα (Neural Networks), το δέντρο απόφασης (Decision Trees), τα Support Vector Machines (SVMs), τον Naive Bayes και τον k -κοντινότερο γείτονα (k -nearest Neighbor - k NN) [13].

- **Ταξινόμηση χωρίς επίβλεψη (Unsupervised Classification)**

Η ταξινόμηση χωρίς επίβλεψη είναι η περίπτωση όπου τα αποτελέσματα (ομαδοποιήσεις εικονοστοιχείων με κοινά χαρακτηριστικά) βασίζονται στην ανάλυση μιας εικόνας από το λογισμικό, χωρίς ο χρήστης να παρέχει δείγματα κλάσεων. Ο υπολογιστής χρησιμοποιεί τεχνικές για να καθορίσει ποια εικονοστοιχεία σχετίζονται μεταξύ τους και τα ομαδοποιεί σε κλάσεις. Ο χρήστης μπορεί να καθορίσει τον αλγόριθμο που θα χρησιμοποιήσει το λογισμικό και τον επιθυμητό αριθμό κλάσεων εξόδου, αλλά κατά τα άλλα δεν βοηθά στη διαδικασία ταξινόμησης. Ωστόσο, ο χρήστης πρέπει να έχει γνώση της περιοχής που ταξινομείται όταν οι ομαδοποιήσεις εικονοστοιχείων με κοινά χαρακτηριστικά που παράγει ο υπολογιστής πρέπει να σχετίζονται με πραγματικά χαρακτηριστικά στο έδαφος. Ορισμένοι από τους πιο συνηθισμένους αλγόριθμους που χρησιμοποιούνται στη μάθηση χωρίς επίβλεψη περιλαμβάνουν την ανάλυση συστάδων (cluster analysis), την ανίχνευση ανωμαλιών (anomaly detection), τα νευρωνικά δίκτυα και προσεγγίσεις για την εκμάθηση μοντέλων λανθάνουσας (latent) μεταβλητής [14].

Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Network, CNN ή ConvNet) είναι ένα ειδικό είδος νευρωνικών δικτύων πολλαπλών στρωμάτων, σχεδιασμένα να αναγνωρίζουν οπτικά μοτίβα απευθείας από εικόνες εικονοστοιχείων με ελάχιστη προεπεξεργασία. Πρόκειται για μια ειδική αρχιτεκτονική των τεχνητών νευρωνικών δικτύων. Τα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούν ορισμένα από τα χαρακτηριστικά του οπτικού φλοιού και ως εκ τούτου έχουν επιτύχει κορυφαία αποτελέσματα σε εργασίες όρασης υπολογιστών. Τα συνελικτικά νευρωνικά δίκτυα αποτελούνται από δύο πολύ απλά στοιχεία, τα συνελικτικά στρώματα (convolutional layers) και τα στρώματα συγκέντρωσης (pooling layers). Διακρίνονται σχεδόν άπειροι τρόποι για την οργάνωση αυτών των επιπέδων για ένα δεδομένο πρόβλημα όρασης υπολογιστή. Τα στοιχεία του συνελικτικού νευρωνικού δικτύου, όπως τα συνελικτικά στρώματα και τα στρώματα συγκέντρωσης, είναι σχετικά εύκολα κατανοητά. Το δύσκολο μέρος της χρήσης των νευρωνικών δικτύων συνελικτικού τύπου στην πράξη είναι ο τρόπος σχεδιασμού αρχιτεκτονικών μοντέλων που χρησιμοποιούν με τον καλύτερο δυνατό τρόπο αυτά τα απλά στοιχεία. Ο λόγος για τον οποίο το συνελικτικό νευρωνικό δίκτυο είναι εξαιρετικά δημοφιλές είναι λόγω της αρχιτεκτονικής του, κατά την οποία δεν υπάρχει ανάγκη εξαγωγής χαρακτηριστικών. Το σύστημα μαθαίνει να κάνει εξαγωγή χαρακτηριστικών με κεντρική ιδέα την χρησιμοποίηση συνέλιξης της εικόνας και φίλτρων για την δημιουργία αμετάβλητων χαρακτηριστικών, τα οποία μεταβιβάζονται στο επόμενο επίπεδο. Τα χαρακτηριστικά στο επόμενο στρώμα συνελίσσονται με διαφορετικά φίλτρα για τη δημιουργία πιο αμετάβλητων και αφηρημένων χαρακτηριστικών και η διαδικασία συνεχίζεται έως ότου προκύψει το τελικό χαρακτηριστικό/έξοδος (output) το οποίο είναι αμετάβλητο σε αποκρύψεις (το αντικείμενο που εντοπίζουμε βρίσκεται πίσω από κάτι άλλο). Οι πιο συχνά χρησιμοποιούμενες αρχιτεκτονικές των νευρωνικών δικτύων συνελικτικής ανάλυσης είναι οι LeNet, AlexNet, ZFNet, GoogLeNet, VGGNet και ResNet [15].

Τεχνητό νευρωνικό δίκτυο (Artificial Neural Network – ANN)

Εμπνευσμένα από τις ιδιότητες των βιολογικών νευρωνικών δικτύων, τα ANN είναι αλγόριθμοι στατιστικής μάθησης και χρησιμοποιούνται για ποικίλες εργασίες, από σχετικά απλές εργασίες ταξινόμησης μέχρι την όραση υπολογιστών και την αναγνώριση ομιλίας. Τα τεχνητά νευρωνικά δίκτυα υλοποιούνται ως ένα σύστημα διασυνδεδεμένων στοιχείων επεξεργασίας, που ονομάζονται κόμβοι (nodes), τα οποία είναι λειτουργικά ανάλογα με τους βιολογικούς νευρώνες. Οι συνδέσεις μεταξύ διαφορετικών κόμβων έχουν αριθμητικές τιμές, που ονομάζονται βάρη, και μεταβάλλοντας τις τιμές αυτές με συστηματικό τρόπο, το δίκτυο είναι τελικά σε θέση να προσεγγίσει την επιθυμητή λειτουργία. Τα κρυφά στρώματα (hidden layers) μπορούν να θεωρηθούν ως μεμονωμένοι ανιχνευτές χαρακτηριστικών, οι οποίοι αναγνωρίζουν όλο και πιο σύνθετα μοτίβα στα δεδομένα καθώς αυτά διαδίδονται σε όλο το δίκτυο. Αυτή η ιεραρχία επιτρέπει στο δίκτυο να αναγνωρίζει τελικά πολύ σύνθετα αντικείμενα. Οι διάφοροι τύποι τεχνητών νευρωνικών δικτύων είναι το συνελικτικό νευρωνικό δίκτυο (CNN), το νευρωνικό δίκτυο τροφοδότησης (feedforward), το πιθανοτικό (probabilistic) νευρωνικό δίκτυο, το νευρωνικό δίκτυο χρονικής καθυστέρησης (time delay), το βαθύ δίκτυο

στοίβαξης (stacking), το δίκτυο συναρτήσεων ακτινικής βάσης (radial basis function network) και το επαναλαμβανόμενο (recurrent) νευρωνικό δίκτυο [16].

Support Vector Machines

Τα SVMs είναι ισχυροί αλλά ευέλικτοι αλγόριθμοι μηχανικής μάθησης με επίβλεψη, οι οποίοι χρησιμοποιούνται τόσο για ταξινόμηση (classification) όσο και για παλινδρόμηση (regression). Οι μηχανές διανυσμάτων υποστήριξης έχουν μοναδικό τρόπο υλοποίησης σε σύγκριση με τους αλγορίθμους μηχανικής μάθησης. Είναι εξαιρετικά δημοφιλείς λόγω της ικανότητάς τους να χειρίζονται πολλαπλές συνεχείς και κατηγορικές μεταβλητές. Το μοντέλο μηχανής διανυσμάτων υποστήριξης είναι ουσιαστικά μια αναπαράσταση των διαφόρων κλάσεων σε ένα υπερεπίπεδο στον πολυδιάστατο χώρο. Το υπερεπίπεδο θα δημιουργηθεί με επαναληπτικό τρόπο από τη μηχανή διανυσμάτων υποστήριξης, έτσι ώστε να ελαχιστοποιηθεί το σφάλμα. Ο στόχος είναι να χωριστούν τα σύνολα δεδομένων σε κλάσεις για να βρεθεί ένα μέγιστο οριακό υπερεπίπεδο. Δημιουργεί ένα υπερεπίπεδο (hyper-plane) ή ένα σύνολο υπερεπιπέδων σε έναν χώρο υψηλών διαστάσεων και ο καλός διαχωρισμός μεταξύ των δύο κλάσεων επιτυγχάνεται από το υπερεπίπεδο που έχει τη μεγαλύτερη απόσταση από το πλησιέστερο σημείο δεδομένων εκπαίδευσης κάθε κλάσης. Η πραγματική ισχύς αυτού του αλγορίθμου εξαρτάται από τη συνάρτηση πυρήνα που χρησιμοποιείται. Οι πιο συχνά χρησιμοποιούμενοι πυρήνες είναι ο γραμμικός πυρήνας, ο γκαουσιανός πυρήνας και ο πολυωνυμικός πυρήνας [17].

KNN

Η μέθοδος KNN είναι μια μη παραμετρική μέθοδος που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση. Και στις δύο περιπτώσεις, η είσοδος αποτελείται από τα k πλησιέστερα παραδείγματα εκπαίδευσης στο χώρο χαρακτηριστικών (feature space). Είναι μακράν ο απλούστερος αλγόριθμος. Είναι ένας μη παραμετρικός, τεμπέλικος (lazy) αλγόριθμος μάθησης, όπου η συνάρτηση προσεγγίζεται μόνο τοπικά και όλοι οι υπολογισμοί αναβάλλονται μέχρι την αξιολόγηση της συνάρτησης. Αυτός ο αλγόριθμος βασίζεται απλώς στην απόσταση μεταξύ των διανυσμάτων χαρακτηριστικών και ταξινομεί άγνωστα σημεία δεδομένων βρίσκοντας την πιο κοινή, συνηθισμένη (common) κλάση μεταξύ των k κοντινότερων παραδειγμάτων. Για να εφαρμοστεί η ταξινόμηση k -κοντινότερων γειτόνων, πρέπει να οριστεί μια μετρική απόστασης, όπου οι συνήθεις επιλογές περιλαμβάνουν την ευκλείδεια απόσταση και την απόσταση Manhattan. Η έξοδος είναι η ένταξη σε μια κλάση. Ένα αντικείμενο ταξινομείται από μια ψηφοφορία πλήθους των γειτόνων του, με το αντικείμενο να κατατάσσεται στην κλάση που είναι πιο κοινή μεταξύ των k πλησιέστερων γειτόνων του (το k είναι ένας θετικός ακέραιος αριθμός, συνήθως μικρός) [18].

Αλγόριθμος Naive Bayes

Οι ταξινομητές Naive Bayes είναι μια συλλογή αλγορίθμων ταξινόμησης που βασίζονται στο θεώρημα του Bayes.. Ο Naive Bayes είναι μια απλή τεχνική για την κατασκευή ταξινομητών που είναι μοντέλα που αποδίδουν ετικέτες κλάσης σε περιπτώσεις προβλήματος, οι οποίες αναπαρίστανται ως διανύσματα τιμών χαρακτηριστικών, όπου οι ετικέτες κλάσης προέρχονται από κάποιο πεπερασμένο σύνολο. Όλοι

οι ταξινομητές Naive Bayes υποθέτουν ότι η τιμή του συγκεκριμένου χαρακτηριστικού είναι ανεξάρτητη από την τιμή οποιουδήποτε άλλου χαρακτηριστικού, δεδομένης της μεταβλητής κλάσης. Ο αλγόριθμος Naive Bayes είναι ένας γρήγορος, εξαιρετικά κλιμακούμενος αλγόριθμος, ο οποίος μπορεί να χρησιμοποιηθεί για ταξινόμηση δυαδικών και πολλαπλών κλάσεων. Εξαρτάται από την πραγματοποίηση μιας σειράς μετρήσεων. Είναι μια δημοφιλής επιλογή για την ταξινόμηση κειμένου, την ταξινόμηση μηνυμάτων spam κ.λπ. Μπορεί εύκολα να εκπαιδευτεί σε μικρό σύνολο δεδομένων. Έχει περιορισμούς, καθώς θεωρεί όλα τα χαρακτηριστικά ως άσχετα μεταξύ τους, οπότε δεν μπορεί να μάθει τη σχέση μεταξύ των χαρακτηριστικών. Το Naive Bayes μπορεί να μάθει τη σημασία μεμονωμένων χαρακτηριστικών, αλλά δεν μπορεί να καθορίσει τη σχέση μεταξύ των χαρακτηριστικών. Διαφορετικοί τύποι αλγορίθμων Naive Bayes είναι ο Gaussian Naive Bayes, ο πολυωνυμικός Naive Bayes και ο Bernoulli Naive Bayes [19].

1.3 Βαθιά Μάθηση (Deep Learning)

1.3.1 Εισαγωγή στην βαθιά μάθηση

Επειδή οι αρχιτεκτονικές βαθιάς μάθησης, ιδίως οι convolutional neural net (convnet), έχουν προσελκύσει πρωτοφανή προσοχή στην επεξεργασία ιατρικών εικόνων, υπάρχει η τάση να παραβλέπεται η πιθανή συμβολή των συμβατικών τεχνικών επεξεργασίας εικόνας. Η γοητεία των νέων αρχιτεκτονικών ConvNet είναι η απλοποίηση του έργου της επεξεργασίας εικόνας. Αλλά αυτή η διευκόλυνση έχει κόστος, κυρίως στη ζήτηση για περισσότερη εκπαίδευση, και εδώ θα αναφερθεί η υπόθεση ότι υπάρχει θέση στην επεξεργασία εικόνας για περισσότερες συμβατικές τεχνικές υπολογιστικής όρασης. Η συγκεκριμένη εργασία επικεντρώνεται στην άνοδο της βαθιάς μάθησης και ειδικότερα στην αρχιτεκτονική ConvNet, τη σχέση μεταξύ της πολυπλοκότητας της εικόνας και της αρχιτεκτονικής επεξεργασίας εικόνας, και εξετάζει τον κανόνα της συγχώνευσης των συμβατικών τεχνικών και της βαθιάς μάθησης. Χρησιμοποιούνται δύο μέτρα σύγκρισης για να οριστεί η ισορροπία μάθησης (μέγεθος συνόλου δεδομένων στο οποίο οι τεχνικές βαθιάς μάθησης αποκτούν υπεροχή) ως συνάρτηση της πολυπλοκότητας εικόνας. Διερευνώνται οι καταστάσεις όπου η συγχώνευση νέων και συμβατικών τεχνικών επεξεργασίας εικόνας προσφέρει την καλύτερη τελική επεξεργασία εικόνας [20].

1.3.2 Η άνοδος της βαθιάς μάθησης

Τα υπολογιστικά μοντέλα βαθιάς μάθησης αποτελούνται από μία σειρά επεξεργαστικών στρωμάτων (processing layers) που λειτουργούν ανεξάρτητα σε αριθμητικά δεδομένα, ώστε να μάθουν ξεχωριστά ιεραρχικές παρουσιάσεις δεδομένων. Η βαθιά μάθηση μπορεί να ανακαλύψει περίπλοκες δομές σε μεγάλα σύνολα δεδομένων χρησιμοποιώντας τον αλγόριθμο backpropagation για να υποδείξει πώς μια μηχανή θα πρέπει να αλλάζει τις εσωτερικές της παραμέτρους. Δεδομένου ότι η αρχιτεκτονική βαθιάς μάθησης περιλαμβάνει στρώματα κόμβων που ενημερώνουν λειτουργικές παραμέτρους με τη σειρά, αποτελεί ένα τύπο νευρωνικού δικτύου. Τα μοντέλα βαθιάς μάθησης διαφέρουν από άλλα νευρωνικά δίκτυα χρησιμοποιώντας ένα βαθύ γράφημα με πολλαπλά στρώματα επεξεργασίας ενός μικρού αριθμού κόμβων, σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα που αποτελούνται από λίγα

επίπεδα με μεγαλύτερο αριθμό κόμβων. Η αυξανόμενη χρήση της βαθιάς μάθησης στην έρευνα μπορεί να αποδοθεί στις προόδους σε διάφορους τομείς, όπως στην ανάπτυξη μεγάλων συνόλων δεδομένων, που αποκαλούνται «big data», στην δραματική αύξηση της υπολογιστικής ισχύος, και στην επιθυμία για επαναπροσδιορισμό των νευρωνικών δικτύων.

Μια αρχιτεκτονική βαθιάς μάθησης που έχει σημειώσει σημαντική επιτυχία σε προκλήσεις αναγνώρισης εικόνων είναι το νευρωνικό δίκτυο συνελίξεων (ConvNet). Η βασική αρχιτεκτονική ConvNet συνδυάζει δύο έννοιες: τον μαθηματικό τελεστή συνέλιξης και ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο. Ένα ή περισσότερα στρώματα συνέλιξης προστίθενται συνήθως σε ένα πλήρως συνδεδεμένο δίκτυο. Η προσθήκη στρωμάτων συνέλιξης σε ένα τυπικό νευρωνικό δίκτυο και η δυνατότητα στον αλγόριθμο εκπαίδευσης με οπισθοδιάδοση (back-propagation) να ενημερώνει όχι μόνο τα βάρη του πλήρως συνδεδεμένου νευρωνικού δικτύου αλλά και τα στοιχεία των πυρήνων συνέλιξης 2D, επιτρέπει στο ConvNet να χρησιμοποιεί απευθείας εικόνες ως εισόδους και απαλλαγθεί από την ανάγκη για τον χειροκίνητο προσδιορισμό του καλύτερου πυρήνα συνέλιξης. Κατά τη διάρκεια της περιόδου εκπαίδευσης, ο καλύτερος πυρήνας συνέλιξης θα βελτιώνεται συνεχώς. Αυτό επέτρεψε στους ερευνητές να επικεντρωθούν στη βελτιστοποίηση της αρχιτεκτονικής του, χωρίς να απαιτείται η συμβατική χειροκίνητη ρύθμιση της εξαγωγής χαρακτηριστικών. Αυτός ο αριθμός των βαρών (παραμέτρων) στο συνελικτικό νευρωνικό δίκτυο είναι μεγάλος και παρουσιάζει υπολογιστικό βάρος στον αλγόριθμο εκπαίδευσης back-propagation. Οι υψηλές υπολογιστικές απαιτήσεις για τη βελτιστοποίηση τόσων πολλών παραμέτρων δεν είναι η μόνη ανησυχία. Εξαιτίας του αριθμού των παραμέτρων, απαιτείται μεγάλος αριθμός δειγμάτων εκπαίδευσης για την επιτυχή εκπαίδευση και γενίκευση. Η υψηλή υπολογιστική απαίτηση για να μπορέσει να χρησιμοποιηθεί η βαθιά μάθηση έχει ελαφρυνθεί κάπως από τη δραματική αύξηση της υπολογιστικής ισχύος από τη συνηθισμένη πλέον χρήση πολλαπλών πυρήνων (CPU) στους σημερινούς επεξεργαστές και στις εξειδικευμένες μονάδες επεξεργασίας γραφικών (GPU).

Η GPU προέκυψε λόγω των απαιτήσεων των παικτών ηλεκτρονικών παιχνιδιών για πιο λεπτομερή γραφικά. Οι προγραμματιστές αυτών των GPU σχεδίασαν αυτούς τους επεξεργαστές ώστε να περιλαμβάνουν εκατοντάδες, πυρήνες που είναι ειδικά σχεδιασμένοι για να εκτελούν γρήγορα και αποτελεσματικά πράξεις πινάκων κινητής υποδιαστολής σε μια προσπάθεια να απαλλάξουν την CPU από το βάρος. Ένα απροσδόκητο αποτέλεσμα ήταν ότι οι GPU μπορούσαν να αξιοποιηθούν για μηχανική μάθηση, καθώς τα νευρωνικά δίκτυα μπορούσαν επίσης να εκφραστούν με μια ακολουθία πράξεων πινάκων κινητής υποδιαστολής. Οι αλγόριθμοι μηχανικής μάθησης άρχισαν να αναπτύσσονται και να υλοποιούνται με παράλληλο τρόπο για να επωφεληθούν από αυτές οι GPU. Αυτοί οι παράλληλοι αλγόριθμοι θα μπορούσαν τώρα να αξιοποιηθούν σε συστοιχίες CPU ή ιδανικά σε GPU. Ως αποτέλεσμα της παράλληλης υλοποίησης των αλγορίθμων μηχανικής μάθησης μαζί με τις γρήγορες λειτουργίες κινητής υποδιαστολής μέσω της χρήσης των GPU, είναι δυνατή η μείωση των χρόνων εκπαίδευσης, όταν συγκρίνεται μία CPU [21].

ΚΕΦΑΛΑΙΟ 2: Μεθοδολογίες

2.1 Μοντέλα CNN και Μεθοδολογίες

2.1.1 Ανασκόπηση μοντέλων CNN

Τα τελευταία χρόνια στον κλάδο της πληροφορικής, υπάρχει τεράστια ζήτηση για ένα συγκεκριμένο σύνολο δεξιοτήτων, γνωστό ως Deep Learning. Η βαθιά μάθηση είναι ένα υποσύνολο της μηχανικής μάθησης που αποτελείται από αλγορίθμους που είναι εμπνευσμένοι από τη λειτουργία του ανθρώπινου εγκεφάλου ή των νευρωνικών δικτύων. Μαθαίνει στον υπολογιστή να κάνει αυτό που έρχεται φυσικά στους ανθρώπους. Στη βαθιά μάθηση, υπάρχουν διάφοροι τύποι μοντέλων όπως τα Τεχνητά Νευρωνικά Δίκτυα (ANN), οι Αυτοκωδικοποιητές (Autoencoders), τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent-NN) και η Ενισχυτική Μάθηση (Reinforcement Learning). Υπάρχει ακόμα ένα συγκεκριμένο μοντέλο που έχει συμβάλει πολύ στον τομέα της όρασης υπολογιστών και της ανάλυσης εικόνων, το οποίο είναι τα Συνελκτικά Νευρωνικά Δίκτυα (CNN) ή αλλιώς τα ConvNet. Το CNN είναι πολύ χρήσιμο καθώς ελαχιστοποιεί την ανθρώπινη προσπάθεια με τον αυτόματο εντοπισμό των χαρακτηριστικών [22].

Τα CNN είναι μια κατηγορία βαθιών νευρωνικών δικτύων που μπορούν να αναγνωρίζουν και να ταξινομούν συγκεκριμένα χαρακτηριστικά από εικόνες και χρησιμοποιούνται ευρέως για την ανάλυση οπτικών εικόνων. Οι εφαρμογές τους κυμαίνονται από την αναγνώριση εικόνων και βίντεο, την ταξινόμηση εικόνων, την ανάλυση ιατρικών εικόνων, την όραση υπολογιστών και την επεξεργασία φυσικής γλώσσας.

Το CNN έχει υψηλή ακρίβεια και εξαιτίας αυτού του γεγονότος είναι χρήσιμο στην αναγνώριση εικόνων. Η αναγνώριση εικόνων έχει ευρύ φάσμα χρήσεων σε διάφορους κλάδους, όπως η ανάλυση ιατρικών εικόνων, η τηλεφωνία, η ασφάλεια, τα συστήματα συστάσεων κ.λπ. Ο όρος «συνέλιξη» στο CNN υποδηλώνει τη μαθηματική λειτουργία της συνέλιξης, η οποία είναι ένα ειδικό είδος γραμμικής πράξης όπου δύο συναρτήσεις πολλαπλασιάζονται για να παραχθεί μια τρίτη συνάρτηση που εκφράζει πώς το σχήμα (shape) της μιας συνάρτησης τροποποιείται από την άλλη. Με απλά λόγια, δύο εικόνες που μπορούν να αναπαρασταθούν ως πίνακες πολλαπλασιάζονται για να δώσουν μια έξοδο που χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από την εικόνα.

2.1.2 Βασικά στοιχεία της αρχιτεκτονικής μοντέλων CNN

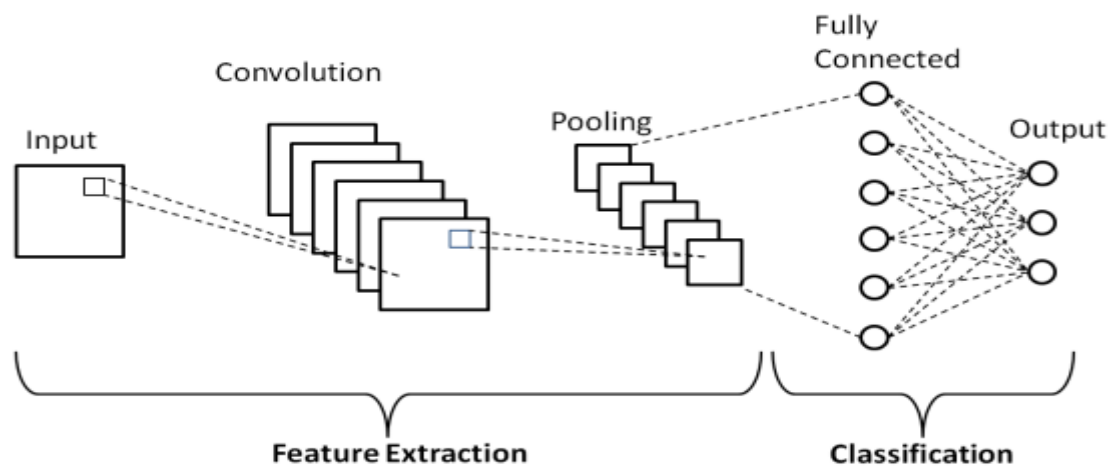
Τα CNN είναι μοντέλα βαθιάς μάθησης που εξάγουν χαρακτηριστικά από εικόνες χρησιμοποιώντας συνεπαγωγικά στρώματα, ακολουθούμενα από συσσωρευτικά και πλήρως συνδεδεμένα στρώματα για εργασίες όπως η ταξινόμηση εικόνων. Διακρίνονται στην καταγραφή χωρικών ιεραρχιών και μοτίβων, γεγονός που τα καθιστά ιδανικά για την ανάλυση οπτικών δεδομένων [23].

Η αρχιτεκτονική των CNN αποτελείται από δύο βασικά μέρη

- Ένα εργαλείο συνέλιξης που διαχωρίζει και προσδιορίζει τα διάφορα χαρακτηριστικά της εικόνας για ανάλυση σε μια διαδικασία που ονομάζεται εξαγωγή χαρακτηριστικών. Το δίκτυο εξαγωγής

χαρακτηριστικών αποτελείται από πολλά ζεύγη στρωμάτων συνέλιξης (convolutional layers) ή στρωμάτων συγκέντρωσης (pooling layers).

- Ένα πλήρως συνδεδεμένο στρώμα (Fully Connected Layer – FC) που χρησιμοποιεί την έξοδο από τη διαδικασία συνέλιξης και προβλέπει την κατηγορία της εικόνας με βάση τα χαρακτηριστικά που εξήχθησαν στα προηγούμενα στάδια. Αυτό το μοντέλο εξαγωγής χαρακτηριστικών CNN στοχεύει στη μείωση του αριθμού των χαρακτηριστικών που υπάρχουν σε ένα σύνολο δεδομένων (dataset). Δημιουργεί νέα χαρακτηριστικά τα οποία συνοψίζουν τα υπάρχοντα χαρακτηριστικά που περιέχονται σε ένα αρχικό σύνολο χαρακτηριστικών. Υπάρχουν πολλά επίπεδα CNN όπως φαίνεται στη βασική αρχιτεκτονική CNN στην Εικόνα 2.1.2.1.



Εικόνα 2.1.2.1: Βασικά Χαρακτηριστικά Αρχιτεκτονικής CNN [25].

Υπάρχουν τρεις τύποι αρχιτεκτονικής CNN, οι οποίοι είναι τα στρώματα συνελκτικής ανάλυσης, τα στρώματα συγκέντρωσης (Pooling Layers) και τα στρώματα πλήρους σύνδεσης. Όταν αυτά τα στρώματα στοιβάζονται, σχηματίζεται μια αρχιτεκτονική CNN. Εκτός από αυτά τα τρία στρώματα, υπάρχουν δύο ακόμη σημαντικές παράμετροι, οι οποίες είναι το στρώμα εγκατάλειψης (Dropout Layer) και η συνάρτηση ενεργοποίησης (Activation Function) [24].

- **Συνελκτικό στρώμα (Convolutional Layer)**

Αυτό το στρώμα είναι το πρώτο στρώμα που χρησιμοποιείται για την εξαγωγή των διαφόρων χαρακτηριστικών από τις εικόνες εισόδου. Σε αυτό το στρώμα εκτελείται η μαθηματική πράξη της συνέλιξης μεταξύ της εικόνας εισόδου και του φίλτρου συγκεκριμένου μεγέθους $M \times M$. Ολισθαίνοντας το φίλτρο πάνω στην εικόνα εισόδου, λαμβάνεται το γινόμενο μεταξύ του φίλτρου και των τμημάτων της εικόνας εισόδου σε σχέση με το μέγεθος του φίλτρου ($M \times M$).

Η έξοδος ονομάζεται χάρτης χαρακτηριστικών (Feature Map), ο οποίος του δίνει πληροφορίες για την εικόνα, όπως οι γωνίες και οι ακμές. Αργότερα, αυτός ο χάρτης χαρακτηριστικών τροφοδοτείται σε άλλα επίπεδα για να μάθει διάφορα άλλα χαρακτηριστικά της εικόνας εισόδου. Το στρώμα συνέλιξης στο CNN μεταβιβάζει το αποτέλεσμα στο επόμενο στρώμα αφού εφαρμόσει

τη λειτουργία συνέλιξης στην είσοδο. Τα στρώματα συνέλιξης στο CNN ωφελούν πολύ, καθώς διασφαλίζουν ότι η χωρική σχέση μεταξύ των εικονοστοιχείων παραμένει ανέπαφη.

- **Στρώμα συγκέντρωσης (Pooling Layers)**

Στις περισσότερες περιπτώσεις, ένα στρώμα συνέλιξης ακολουθείται από ένα στρώμα συγκέντρωσης. Ο πρωταρχικός στόχος αυτού του στρώματος είναι να μειωθεί το μέγεθος του χάρτη συνελκτικών χαρακτηριστικών για να μειωθεί το υπολογιστικό κόστος. Αυτό πραγματοποιείται με τη μείωση των συνδέσεων μεταξύ των στρωμάτων και λειτουργεί ανεξάρτητα σε κάθε χάρτη χαρακτηριστικών. Ανάλογα με τη μέθοδο που χρησιμοποιείται, υπάρχουν διάφοροι τύποι πράξεων συγκέντρωσης. Στο Max Pooling, το μεγαλύτερο στοιχείο λαμβάνεται από τον χάρτη χαρακτηριστικών. Το Average Pooling υπολογίζει το μέσο όρο των στοιχείων σε ένα τμήμα της εικόνας προκαθορισμένου, από το χρήστη, μεγέθους. Στο Sum Pooling υπολογίζεται το συνολικό άθροισμα των στοιχείων στο προκαθορισμένο τμήμα. Το Pooling Layer χρησιμεύει συνήθως ως γέφυρα μεταξύ του Convolutional Layer και του επιπέδου FC. Αυτό το μοντέλο CNN γενικεύει τα χαρακτηριστικά που εξάγονται από το στρώμα συνέλιξης και βοηθά τα δίκτυα να αναγνωρίζουν τα χαρακτηριστικά ανεξάρτητα. Με τη βοήθεια αυτού, οι υπολογισμοί μειώνονται δραματικά σε ένα δίκτυο, γεγονός που χαμηλώνει τις απαιτήσεις ισχύος του συστήματος αλλά και του χρόνου που αφιερώνεται για την εκτέλεση των πράξεων.

- **Πλήρως συνδεδεμένο στρώμα (Fully Connected Layer)**

Το FC αποτελείται από τα βάρη και τις τάσεις (biases) μαζί με τους νευρώνες και χρησιμοποιείται για τη σύνδεση των νευρώνων μεταξύ δύο διαφορετικών στρωμάτων. Αυτά τα στρώματα τοποθετούνται συνήθως πριν από το στρώμα εξόδου και αποτελούν τα τελευταία στρώματα μιας αρχιτεκτονικής CNN. Σε αυτό, η εικόνα εισόδου από τα προηγούμενα στρώματα ισοπεδώνεται και τροφοδοτείται στο στρώμα FC. Το πεπλατυσμένο (flattened) διάνυσμα στη συνέχεια περνάει από μερικά ακόμη στρώματα FC, όπου συνήθως το διαδέχονται πράξεις με μαθηματικές συναρτήσεις. Σε αυτό το στάδιο αρχίζει η διαδικασία ταξινόμησης (classification). Ο λόγος για τον οποίο συνδέονται δύο στρώματα είναι ότι δύο πλήρως συνδεδεμένα στρώματα θα αποδώσουν καλύτερα από ένα μόνο συνδεδεμένο στρώμα.

- **Στρώμα Εγκατάλειψης (Dropout Layer)**

Συνήθως, όταν όλα τα χαρακτηριστικά συνδέονται στο στρώμα FC, μπορεί να προκληθεί υπερπροσαρμογή (overfitting) στο σύνολο δεδομένων εκπαίδευσης (training data). Η υπερπροσαρμογή συμβαίνει όταν ένα συγκεκριμένο μοντέλο λειτουργεί τόσο καλά στα δεδομένα εκπαίδευσης προκαλώντας αρνητικό αντίκτυπο στην απόδοση του μοντέλου όταν χρησιμοποιείται σε νέα δεδομένα. Για να ξεπεραστεί αυτό το πρόβλημα, χρησιμοποιείται ένα στρώμα εγκατάλειψης, όπου μερικοί νευρώνες εγκαταλείπονται από το νευρωνικό δίκτυο κατά τη διαδικασία εκπαίδευσης με αποτέλεσμα τη μείωση του μεγέθους του μοντέλου. Άμα το επίπεδο εγκατάλειψης είναι προκαθορισμένο στο 0.3, το 30% των κόμβων αποσύρεται τυχαία από το

νευρωνικό δίκτυο. Το Dropout έχει ως αποτέλεσμα τη βελτίωση της απόδοσης του μοντέλου μηχανικής μάθησης, καθώς αποτρέπει την υπερβολική προσαρμογή κάνοντας το δίκτυο απλούστερο, αποβάλλοντας νευρώνες από τα νευρωνικά δίκτυα κατά τη διάρκεια της εκπαίδευσης.

- **Συνάρτηση ενεργοποίησης**

Μία από τις σημαντικότερες παραμέτρους του μοντέλου CNN είναι η συνάρτηση ενεργοποίησης. Χρησιμοποιούνται για την εκμάθηση και την προσέγγιση κάθε είδους συνεχούς και σύνθετης (Complex) σχέσης μεταξύ των μεταβλητών του δικτύου. Με απλά λόγια, αποφασίζει ποιες πληροφορίες του μοντέλου θα πρέπει να προχωρήσουν και ποιες όχι στο τέλος του δικτύου. Προσθέτει μη γραμμικότητα (non-linearity) στο δίκτυο. Υπάρχουν διάφορες ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης, όπως οι συναρτήσεις Rectified Linear Unit (ReLU), SoftMax, TanH και Sigmoid. Για ένα μοντέλο δυαδικής ταξινόμησης CNN, προτιμώνται οι συναρτήσεις Sigmoid και SoftMax, ενώ για ταξινόμηση πολλαπλών κλάσεων, γενικά χρησιμοποιείται η SoftMax. Με απλά λόγια, οι συναρτήσεις ενεργοποίησης σε ένα μοντέλο CNN καθορίζουν αν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι. Αποφασίζει αν η είσοδος στο έργο είναι σημαντική ή όχι για την πρόβλεψη χρησιμοποιώντας μαθηματικές πράξεις [26].

Σημασία της συνάρτησης ενεργοποίησης ReLU στο CNN:

Η ReLU είναι μια δημοφιλής συνάρτηση ενεργοποίησης που χρησιμοποιείται στα CNN δίκτυα. Εισάγει τη μη γραμμικότητα με την άμεση έξοδο της εισόδου εάν είναι θετική και μηδενική σε αντίθετη περίπτωση, βοηθώντας τα μοντέλα να μαθαίνουν σύνθετα μοτίβα.

2.1.3 Δεδομένα και διαχωρισμός (Εκπαίδευση, Επικύρωση, Δοκιμή)

Στα Συνελκτικά Νευρωνικά Δίκτυα είναι ζωτικής σημασίας η ακριβής αξιολόγηση της απόδοσής τους, με σκοπό την διασφάλιση της γενίκευσης σε νέα, αθέατα δεδομένα. Για να επιτευχθεί αυτό, ένα σύνολο δεδομένων χωρίζεται σε τρία υποσύνολα, το σύνολο εκπαίδευσης, επικύρωσης και δοκιμής.

- **Σύνολο εκπαίδευσης (Training Set):**

Είναι το τμήμα του συνόλου δεδομένων που προορίζεται για την προσαρμογή του μοντέλου. Με άλλα λόγια, το μοντέλο βλέπει και μαθαίνει από τα δεδομένα του συνόλου εκπαίδευσης για να βελτιώσει άμεσα τις παραμέτρους του. Για να μεγιστοποιηθεί η απόδοση του μοντέλου, το σύνολο εκπαίδευσης πρέπει να είναι αρκετά μεγάλο ώστε να αποδίδει ουσιαστικά αποτελέσματα και να είναι αντιπροσωπευτικό του συνόλου των δεδομένων. Αυτό θα επιτρέψει στο εκπαιδευμένο μοντέλο να προβλέψει νέα δεδομένα που ενδέχεται να εμφανιστούν στο μέλλον.

- **Σύνολο επικύρωσης (Validation Set):**

Αποτελεί το σύνολο των δεδομένων που χρησιμοποιούνται για την αξιολόγηση και τη λεπτομερή ρύθμιση του μοντέλου μηχανικής μάθησης κατά τη διάρκεια της εκπαίδευσης, βοηθώντας στην αξιολόγηση της απόδοσης του μοντέλου και στην πραγματοποίηση προσαρμογών. Η αξιολόγηση του εκπαιδευμένου μοντέλου στο σύνολο επικύρωσης δίνει πληροφορίες σχετικά με την ικανότητά του να

λειτουργεί σε νέα δεδομένα. Το σύνολο επικύρωσης είναι επίσης απαραίτητο για τον συντονισμό των υπερπαραμέτρων (hyperparameter tuning). Οι υπερπαραμέτροι είναι ρυθμίσεις που ελέγχουν τη συμπεριφορά του μοντέλου, όπως ο ρυθμός μάθησης ή η ισχύς της κανονικοποίησης (normalization).

- **Σύνολο δοκιμών (Test Set):**

Είναι το σύνολο δεδομένων που χρησιμοποιείται για την αξιολόγηση της τελικής απόδοσης του εκπαιδευμένου μοντέλου. Χρησιμεύει ως ένα μέτρο αξιολόγησης του μοντέλου κατά την γενίκευση σε νέα δεδομένα σε πραγματικές συνθήκες. Διατηρώντας το σύνολο δοκιμών ξεχωριστά καθ' όλη τη διάρκεια της διαδικασίας, προκύπτει ένα αξιόπιστο μέτρο σύγκρισης της απόδοσης του μοντέλου. Αυτή η αξιολόγηση επιτρέπει να προσδιοριστεί αν το εκπαιδευμένο μοντέλο έχει μάθει επιτυχώς τα σχετικά πρότυπα και μπορεί να κάνει ακριβείς προβλέψεις πέρα από τα πλαίσια εκπαίδευσης και επικύρωσης [27].

2.1.4 Πλεονεκτήματα των CNN από παραδοσιακές μεθόδους αναγνώρισης εικόνας

Τα CNN δίκτυα έχουν φέρει επανάσταση στον τομέα της επεξεργασίας εικόνας και της όρασης υπολογιστών. Μερικά από τα σημαντικότερα πλεονεκτήματα των CNN έναντι των παραδοσιακών μεθόδων επεξεργασίας εικόνας είναι τα εξής:

- **Εκμάθηση χαρακτηριστικών**

Αυτόματη εξαγωγή χαρακτηριστικών: Σε αντίθεση με τις παραδοσιακές μεθόδους που βασίζονται στη χειροκίνητη εξαγωγή χαρακτηριστικών, τα CNN εξάγουν αυτόματα χαρακτηριστικά από τα δεδομένα. Η δυνατότητα εκμάθησης ιεραρχικών αναπαραστάσεων χαρακτηριστικών απευθείας από τις ακατέργαστες εικόνες βελτιώνει την απόδοση και μειώνει την ανάγκη για εξειδικευμένες γνώσεις στην ανάπτυξη χαρακτηριστικών.

Ιεραρχικά χαρακτηριστικά: Το CNN μαθαίνει χαρακτηριστικά σε διάφορα επίπεδα αφαίρεσης. Στο αρχικό επίπεδο, μαθαίνονται χαρακτηριστικά χαμηλού επιπέδου, όπως ακμές και υφές, και σε βαθύτερα επίπεδα, μαθαίνονται χαρακτηριστικά υψηλού επιπέδου, όπως τμήματα αντικειμένων.

- **Επεξεργασία δεδομένων υψηλών διαστάσεων**

Επεκτασιμότητα: Τα CNN μπορούν να επεξεργάζονται αποτελεσματικά δεδομένα υψηλών διαστάσεων. Οι παραδοσιακές μέθοδοι συχνά αποτυγχάνουν να χειριστούν την υψηλή διαστατικότητα των δεδομένων εικόνας, αλλά τα CNN έχουν σχεδιαστεί για να χειρίζονται μεγάλα, υψηλής διάστασης δεδομένα εισόδου.

- **Ιδιότητες Αμεταβλητότητας**

Μη μεταφραστική μεταβλητότητα: Λόγω των λειτουργιών συνέλιξης και συγκέντρωσης, τα CNN έχουν ενσωματωμένη μεταφραστική ικανότητα, πράγμα που σημαίνει ότι μπορούν να αναγνωρίζουν αντικείμενα ανεξάρτητα από τη θέση τους στην εικόνα.

Ανθεκτικότητα στις παραμορφώσεις: Τα CNN είναι ανθεκτικά σε διάφορες παραμορφώσεις και μεταβολές στις εικόνες, όπως περιστροφές, κλιμάκωση και διάφορους συγγενείς μετασχηματισμούς.

- **Βελτιωμένη ακρίβεια και απόδοση**

Υψηλή ακρίβεια: Τα CNN έχουν επιτύχει κορυφαίες επιδόσεις σε διάφορες εργασίες αναγνώρισης εικόνων. Ξεπερνούν τις παραδοσιακές μεθόδους σε ακρίβεια, ιδίως για πολύπλοκες εργασίες ταξινόμησης και ανίχνευσης εικόνων.

Γενίκευση: Τα CNN γενικεύουν καλύτερα σε νέα δεδομένα σε σύγκριση με τις παραδοσιακές τεχνικές, υπό την προϋπόθεση ότι εκπαιδεύονται σε αρκετά μεγάλα και διαφορετικά σύνολα δεδομένων.

- **Ευελιξία και προσαρμοστικότητα**

Προσαρμοστικότητα: Τα CNN μπορούν να προσαρμοστούν σε ένα ευρύ φάσμα εργασιών επεξεργασίας εικόνας, συμπεριλαμβανομένης της ταξινόμησης εικόνων, της ανίχνευσης αντικειμένων, της κατάτμησης και άλλων. Οι παραδοσιακές μέθοδοι απαιτούν συχνά προσαρμογές ειδικά για κάθε εργασία.

Μάθηση μεταφοράς: Τα προ-εκπαιδευμένα CNN μπορούν να ρυθμιστούν λεπτομερώς για συγκεκριμένες εργασίες, επιτρέποντας την αποτελεσματική χρήση προ υπαρχόντων μοντέλων και μειώνοντας την ανάγκη για εκτεταμένη εκπαίδευση σε νέες εργασίες.

- **Μειωμένη ανάγκη για προεπεξεργασία**

Ελάχιστη προεπεξεργασία: Τα CNN απαιτούν λιγότερη προεπεξεργασία των εικόνων σε σύγκριση με τις παραδοσιακές μεθόδους, οι οποίες συχνά απαιτούν σημαντική προεπεξεργασία για την εξαγωγή ουσιαστικών χαρακτηριστικών.

- **Προσέγγιση με γνώμονα τα δεδομένα**

Μάθηση από τα δεδομένα: Τα CNN μαθαίνουν απευθείας από τα δεδομένα, γεγονός που τους επιτρέπει να βελτιώνονται συνεχώς καθώς διατίθενται περισσότερα δεδομένα. Οι παραδοσιακές τεχνικές βασίζονται συχνά σε χειροποίητους κανόνες που δεν βελτιώνονται με πρόσθετα δεδομένα.

- **Παράλληλη επεξεργασία**

Αποδοτικός υπολογισμός: Το σύγχρονο υλικό (GPU) και τα πλαίσια λογισμικού είναι βελτιστοποιημένα για τις δυνατότητες παράλληλης επεξεργασίας των CNN, καθιστώντας τα πιο αποδοτικά από υπολογιστική άποψη για εργασίες επεξεργασίας εικόνας μεγάλης κλίμακας σε σύγκριση με τις παραδοσιακές τεχνικές [28].

2.2 Προ-εκπαιδευμένες αρχιτεκτονικές CNN

2.2.1 Εισαγωγή

Τα προ-εκπαιδευμένα μοντέλα CNN (Convolutional Neural Network) είναι νευρωνικά δίκτυα που έχουν εκπαιδευτεί σε ένα μεγάλο σύνολο δεδομένων, συνήθως για μια γενική εργασία αναγνώρισης εικόνων, και διατίθενται για χρήση από άλλους. Αυτά τα μοντέλα έχουν ήδη μάθει να αναγνωρίζουν ένα ευρύ φάσμα χαρακτηριστικών και μοτίβων σε εικόνες. Η σημασία των προ-εκπαιδευμένων

μοντέλων CNN έγκειται στην ικανότητά τους να χρησιμεύουν ως ισχυρό θεμέλιο για μια ποικιλία εργασιών υπολογιστικής όρασης.

• **Μάθηση Μεταφοράς (Transfer Learning - TL)**

Τα προ-εκπαιδευμένα μοντέλα μπορούν να χρησιμοποιηθούν ως σημείο εκκίνησης για τις εργασίες, ακόμη και αν αυτές οι εργασίες είναι αρκετά διαφορετικές από την αρχική εργασία εκπαίδευσης. Η έννοια αυτή είναι γνωστή ως *Transfer Learning*. Χρησιμοποιώντας προ-εκπαιδευμένα μοντέλα, μπορεί να εξοικονομηθεί σημαντικός χρόνος και υπολογιστικοί πόροι που διαφορετικά θα απαιτούνταν για την εκπαίδευση του νευρωνικού δικτύου από την αρχή. Το προ-εκπαιδευμένο μοντέλο βελτιστοποιείται στο συγκεκριμένο σύνολο δεδομένων, προσαρμοζόμενο στο συγκεκριμένο πρόβλημα, το οποίο συχνά απαιτεί πολύ λιγότερα δεδομένα.

• **Εξαγωγή χαρακτηριστικών (Feature Extraction)**

Τα προ-εκπαιδευμένα μοντέλα χρησιμεύουν ως εξαιρετικοί εξαγωγείς χαρακτηριστικών. Τα μαθημένα χαρακτηριστικά μπορούν να χρησιμοποιηθούν από τα ενδιαμέσα στρώματα του δικτύου ως βάση για άλλα μοντέλα μηχανικής μάθησης ή για διάφορες εργασίες όρασης υπολογιστών (CV).

• **Βελτιωμένη απόδοση (Improved Performance)**

Τα προ-εκπαιδευμένα μοντέλα είναι συνήθως αρχιτεκτονικές τελευταίας τεχνολογίας που έχουν εκπαιδευτεί σε μαζικά και ποικίλα σύνολα δεδομένων. Αυτό σημαίνει ότι έχει ήδη γίνει μάθηση για την αναγνώριση ενός ευρέος φάσματος αντικειμένων (objects), υφών (textures) και μοτίβων (patterns), γεγονός που μπορεί να ενισχύσει σημαντικά την απόδοση της εφαρμογής, ειδικά αν το σύνολο δεδομένων είναι περιορισμένο.

• **Μειωμένη υπερπροσαρμογή (Reduced Overfitting)**

Η εκπαίδευση σε μεγάλα σύνολα δεδομένων έχει ήδη πραγματοποιηθεί για τα προ-εκπαιδευμένα μοντέλα, γεγονός που συχνά βοηθά στην αποφυγή της υπερπροσαρμογής όταν γίνεται λεπτομερής ρύθμιση στη συγκεκριμένη εργασία, καθώς χρησιμοποιούνται ουσιαστικές αρχικοποιήσεις βαρών (weights).

• **Ευρεία εφαρμογή (Applicability)**

Τα προ-εκπαιδευμένα μοντέλα μπορούν να χρησιμοποιηθούν σε ένα ευρύ φάσμα εργασιών υπολογιστικής όρασης, συμπεριλαμβανομένης της ταξινόμησης εικόνων, της ανίχνευσης αντικειμένων, της κατάτμησης εικόνων και άλλων. Προ-εκπαιδευμένα μοντέλα είναι διαθέσιμα για διάφορες αρχιτεκτονικές, συμπεριλαμβανομένων των VGG, ResNet.

• **Κοινοτικές συνεισφορές (Community Contributions)**

Τα προ-εκπαιδευμένα μοντέλα μοιράζονται συχνά στην κοινότητα της μηχανικής μάθησης, διευκολύνοντας την πρόσβαση των ερευνητών, των προγραμματιστών και των επαγγελματιών στις τελευταίες εξελίξεις στην όραση υπολογιστών χωρίς να ξεκινούν από το μηδέν.

2.2.2 Οι Δημοφιλέστερες αρχιτεκτονικές CNN

1. Αρχιτεκτονική μοντέλου VGG-16

Η αρχιτεκτονική VGG16 είναι ένα μοντέλο συνελκτικού νευρωνικού δικτύου που αναπτύχθηκε από την Ομάδα Οπτικής Γεωμετρίας (Visual Geometry Group – VGG) του Πανεπιστημίου της Οξφόρδης, γνωστή για την υπεροχή τους σε εργασίες ταξινόμησης εικόνων. Αποτελεί μέλος της οικογένειας μοντέλων VGG, με τα ακόλουθα αρχιτεκτονικά χαρακτηριστικά:

Στρώμα εισόδου (Input Layer)

Το VGG16 δέχεται εικόνες εισόδου σταθερής ανάλυσης, η οποία συνήθως ορίζεται σε 224x224 pixels. Αυτές οι εικόνες εισόδου αποτελούνται συμβατικά από τρία χρωματικά κανάλια, που αντιστοιχούν στον χρωματικό χώρο Κόκκινο, Πράσινο και Μπλε (RGB).

Συνελκτικά στρώματα (Convolutional Layers)

Το VGG16 περιλαμβάνει 13 επίπεδα συνελίξεων, καθένα από τα οποία ακολουθείται διαδοχικά από μια συνάρτηση ενεργοποίησης ReLU. Αυτά τα επίπεδα συνελίξεων είναι υπεύθυνα για την εξαγωγή ιεραρχικών χαρακτηριστικών σε διάφορες κλίμακες εντός της εικόνας εισόδου.

Στρώματα μέγιστης συγκέντρωσης (Max-Pooling Layers)

Μετά από κάθε ζεύγος διαδοχικών στρωμάτων συνελκτικής επεξεργασίας, ενσωματώνεται ένα στρώμα μέγιστης συγκέντρωσης. Αυτά τα στρώματα εξυπηρετούν τον σκοπό της μείωσης της δειγματοληψίας, μειώνοντας τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών, διατηρώντας παράλληλα σημαντικές πληροφορίες χαρακτηριστικών.

Πλήρως συνδεδεμένα στρώματα (Fully-Connected Layers)

Το VGG16 ενσωματώνει τρία πλήρως συνδεδεμένα στρώματα. Τα δύο αρχικά στρώματα είναι εξοπλισμένα με 4096 νευρώνες το καθένα, ενώ το τελευταίο στρώμα περιέχει 1000 νευρώνες. Το τελικό στρώμα ευθυγραμμίζεται με τον αριθμό των κλάσεων στο σύνολο δεδομένων ImageNet και αναλαμβάνει το έργο της αφαίρεσης και ταξινόμησης χαρακτηριστικών υψηλού επιπέδου.

Στρώμα Softmax (SoftMax Layer)

Το καταληκτικό στρώμα είναι ένα στρώμα Softmax, υπεύθυνο για την ανάθεση πιθανοτήτων κλάσης στις διάφορες κατηγορίες στο πλαίσιο του προβλήματος ταξινόμησης.

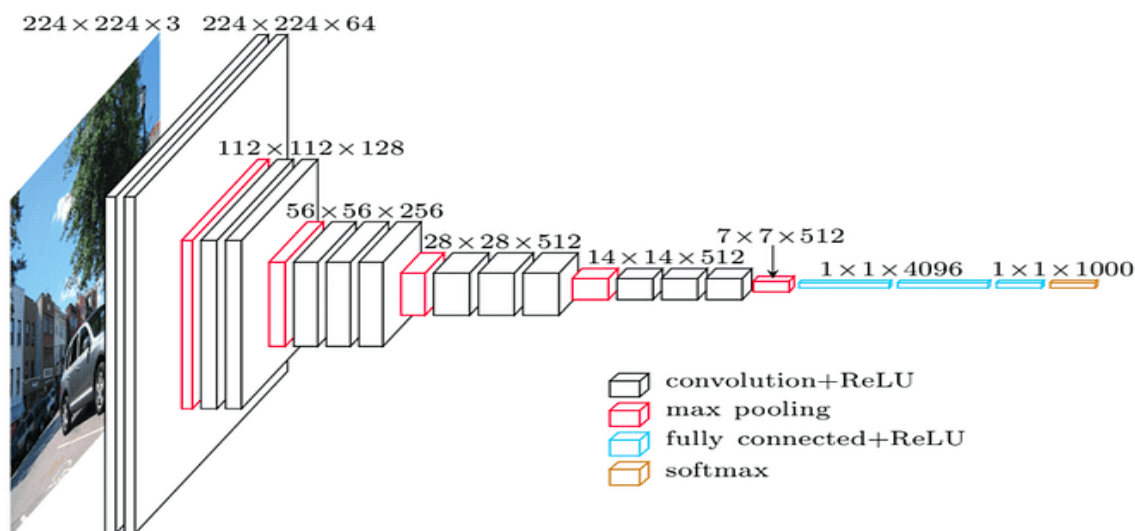
Επίπεδο επιπεδοποίησης (Flatten Layer)

Πριν από τα πλήρως συνδεδεμένα στρώματα, οι χάρτες χαρακτηριστικών συνήθως επιπεδοποιούνται σε ένα μονοδιάστατο διάνυσμα, καθιστώντας τους έτσι κατάλληλους για εισαγωγή στα πλήρως συνδεδεμένα στρώματα.

Πλεονεκτήματα

Το VGG16, αναγνωρισμένο για την απλή αλλά ισχυρή αρχιτεκτονική του, χαρακτηρίζεται από το σημαντικό βάθος και τον σημαντικό αριθμό παραμέτρων. Κατά συνέπεια, χρησιμεύει συχνά ως θεμελιώδες μοντέλο για την εξαγωγή χαρακτηριστικών ή τη λεπτομερή ρύθμιση σε συγκεκριμένες

εργασίες, αντί να ξεκινά την εκπαίδευση από το μηδέν. Η αρχιτεκτονική VGG16 έχει διαδραματίσει καθοριστικό ρόλο στην πρόοδο του πεδίου της βαθιάς μάθησης για την όραση υπολογιστών, προσφέροντας θεμελιώδεις γνώσεις σχετικά με τα αρχιτεκτονικά στοιχεία που διέπουν τα πολύπλοκα νευρωνικά δίκτυα σε εργασίες που σχετίζονται με την αναγνώριση εικόνας.



Εικόνα 2.2.2.1: Οπτικοποίηση Αρχιτεκτονικής VGG16 [29].

2. Αρχιτεκτονική μοντέλου VGG-19

Η αρχιτεκτονική VGG-19, μια επέκταση του αρχικού μοντέλου VGG, αναπτύχθηκε από την Ομάδα Οπτικής Γεωμετρίας (Visual Geometry Group- VGG) του Πανεπιστημίου της Οξφόρδης (University Of Oxford) και αποτελεί σημαντική πρόοδο στον τομέα των βαθιών νευρωνικών δικτύων με συνελκτικό τρόπο (CNN) για εργασίες αναγνώρισης εικόνας.

Στρώμα εισόδου (Input Layer)

Το VGG-19, όπως και ο πρόδρομός του VGG-16, έχει σχεδιαστεί για να επεξεργάζεται εικόνες εισόδου τυποποιημένου μεγέθους, το οποίο συνήθως ορίζεται σε 224×224 εικονοστοιχεία, χρησιμοποιώντας τα συμβατικά τρία χρωματικά κανάλια (RGB). Αυτή η τυποποίηση προωθεί τη συμβατότητα με διαφορετικά σύνολα δεδομένων εικόνων.

Συνελκτικά επίπεδα (Convolutional Layers)

Το VGG-19 παρουσιάζει ένα βαθύτατο βάθος με 16 διαδοχικά στρώματα συνελκτικής επεξεργασίας, καθένα από τα οποία ακολουθείται από μια συνάρτηση ενεργοποίησης ReLU. Αυτά τα στρώματα συνελκτικής επεξεργασίας συμβάλλουν στην καταγραφή περίπλοκων οπτικών χαρακτηριστικών, όπως ακμές, υφές και μοτίβα, σε διάφορες χωρικές κλίμακες.

Επίπεδα Μέγιστης Συγκέντρωσης (Max Pooling Layers)

Μετά από κάθε ζεύγος διαδοχικών στρωμάτων συνελκτικής επεξεργασίας, το VGG-19 ενσωματώνει στρώματα μέγιστης συγκέντρωσης. Αυτά τα στρώματα εξυπηρετούν τον διπλό σκοπό της μείωσης της

χωρικής διάστασης και της διατήρησης των χαρακτηριστικών, διευκολύνοντας την ικανότητα του δικτύου να γενικεύει αποτελεσματικά.

Στρώμα Softmax (Softmax Layer)

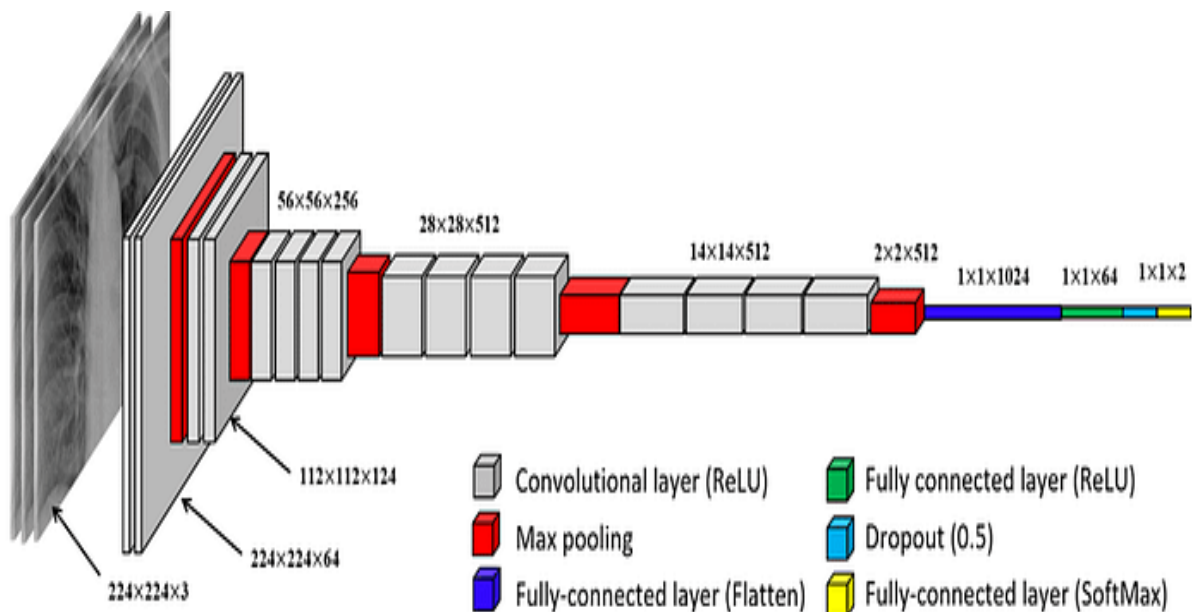
Το τελευταίο αρχιτεκτονικό στοιχείο, το στρώμα Softmax, είναι υπεύθυνο για την ανάθεση πιθανοτήτων κλάσης στην έξοδο του δικτύου. Μετατρέπει τις ακατέργαστες προβλέψεις του μοντέλου σε μια κατανομή πιθανοτήτων σε πιθανές κατηγορίες εικόνων.

Επίπεδο επιπεδοποίησης (Flatten Layer)

Πριν από τη σύνδεση με τα πλήρως συνδεδεμένα στρώματα, χρησιμοποιείται ένα στρώμα επιπεδοποίησης για να μετατρέψει τους χάρτες χαρακτηριστικών από τα προηγούμενα στρώματα συνελκτικής και συγκέντρωσης σε ένα μονοδιάστατο διάνυσμα. Αυτό διευκολύνει την απρόσκοπτη ενσωμάτωση των χαρακτηριστικών για την ταξινόμηση.

Πλεονεκτήματα

Το VGG-19 διακρίνεται για την αρχιτεκτονική πολυπλοκότητά του και τη βαθιά του ικανότητα εκμάθησης χαρακτηριστικών. Το βάθος του επιτρέπει την καταγραφή πλούσιων οπτικών αναπαραστάσεων, αλλά το πλεονέκτημα απαιτεί σημαντικούς υπολογιστικούς πόρους για την εκπαίδευση. Συνεπώς, το VGG-19 χρησιμοποιείται συχνά ως θεμελιώδες μοντέλο για εργασίες που περιλαμβάνουν μάθηση μεταφοράς και λεπτομερή ρύθμιση σε σύνολα δεδομένων.



Εικόνα 2.2.2.2: Οπτικοποίηση Αρχιτεκτονικής VGG19 [29] .

3. Αρχιτεκτονική Μοντέλου ResNet

Η αρχιτεκτονική ResNet, η οποία παρουσιάστηκε ως “Deep Residual Learning for Image Recognition” από μια ομάδα ερευνητών της Microsoft Research, αποτελεί πρωτοποριακή εξέλιξη στο πεδίο των CNN δικτύων. Σχεδιάστηκε ως λύση στις προκλήσεις που αντιμετωπίζονται κατά την εκπαίδευση εξαιρετικά βαθιών νευρωνικών δικτύων, όπου το πρόβλημα των εξαφανιζόμενων κλίσεων αποτελεί σημαντικό εμπόδιο.

Στρώμα εισόδου (Input Layer)

Το δίκτυο ξεκινά με ένα στρώμα εισόδου που δέχεται τυπικές εικόνες RGB, συνήθως διαστάσεων 224x224 pixel.

Αρχικά στρώματα συνέλιξης (Initial Convolutional Layers)

Αρχικά χρησιμοποιείται ένα μόνο επίπεδο συνελκτικής ανάλυσης 7x7 με 64 φίλτρα, ακολουθούμενο από ομαδοποίηση δέσμης και ενεργοποίηση διορθωμένης γραμμικής μονάδας (ReLU).

Στρώματα μέγιστης συγκέντρωσης (Max Pooling Layers)

Μετά την αρχική συνέλιξη, εφαρμόζεται ένα στρώμα μέγιστης συγκέντρωσης για τη μείωση των χωρικών διαστάσεων των χαρτών χαρακτηριστικών

Υπολειμματικά μπλοκ (Residual Blocks)

Τα θεμελιώδη δομικά στοιχεία του ResNet είναι τα υπολειμματικά μπλοκ, τα οποία στοιβάζονται στρατηγικά για την κατασκευή βαθιών δικτύων. Κάθε υπολειμματικό μπλοκ αποτελείται από δύο στρώματα συνέλιξης. Η καθοριστική καινοτομία εδώ είναι η συμπερίληψη μιας σύνδεσης παράκαμψης ή συντόμευσης που παρακάμπτει ένα ή περισσότερα ενδιάμεσα στρώματα. Αυτή η σύνδεση επιτρέπει την άμεση προσθήκη της εισόδου στην έξοδο του υπολειμματικού μπλοκ. Η έννοια αυτή διατυπώνεται μαθηματικά ως εξής: $\text{έξοδος} = \text{είσοδος} + F(\text{είσοδος})$, όπου $F(\text{είσοδος})$ είναι ο μετασχηματισμός που εισάγεται στο υπολειμματικό μπλοκ.

Αρχιτεκτονικές συμφόρησης (Bottleneck Architectures)

Σε σενάρια που απαιτούν βαθύτερα δίκτυα, η αρχιτεκτονική bottleneck υιοθετείται συχνά εντός των υπολειπόμενων μπλοκ. Αυτή η σχεδίαση ενσωματώνει τρία επίπεδα συνέλιξης: 1x1, 3x3 και 1x1 συνέλιξεις. Οι 1x1 συνέλιξεις χρησιμοποιούνται στρατηγικά για τη μείωση της διάστασης του χάρτη χαρακτηριστικών πριν από την εφαρμογή των 3x3 συνέλιξεων και στη συνέχεια την επαναφορά του. Αυτή η αρχιτεκτονική επιλογή εξισορροπεί την υπολογιστική απόδοση και τη χωρητικότητα του δικτύου.

Στοιβάξη μπλοκ (Stacking of Blocks)

Το ResNet υποστηρίζει τη στοιβάξη πολλαπλών υπολειμματικών μπλοκ για την επίτευξη εντυπωσιακού βάθους δικτύου. Ο αριθμός και το βάθος αυτών των μπλοκ μπορεί να ποικίλλει ανάλογα με τη συγκεκριμένη παραλλαγή του ResNet, όπως το ResNet-18 ή το ResNet-50.

Συγκέντρωση του παγκόσμιου μέσου όρου (Global Average Pooling)

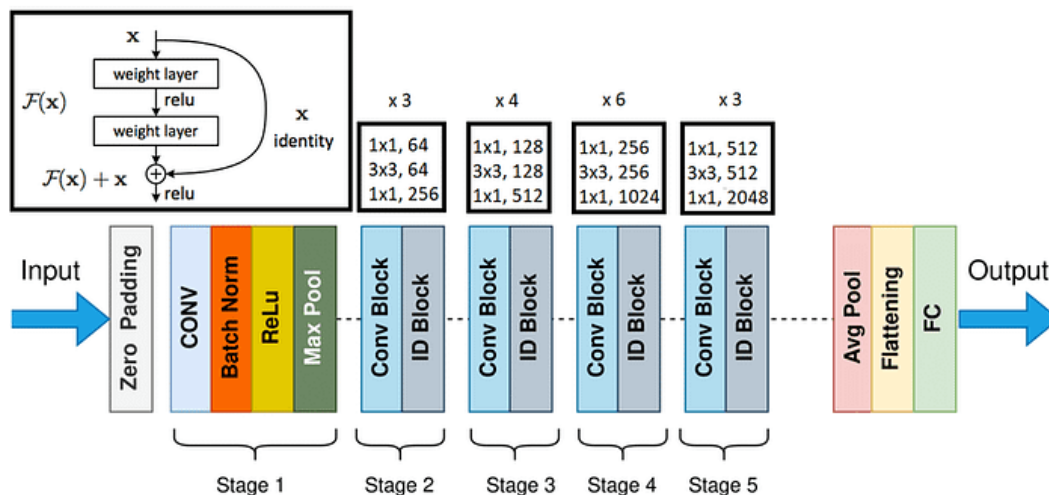
Τα παραδοσιακά πλήρως συνδεδεμένα στρώματα αντικαθίστανται στο ResNet από τη συγκέντρωση του ολικού μέσου όρου (global average pooling). Αυτό το στρώμα συγκέντρωσης υπολογίζει τον μέσο όρο σε αυτές τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών, δίνοντας ένα μονοδιάστατο διάνυσμα που χρησιμεύει άμεσα για την ταξινόμηση.

Στρώμα εξόδου (Output Layer)

Τα χαρακτηριστικά γνωρίσματα με συγκέντρωση του παγκόσμιου μέσου όρου τροφοδοτούνται στη συνέχεια στο στρώμα εξόδου, τυπικά ένα πλήρως συνδεδεμένο στρώμα με ενεργοποίηση Softmax, κατάλληλο για εργασίες ταξινόμησης.

Πλεονεκτήματα

Η αρχιτεκτονική καινοτομία του ResNet, ιδίως η ενσωμάτωση των υπολειμματικών μπλοκ, έχει εγκαινιάσει μια μετασχηματιστική εποχή στη βαθιά μάθηση και την όραση υπολογιστών. Διευκόλυνε την εκπαίδευση εξαιρετικά βαθιών δικτύων, με αποτέλεσμα ουσιαστικές εξελίξεις στην αναγνώριση εικόνων και σε ένα ευρύ φάσμα εφαρμογών όρασης υπολογιστών



Εικόνα 2.2.2.3: Οπτικοποίηση Αρχιτεκτονικής Resnet [29] .

4. Αρχιτεκτονική Μοντέλου AlexNet

Το AlexNet είναι μια πρωτοποριακή αρχιτεκτονική ανάλυση συνελκτικών νευρωνικών δικτύων (CNN) που επινοήθηκε από τον Alex Krizhevsky, Ilya Sutskever και Geoffrey Hinton, η οποία έχει μεταμορφώσει σημαντικά το τοπίο της όρασης υπολογιστών και της ταξινόμησης εικόνων. Τα αρχιτεκτονικά του στοιχεία είναι σχολαστικά δομημένα, καθιστώντας το έναν σημαντικό πρόδρομο της σύγχρονης εποχής των μοντέλων βαθιάς μάθησης.

Στρώμα εισόδου (Input Layer)

Το AlexNet ξεκινά την ανάλυσή του με ένα στρώμα εισόδου προσαρμοσμένο για να δέχεται εικόνες RGB. Οι εικόνες εισόδου αλλάζουν συμβατικά το μέγεθός του σε 227x227 pixel και διατηρούν τα τρία τυπικά κανάλια χρώματος: Κόκκινο, πράσινο και μπλε.

Συνελκτικά Στρώματα (Convolutional Layers)

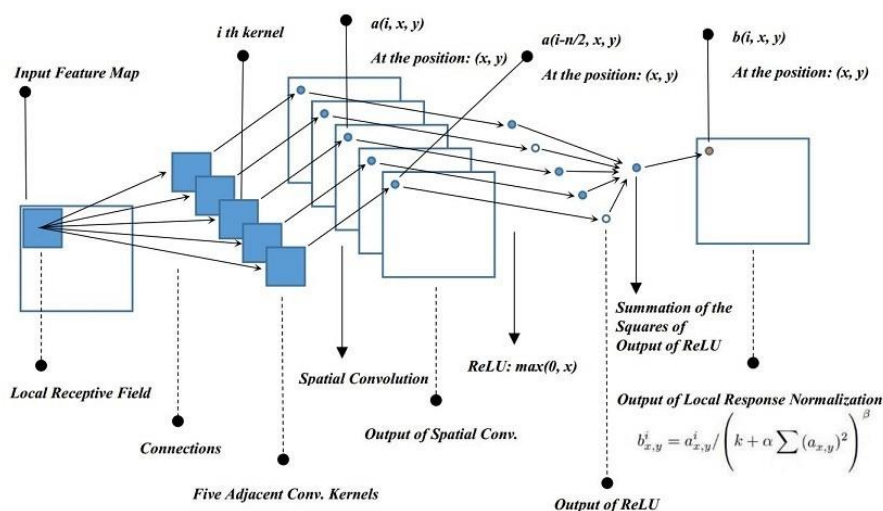
Ο πυρήνας του AlexNet αποτελείται από πέντε συνελκτικά στρώματα. Τα στρώματα αυτά λειτουργούν συντονισμένα για την εξαγωγή του εκτεταμένου φάσματος χαρακτηριστικών εικόνας, τόσο χαμηλού όσο και υψηλού επιπέδου. Για να εισαχθεί μη γραμμικότητα στο μοντέλο, οι συναρτήσεις ενεργοποίησης ReLU ακολουθούν κάθε συνελκτικό στρώμα.

Στρώματα μέγιστης συγκέντρωσης (Max-Pooling Layers)

Μετά τα δύο πρώτα στρώματα συνελκτικής επεξεργασίας, χρησιμοποιούνται στρώματα μέγιστης συγκέντρωσης. Το max pooling εφαρμόζει χωρική μείωση της δειγματοληψίας (spatial down-sampling) των χαρτών χαρακτηριστικών, προωθώντας έτσι τη μεταφραστική αμεταβλητότητα και μειώνοντας τον αριθμό των παραμέτρων.

Κανονικοποίηση τοπικής απόκρισης (Local Response Normalization – LRN)

Το AlexNet ενσωματώνει την κανονικοποίηση της τοπικής απόκρισης μετά το πρώτο και το δεύτερο συνελκτικό στρώμα. Αυτή η τεχνική κανονικοποίησης ενισχύει τις αποκρίσεις των ισχυρά ενεργοποιημένων νευρώνων, αυξάνοντας την ικανότητα γενίκευσης του δικτύου [29].



Εικόνα 2.2.2.4: Οπτικοποίηση LRN [30]

Στρώμα εγκατάλειψης (Dropout)

Η απομάκρυνση εφαρμόζεται με σύνεση στα πλήρως συνδεδεμένα στρώματα του AlexNet κατά τη διάρκεια της εκπαίδευσης. Αυτή η τεχνική, η οποία απενεργοποιεί τυχαία ένα μέρος των νευρώνων, καταπολεμά την υπερπροσαρμογή και αυξάνει την ευρωστία του μοντέλου.

Πλήρως συνδεδεμένα στρώματα (Fully-Connected Layers)

Η αρχιτεκτονική διαθέτει τρία πλήρως συνδεδεμένα στρώματα, καθένα από τα οποία φιλοξενεί 4096 νευρώνες, με αποκορύφωμα ένα στρώμα Softmax. Αυτά τα στρώματα φέρουν την ευθύνη της εξαγωγής χαρακτηριστικών υψηλού επιπέδου και της ταξινόμησης.

Στρώμα Softmax

Η τελευταία βαθμίδα στο AlexNet καταλαμβάνεται από ένα στρώμα Softmax, το οποίο υπολογίζει τις πιθανότητες κλάσης. Λειτουργεί ως ο αγωγός μέσω του οποίου οι ακατέργαστες έξοδοι του μοντέλου μετατρέπονται σε μια κατανομή πιθανοτήτων που καλύπτει τις κατηγορίες ταξινόμησης.

Επίπεδο επιπεδοποίησης (Flatten Layer)

Πριν από τη διασύνδεση με τα πλήρως συνδεδεμένα στρώματα, οι χάρτες χαρακτηριστικών υποβάλλονται σε μετασχηματισμό σε μονοδιάστατο διάνυσμα για απρόσκοπτη ενσωμάτωση.

Ενίσχυση δεδομένων (Data Augmentation)

Κατά τη διάρκεια της εκπαίδευσης, υιοθετούνται στρατηγικές επαύξησης δεδομένων, όπως η τυχαία περικοπή (random cropping) και η οριζόντια αναστροφή (horizontal flipping), για να ενισχυθεί η ανθεκτικότητα του δικτύου και να περιοριστεί η υπερπροσαρμογή.

Πλεονεκτήματα

Οι αρχιτεκτονικές αρχές του AlexNet, οι οποίες χαρακτηρίζονται από βαθιά στρώματα συνελκτικού δικτύου, έχουν προσδώσει εξαιρετική αποτελεσματικότητα στο πεδίο της ταξινόμησης εικόνων. Τα πρωτοποριακά επιτεύγματά του άφησαν ανεξίτηλο αποτύπωμα στην αναβίωση των νευρωνικών δικτύων, αναδιαμορφώνοντας την πορεία της βαθιάς μάθησης και δημιουργώντας την ανάπτυξη μεταγενέστερων αρχιτεκτονικών CNN που τώρα ηγούνται στην επίτευξη κορυφαίων επιδόσεων σε διάφορους υπολογισμούς [29].

2.2.3 Μάθηση Μεταφοράς

Η μάθηση μεταφοράς είναι μια δημοφιλής μέθοδος στην όραση υπολογιστών, επειδή επιτρέπει τη δημιουργία ακριβών μοντέλων με τρόπο που εξοικονομεί χρόνο. Με τη μάθηση μεταφοράς, αντί να ξεκινήσει η διαδικασία μάθησης από το μηδέν, η διαδικασία ξεκινά από πρότυπα που έχουν μάθει κατά την επίλυση ενός διαφορετικού προβλήματος. Αυτή η προσέγγιση αξιοποιεί προηγούμενες μάθησης και αποφεύγει να ξεκινήσει από το μηδέν.

Στην όραση υπολογιστών, η μάθηση μεταφοράς υλοποιείται συνήθως μέσω της χρήσης προ-εκπαιδευμένων μοντέλων. Ένα προ-εκπαιδευμένο μοντέλο είναι ένα μοντέλο που εκπαιδεύτηκε σε ένα μεγάλο σύνολο δεδομένων αναφοράς για την επίλυση του προβλήματος παρόμοιου με αυτό που πρέπει να επιλυθεί. Λόγω του υπολογιστικού κόστους της εκπαίδευσης τέτοιων μοντέλων, είναι κοινή πρακτική η εισαγωγή και χρήση μοντέλων από τη δημοσιευμένη βιβλιογραφία (π.χ. VGG) [31].

2.2.4 Λεπτομερής ρύθμιση (Fine-Tuning) μέσω Transfer Learning

Η ανάλυση ιατρικών εικόνων περιλαμβάνει την εξαγωγή ζωτικών πληροφοριών από ιατρικές εικόνες για διαγνωστικούς και θεραπευτικούς σκοπούς. Ενώ οι εξελίξεις στις τεχνολογίες απεικόνισης διευκολύνουν την έγκαιρη ανίχνευση και τον ακριβή εντοπισμό ασθενειών, η χειροκίνητη ανάλυση παραμένει επίπονη και επιρρεπής σε σφάλματα, οδηγώντας σε ασυνεπείς ερμηνείες. Για να ξεπεραστούν αυτές οι προκλήσεις, οι αυτοματοποιημένες τεχνικές, ιδίως εκείνες που χρησιμοποιούν προηγμένους αλγορίθμους μηχανικής μάθησης, έχουν κερδίσει δημοτικότητα για τη βελτίωση της αποτελεσματικότητας και της αξιοπιστίας στην ανάλυση ιατρικών εικόνων. Μια σημαντική πρόκληση σε αυτόν τον τομέα είναι η έλλειψη μεγάλων, σχολιασμένων συνόλων δεδομένων, τα οποία είναι γενικά απαραίτητα για την εκπαίδευση εύρωστων μοντέλων υψηλής απόδοσης. Η μάθηση μεταφοράς αντιμετωπίζει αυτή την πρόκληση με τη χρήση μοντέλων που εκπαιδεύονται σε εκτεταμένα, διαφορετικά σύνολα δεδομένων και την εφαρμογή τους σε συγκεκριμένα ιατρικά πλαίσια.

Στην ανάλυση ιατρικών εικόνων, η μάθηση μεταφοράς έχει αποδείξει την αξία της παρέχοντας ποιοτική υποστήριξη αποφάσεων και μειώνοντας την εξάρτηση από εκτεταμένα μαρκαρισμένα σύνολα ιατρικών δεδομένων, η συλλογή των οποίων είναι συχνά δαπανηρή και χρονοβόρα. Με τη χρήση προ-εκπαιδευμένων μοντέλων, η TL αξιοποιεί τα γενικευμένα χαρακτηριστικά που έχουν διδαχθεί προηγουμένως από διάφορους τύπους εικόνων, επιτρέποντας την αποτελεσματικότερη εξαγωγή χαρακτηριστικών από ιατρικές εικόνες. Τα δημοφιλή μοντέλα TL σε αυτόν τον τομέα περιλαμβάνουν

τα AlexNet, ResNet και VGGNet, τα οποία έχουν αποδειχθεί αποτελεσματικά σε εργασίες όπως η τμηματοποίηση, η αναγνώριση αντικειμένων, η ανάλυση χειρουργικής ροής εργασίας και η κατηγοριοποίηση ασθενειών.

Η λεπτομερής ρύθμιση μπορεί να βελτιώσει την απόδοση των μοντέλων και να μειώσει την ανάγκη για εκτεταμένη εκπαίδευση και σχολιασμό δεδομένων, και έχει χρησιμοποιηθεί ευρέως στην ταξινόμηση ιατρικών εικόνων, όπου τα δεδομένα είναι συχνά σπάνια, μεταβλητά και πολύπλοκα. Ωστόσο, η λεπτομερής ρύθμιση δημιουργεί και ορισμένες προκλήσεις στην ταξινόμηση ιατρικών εικόνων. Μία από αυτές είναι η επιλογή μιας κατάλληλης αρχιτεκτονικής δικτύου που εξισορροπεί την πολυπλοκότητα και την ακρίβεια. Η αρχιτεκτονική του δικτύου, ιδίως τα επίπεδα συνελίξεων και συγκέντρωσης, γνωστά και ως ραχοκοκαλιά, επηρεάζει τον αριθμό των παραμέτρων, το υπολογιστικό κόστος και την ικανότητα αναπαράστασης δεδομένων του δικτύου. Μεγαλύτερες αρχιτεκτονικές, όπως το ResNet, το GoogLeNet, μπορούν να εξάγουν πολύπλοκα χαρακτηριστικά, αλλά απαιτούν περισσότερα δεδομένα και υπολογιστικούς πόρους για τη λεπτομερή ρύθμιση. Η βέλτιστη αρχιτεκτονική εξαρτάται από τη συγκεκριμένη εφαρμογή, τα διαθέσιμα δεδομένα και τους υπολογιστικούς περιορισμούς και μπορεί να διαφέρει μεταξύ διαφορετικών μελετών. Μια άλλη πρόκληση είναι η επιλογή του σχετικού συνόλου δεδομένων προέλευσης (Source Dataset) για τη μάθηση μεταφοράς.

Το αρχικό σύνολο δεδομένων θα πρέπει να έχει υψηλή ποιότητα και ομοιότητα με το σύνολο δεδομένων-στόχο, έτσι ώστε τα χαρακτηριστικά που μαθαίνονται να είναι μεταβιβάσιμα. Το ImageNet, ένα μεγάλο σύνολο δεδομένων με ποικίλες κατηγορίες αντικειμένων, χρησιμοποιείται συνήθως ως αρχικό σύνολο δεδομένων, αλλά μπορεί να μην ανταποκρίνεται πάντα στις απαιτήσεις του συγκεκριμένου τομέα. Επιπλέον, οι διαφορετικές μέθοδοι ιατρικής απεικόνισης, όπως οι ακτίνες X, οι σαρώσεις PET, οι υπέρηχοι, οι μαγνητικές τομογραφίες και οι μικροσκοπήσεις, αποτελούν μια πρόσθετη πρόκληση. Κάθε τρόπος έχει μοναδικά χαρακτηριστικά που επηρεάζουν την επεξεργασία και την ταξινόμηση της εικόνας. Η ανάπτυξη μοντέλων που μπορούν να χειριστούν αποτελεσματικά αυτή την ποικιλομορφία παραμένει ένα δύσκολο έργο [32].

2.3 Δεδομένα και επεξεργασία

2.3.1 Προεπεξεργασία και Ενίσχυση (Augmentation) Δεδομένων

Είναι αναγνωρισμένο γεγονός ότι η καλή ποιότητα (good quality) και η ποικιλία (variety) των δειγμάτων δεδομένων εκπαίδευσης είναι απαραίτητα για την επίτευξη υψηλών ποσοστών επιτυχίας στη λειτουργία του νευρωνικού δικτύου συνελκτικής ανάλυσης. Η απόδοση του εκπαιδευμένου δικτύου εξαρτάται σε μεγάλο βαθμό από το αν τα δεδομένα εκπαίδευσης είναι όσο το δυνατόν πιο ποικίλα και αντιπροσωπευτικά και αν κάθε μαθήσιμο χαρακτηριστικό (learnable feature) είναι παρόν σε μεγάλο αριθμό δειγμάτων δεδομένων εκπαίδευσης. Υπάρχουν περιπτώσεις κατά τις οποίες δεν είναι δυνατόν να συγκεντρωθεί επαρκής αριθμός αρχικών δειγμάτων και απαιτείται η δημιουργία τεχνητών δειγμάτων [33].

2.3.2 Τεχνικές Προεπεξεργασίας (Preprocessing Techniques)

Προεπεξεργασία δεδομένων

Η προεπεξεργασία δεδομένων είναι η διαδικασία αξιολόγησης (evaluation), φιλτραρίσματος (filtering), χειρισμού (manipulation) και κωδικοποίησης των δεδομένων, έτσι ώστε ένας αλγόριθμος μηχανικής μάθησης να μπορεί να τα κατανοήσει και να χρησιμοποιήσει την έξοδο που προκύπτει. Ο κύριος στόχος της προεπεξεργασίας δεδομένων είναι να εξαλειφθούν προβλήματα δεδομένων, όπως οι ελλείπουσες τιμές, να βελτιωθεί η ποιότητα των δεδομένων και να καταστούν τα δεδομένα χρήσιμα για σκοπούς μηχανικής μάθησης.

Σημασία προεπεξεργασίας δεδομένων

Οι αλγόριθμοι με βάση τα δεδομένα (data-driven) είναι στατιστικές εξισώσεις που λειτουργούν με τιμές βάσεων δεδομένων. Δεδομένου ότι μια ποικιλία ανθρώπων, επιχειρηματικών διαδικασιών και εφαρμογών συχνά παράγουν, επεξεργάζονται και αποθηκεύουν δεδομένα του πραγματικού κόσμου, είναι βέβαιο ότι τα δεδομένα θα γίνουν χαοτικά. Αυτό είναι συνήθως αποτέλεσμα χειροκίνητων σφαλμάτων, απρόβλεπτων περιστατικών, τεχνολογικών σφαλμάτων ή διαφόρων άλλων παραγόντων. Οι αλγόριθμοι δεν μπορούν να λαμβάνουν ελλιπή ή θορυβώδη δεδομένα, επειδή συνήθως δεν είναι κατασκευασμένοι για να διαχειρίζονται τις ελλείπουσες τιμές. Και ο θόρυβος διαταράσσει το πραγματικό μοτίβο του δείγματος. Αυτός είναι ο λόγος για τον οποίο απαιτείται προεπεξεργασία δεδομένων για σχεδόν όλους τους τύπους ανάλυσης δεδομένων, επιστήμης δεδομένων και ανάπτυξης τεχνητής νοημοσύνης, ώστε να παράγονται αξιόπιστα, ακριβή και ανθεκτικά ευρήματα για εταιρικές εφαρμογές [34].

Παραδείγματα και τεχνικές προεπεξεργασίας δεδομένων

- **Μετασχηματισμός δεδομένων (Data Transformation)**

Ένα από τα σημαντικότερα στάδια στη φάση της προετοιμασίας είναι ο μετασχηματισμός δεδομένων, ο οποίος μετατρέπει τα δεδομένα από μια μορφή σε μια άλλη. Ορισμένοι αλγόριθμοι απαιτούν την αλλαγή των δεδομένων εισόδου, η οποία αν δεν ολοκληρωθεί, θα υπάρξει κακή απόδοση του μοντέλου και εισαγωγή bias στα δεδομένα. [35]

- **Σχεδιασμός χαρακτηριστικών (Feature Engineering)**

Η στρατηγική σχεδιασμού χαρακτηριστικών χρησιμοποιείται για την παραγωγή καλύτερων χαρακτηριστικών για το σύνολο δεδομένων, τα οποία θα βελτιώσουν την απόδοση του μοντέλου. Συνήθως, χρησιμοποιείται η γνώση του τομέα για την παραγωγή αυτών των χαρακτηριστικών, τα οποία παράγονται χειροκίνητα από υπάρχοντα χαρακτηριστικά, αφού εφαρμοστεί ένας μετασχηματισμός σε αυτά.

- **Ανισόρροπα δεδομένα (Imbalanced Data)**

Ένα από τα πιο διαδεδομένα προβλήματα που μπορεί να αντιμετωπιστούν κατά την κατηγοριοποίηση πραγματικών δεδομένων είναι ότι οι κλάσεις δεν είναι ισορροπημένες (η μία περιέχει περισσότερα δείγματα από την άλλη), με αποτέλεσμα να δημιουργείται σημαντική μεροληψία για το μοντέλο. Με

βάση τα δεδομένα εκπαίδευσής, το 95% του συνόλου δεδομένων αποτελείται από νόμιμες εγγραφές συναλλαγών, ενώ μόλις το 5% αποτελείται από δόλιες συναλλαγές. Με βάση αυτό, το μοντέλο πιθανότατα θα προβλέψει την πλειοψηφική κλάση, αναγνωρίζοντας ως νόμιμες τις δόλιες συναλλαγές.

- **Δεδομένα δειγματοληψίας (Sampling Data)**

Όσο περισσότερα δεδομένα υπάρχουν, τόσο μεγαλύτερη είναι η ακρίβεια του μοντέλου. Παρόλα αυτά, ορισμένοι αλγόριθμοι μηχανικής μάθησης μπορεί να δυσκολεύονται να διαχειριστούν μεγάλο όγκο δεδομένων, με αποτέλεσμα να δημιουργούνται προβλήματα όπως κορεσμός μνήμης και υπολογιστικές αυξήσεις για την ενημέρωση των παραμέτρων του μοντέλου [36].

2.3.3 Τεχνικές Ενίσχυσης (Augmentation Techniques)

Η ενίσχυση δεδομένων είναι μια τεχνική τεχνητής αύξησης του συνόλου εκπαίδευσης με τη δημιουργία τροποποιημένων αντιγράφων ενός συνόλου δεδομένων χρησιμοποιώντας υπάρχοντα δεδομένα. Περιλαμβάνει την πραγματοποίηση μικρών αλλαγών στο σύνολο δεδομένων ή τη χρήση βαθιάς μάθησης για τη δημιουργία νέων δεδομένων. Διακρίνονται δύο είδη ενίσχυσης δεδομένων:

- **Επαυξημένα δεδομένα (Augmented Data)**

Τα επαυξημένα δεδομένα οδηγούνται από τα αρχικά δεδομένα με κάποιες μικρές αλλαγές. Στην περίπτωση της επαύξησης εικόνων, πραγματοποιούμε γεωμετρικούς μετασχηματισμούς και μετασχηματισμούς του χρωματικού χώρου (αναστροφή, αλλαγή μεγέθους, περικοπή, φωτεινότητα, αντίθεση) για να αυξήσουμε το μέγεθος και την ποικιλομορφία του συνόλου εκπαίδευσης.

- **Συνθετικά δεδομένα (Synthetic Data)**

Τα συνθετικά δεδομένα δημιουργούνται τεχνητά χωρίς να χρησιμοποιείται το αρχικό σύνολο δεδομένων. Συχνά χρησιμοποιούνται νευρωνικά δίκτυα βαθιάς μάθησης (Deep Neural Networks – DNNs) και δίκτυα γενίκευσης ανταγωνιστικότητας (Generative Adversarial Networks – GANs) για τη δημιουργία συνθετικών δεδομένων [37].

Τεχνικές Ενίσχυσης Εικόνας (Image Augmentation Techniques):

- Γεωμετρικοί μετασχηματισμοί: τυχαία αναστροφή, περικοπή, περιστροφή, τέντωμα και μεγέθυνση / σμίκρυνση ζουμ εικόνων. Χρειάζεται προσοχή με την εφαρμογή πολλαπλών μετασχηματισμών στις εικόνες, καθώς αυτό μπορεί να μειώσει την απόδοση του μοντέλου.
- Μετασχηματισμοί χρωματικού χώρου: Τυχαία αλλαγή των καναλιών χρώματος RGB, της αντίθεσης και της φωτεινότητας.
- Φίλτρα πυρήνα: Αλλαγή της ευκρίνειας (Sharpness) ή του θολώματος (Blurring) της εικόνας.
- Τυχαία διαγραφή: Διαγραφή κάποιου τμήματος της αρχικής εικόνας.
- Ανάμειξη εικόνων: Ανάμειξη πολλαπλών εικόνων [38].

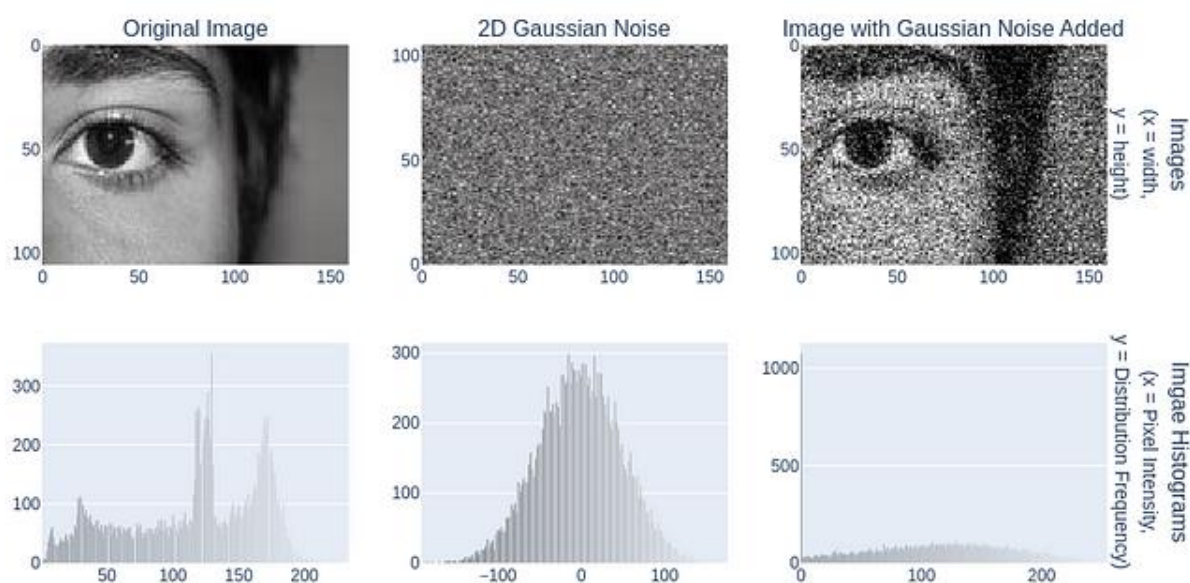
Προσθήκη Θορύβου (Noise Addition)

Η ύπαρξη θορύβου στα δεδομένα είναι αποτέλεσμα ακατάλληλης συλλογής ή αποθήκευσης δεδομένων. Για παράδειγμα, ένα σύνολο δεδομένων σε πίνακες μπορεί να συγκεντρώσει θόρυβο λόγω ανθρώπινων σφαλμάτων, απροσεξία όσον αφορά την ακρίβεια των δεδομένων, αποκοπή δεδομένων,

εσφαλμένη επισήμανση δεδομένων, σφάλματα προγραμματισμού και δόλια παραποίηση δεδομένων. Οι μη ανθρώπινες αιτίες θορύβου περιλαμβάνουν συνήθως προβλήματα με τον αισθητήρα της κάμερας, προβλήματα με το μικρόφωνο και ακατάλληλο φιλτράρισμα που προκαλεί θόρυβο στις εικόνες και τον ήχο.

Στις περισσότερες περιπτώσεις, ο θόρυβος είναι τυχαία κατανομημένος, λόγω του γεγονότος ότι οι διαδικασίες που δημιουργούν το θόρυβο είναι συχνά πολύ πολύπλοκες για να ελεγχθούν και συνεπώς είναι απρόβλεπτες από τη φύση τους. Η τυχαιότητα του θορύβου τον κάνει να διαθέτει στατιστικές ιδιότητες, όπως η κατανομή και η διακύμανση, επιτρέποντας τη χρήση της στατιστικής για τη δημιουργία τεχνητού θορύβου με ορισμένες επιθυμητές ιδιότητες [39].

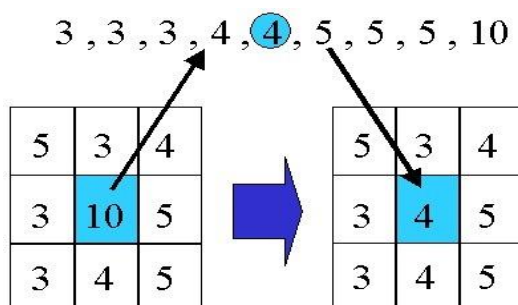
Ο θόρυβος μπορεί να προστεθεί σκόπιμα για την επίδειξη της αντίδρασης του μοντέλου σε δεδομένα με θόρυβο. Τα δεδομένα του πραγματικού κόσμου συχνά περιέχουν κάποια ποσότητα θορύβου. Στις περισσότερες περιπτώσεις είναι σχεδόν αδύνατο να συλλεχθούν δεδομένα χωρίς θόρυβο. Ως εκ τούτου, καθίσταται απαραίτητο για τους αλγόριθμους μηχανικής μάθησης να ανέχονται το θόρυβο και να μαθαίνουν τα υποκείμενα μοτίβα (underlying patterns) που υπάρχουν στα δεδομένα. Δηλαδή, το μοντέλο θα πρέπει να είναι σε θέση να γενικεύει καλά τα θορυβώδη δεδομένα εκπαίδευσης, ώστε να μπορεί να κάνει ικανοποιητικές προβλέψεις στα προφανώς θορυβώδη δεδομένα παραγωγής. Τα συνθετικά δεδομένα συχνά δημιουργούνται για να αντικατοπτρίζουν ιδανικά δεδομένα που δεν περιέχουν θόρυβο και όχι για να αντικατοπτρίζουν δεδομένα του πραγματικού κόσμου που στις περισσότερες περιπτώσεις περιέχουν θόρυβο. Εάν ένα μοντέλο έχει εκπαιδευτεί σε συνθετικά δεδομένα (χωρίς θόρυβο), προσπαθεί να προσαρμοστεί στα δεδομένα εκπαίδευσης και δεν θα γενικευτεί (generalize). Έτσι όταν θα τροφοδοτηθεί με δεδομένα του πραγματικού κόσμου (με θόρυβο), θα αποτύχει να αποδώσει, απλώς και μόνο επειδή δεν είχε ποτέ την ευκαιρία να γενικευτεί κατά τη διάρκεια της εκπαίδευσης. Μερικοί τύποι θορύβου είναι ο Γκαουσιανός θόρυβος (Εικόνα 2.3.3.1), ο Παλμικός θόρυβος, ο θόρυβος αλατιού και πιπεριού και ο θόρυβος ποσοτικοποίησης.



Εικόνα 2.3.3.1: Προσθήκη 2D Θορύβου σε εικόνα που ακολουθεί την γκαουσιανή κατανομή [40]

Εφαρμογή Median Φίλτρου

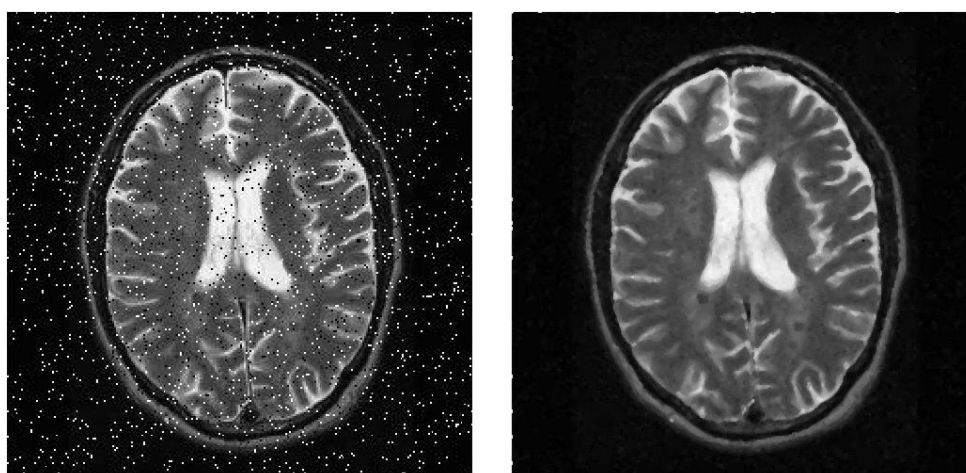
Στον πυρήνα του, το φίλτρο Median είναι μια μη γραμμική τεχνική ψηφιακού φιλτραρίσματος (non-linear digital filtering technique) που χρησιμοποιείται συχνά στην επεξεργασία σήματος και εικόνας. Ο κύριος στόχος αυτού του φίλτρου είναι η μείωση του θορύβου. Ακολουθεί μια μέθοδο αντικατάστασης κάθε εγγραφής στο σήμα ή την εικόνα με τη διάμεσο των γειτονικών εγγραφών. Η γειτονιά γύρω από μια καταχώρηση θα μπορούσε να είναι, για παράδειγμα, τα γύρω εικονοστοιχεία μιας εικόνας ή ένα μικρό χρονικό παράθυρο ψηφιακού ήχου. Με την ταξινόμηση των γειτονικών τιμών στη σειρά και την επιλογή της μεσαίας, το φίλτρο καταστέλλει αποτελεσματικά το θόρυβο χωρίς να εξασθενεί τα αιχμηρά χαρακτηριστικά που έχουν πρωταρχική σημασία για την ακριβή αναπαράσταση, όπως οι ευδιάκριτες ακμές σε μια εικόνα ή οι ακριβείς κυματομορφές σε δεδομένα ήχου.



Εικόνα 2.3.3.2: Διαδικασία Median Filtering σε πίνακα [41]

Επίδραση Median Φίλτρου σε ιατρικές εικόνες

Το φιλτράρισμα διάμεσου χρησιμοποιείται συχνά σε εφαρμογές ιατρικής απεικόνισης για την αφαίρεση του θορύβου από τις εικόνες και τη βελτίωση της ποιότητας της εικόνας. Οι αξονικές τομογραφίες, οι μαγνητικές τομογραφίες και οι ακτίνες X έχουν συχνά θόρυβο λόγω της διαδικασίας απεικόνισης και η μείωση του θορύβου είναι απαραίτητη για την ακριβή ερμηνεία των εικόνων.

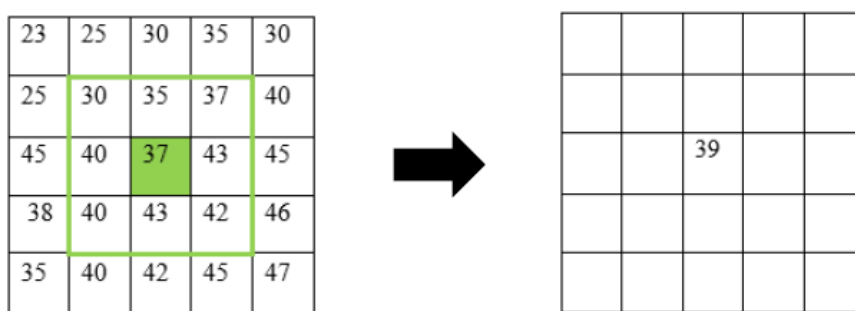


Εικόνα 2.3.3.3: Εφαρμογή Median Φίλτρου σε MRI ανθρώπινου εγκεφάλου με σκοπό την αφαίρεση του τύπου θορύβου αλατιού και πιπεριού (salt – pepper noise) [42]

Εφαρμογή Mean (Moving Average) Φίλτρου

Το φιλτράρισμα του μέσου όρου είναι μια απλή, διαισθητική και εύκολα εφαρμόσιμη μέθοδος εξομάλυνσης εικόνων, δηλαδή μείωσης της διακύμανσης της έντασης μεταξύ ενός εικονοστοιχείου και του επόμενου. Χρησιμοποιείται συχνά για τη μείωση του θορύβου σε εικόνες.

Η ιδέα του φιλτραρίσματος μέσου όρου είναι απλά η αντικατάσταση κάθε τιμής εικονοστοιχείου σε μια εικόνα με τη μέση τιμή (μέσος όρος) των γειτονικών του σημείων, συμπεριλαμβανομένου του εαυτού του. Αυτό έχει ως αποτέλεσμα την εξάλειψη των τιμών των εικονοστοιχείων που δεν αντιπροσωπεύουν το περιβάλλον τους. Το φιλτράρισμα μέσου όρου θεωρείται συνήθως ως φίλτρο συνέλιξης. Όπως και οι άλλες συνελίξεις, βασίζεται σε έναν πυρήνα, ο οποίος αντιπροσωπεύει το σχήμα και το μέγεθος της γειτονιάς που πρέπει να δειγματοληπτηθεί κατά τον υπολογισμό του μέσου όρου. Συχνά χρησιμοποιείται ένας τετραγωνικός πυρήνας 3×3 , αν και μπορούν να χρησιμοποιηθούν μεγαλύτεροι πυρήνες (π.χ. 5×5 τετράγωνα) για πιο σοβαρή εξομάλυνση [43].



Εικόνα 2.3.3.4: Διαδικασία Mean Filtering σε πίνακα [44]

2.4 Αποδοτικότητα μοντέλου (Model Performance)

2.4.1 Αξιολόγηση μοντέλου (Model Evaluation)

Η αξιολόγηση των μοντέλων είναι ένα κρίσιμο βήμα στη ροή εργασίας της μηχανικής μάθησης, μεταξύ άλλων και στα συνελκτικά νευρωνικά δίκτυα (CNN). Παίζει καθοριστικό ρόλο στη διασφάλιση ότι τα μοντέλα που αναπτύσσονται είναι αποτελεσματικά, αξιόπιστα και κατάλληλα για ανάπτυξη εφαρμογών του πραγματικού κόσμου [45].

Σημασία των μετρικών αξιολόγησης

Οι μετρικές αξιολόγησης αποτελούν θεμελιώδη εργαλεία για την ανάπτυξη και των συνεπτυγμένων νευρωνικών δικτύων (CNN). Παρέχουν ποσοτικά μέτρα για την αξιολόγηση του πόσο καλά αποδίδει ένα CNN σε μια δεδομένη εργασία, όπως η ταξινόμηση εικόνων (image classification) και η ανίχνευση αντικειμένων (object detection). Οι λόγοι που τις καθιστούν χρήσιμες αναλύονται παρακάτω.

- **Αξιολόγηση επιδόσεων (Performance Assessment)**

Οι μετρικές αξιολόγησης παρέχουν έναν σαφή και αντικειμενικό τρόπο μέτρησης της απόδοσης των CNN. Μετρικές όπως η ακρίβεια, η ανάκληση και το σκορ F1 προσφέρουν πληροφορίες σχετικά με διάφορες πτυχές της αποτελεσματικότητας του μοντέλου, βοηθώντας να κατανοήσουμε πόσο καλά γενικεύεται σε αόρατα δεδομένα.

- **Σύγκριση μοντέλων (Model Comparison)**

Αυτές οι μετρικές επιτρέπουν τη σύγκριση διαφορετικών μοντέλων ή διαμορφώσεων. Αξιολογώντας πολλαπλά μοντέλα με αυτές τις μετρικές, οι επαγγελματίες μπορούν να προσδιορίσουν αντικειμενικά ποιο μοντέλο αποδίδει καλύτερα για μια συγκεκριμένη εργασία ή σύνολο δεδομένων.

- **Καθοδήγηση της βελτίωσης του μοντέλου (Model Improvement)**

Οι μετρικές είναι απαραίτητες για τον εντοπισμό των δυνατών και αδύναμων σημείων του μοντέλου. Για παράδειγμα, μια χαμηλή ανάκληση μπορεί να υποδηλώνει ότι το μοντέλο χάνει τις σχετικές περιπτώσεις, γεγονός που υποδηλώνει την ανάγκη προσαρμογής της διαδικασίας εκπαίδευσης ή της αρχιτεκτονικής του μοντέλου.

- **Ρύθμιση υπερπαραμέτρων (hyper-parameter tuning)**

Οι μετρικές αξιολόγησης παρέχουν ανατροφοδότηση που είναι κρίσιμη για τη ρύθμιση των υπερπαραμέτρων. Με τη συστηματική προσαρμογή των παραμέτρων και την παρατήρηση των αλλαγών στις μετρικές, μπορεί κανείς να βελτιστοποιήσει το μοντέλο ώστε να επιτύχει την καλύτερη δυνατή απόδοση [46].

2.4.2 Μετρικές Αξιολόγησης (Evaluation Metrics)

Σέ ένα πρόβλημα ταξινόμησης, κύριος στόχος είναι να προβλέψουμε τη μεταβλητή-στόχο η οποία έχει τη μορφή διακριτών τιμών και ονομάζεται κλάση. Για την αξιολόγηση της απόδοσης ενός τέτοιου μοντέλου υπάρχουν μετρικές όπως η Ακρίβεια ταξινόμησης (Classification Accuracy), Περιοχή κάτω από την καμπύλη (Area Under Curve – AUC), F1-score, Ακρίβεια (Precision), Ανάκληση (Recall)/Εναισθησία (Sensitivity), Πίνακας σύγχυσης (Confusion Matrix).

1. Αληθές θετικό (True Positive – TP): Έγινε πρόβλεψη για κάτι θετικό και είναι αλήθεια.
2. Αληθές Αρνητικό (True Negative – TN): Έγινε πρόβλεψη για κάτι αρνητικό και είναι αλήθεια.
3. Ψευδώς θετικό (False Positive – FP): (Σφάλμα τύπου 1): Έγινε πρόβλεψη για κάτι θετικό και είναι ψευδές.
4. Ψευδώς αρνητικό (False Negative – FN): (Σφάλμα τύπου 2): Έγινε πρόβλεψη για κάτι αρνητικό και είναι ψευδές.

- **Ακρίβεια ταξινόμησης (Classification Accuracy)**

Η ακρίβεια ταξινόμησης είναι μια θεμελιώδης μετρική για την αξιολόγηση της απόδοσης του μοντέλου ταξινόμησης, παρέχοντας μια γρήγορη εικόνα του πόσο καλά αποδίδει το μοντέλο όσον αφορά τις σωστές προβλέψεις. Υπολογίζεται ως ο λόγος των σωστών προβλέψεων προς τον συνολικό αριθμό των δειγμάτων εισόδου.

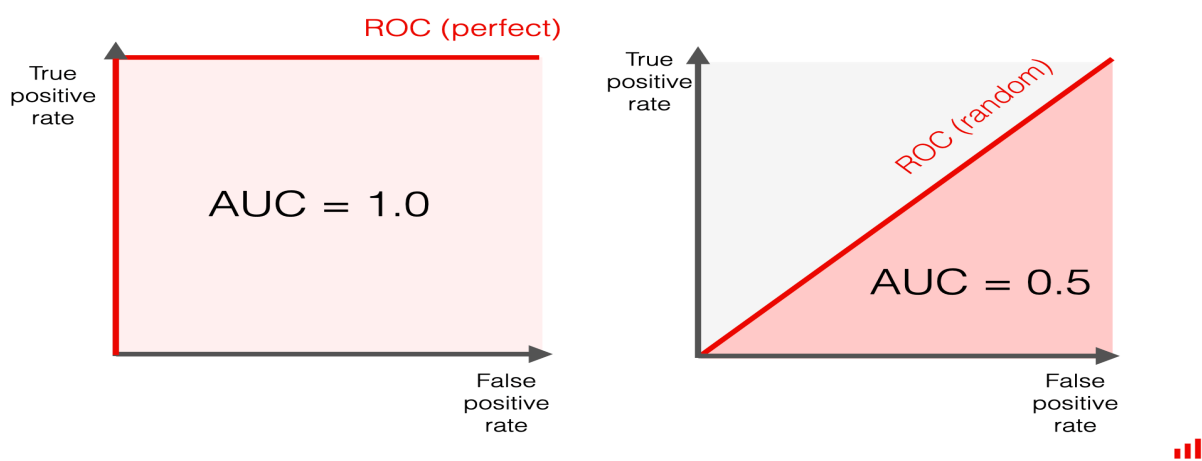
$$\text{Ακρίβεια (Accuracy)} = \frac{\text{Αριθμός σωστών προβλέψεων}}{\text{Συνολικός αριθμός δειγμάτων εισόδου}} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}} \quad (1)$$

Λειτουργεί εξαιρετικά εάν υπάρχει ίσος αριθμός δειγμάτων για κάθε κλάση. Η ακρίβεια ταξινόμησης είναι καλή, αλλά δίνει μία ψευδώς θετική αίσθηση για την επίτευξη υψηλής ακρίβειας. Το πρόβλημα

προκύπτει λόγω της πιθανότητας να υπάρξει λανθασμένη ταξινόμηση δειγμάτων υψηλού όγκου των δευτερευουσών κλάσεων [47].

- **Περιοχή κάτω από την καμπύλη (AUC)**

Είναι μία από τις ευρέως χρησιμοποιούμενες μετρικές και βασικά χρησιμοποιείται για δυαδική ταξινόμηση. Η AUC ενός ταξινομητή ορίζεται ως η πιθανότητα ένας ταξινομητής να κατατάξει ένα τυχαία επιλεγμένο θετικό παράδειγμα υψηλότερα από ένα αρνητικό παράδειγμα. Όπως φαίνεται στα αριστερά της εικόνας 2.4.2.1, η καμπύλη ROC για ένα τέλειο μοντέλο θα πήγαινε από το (0,0) στο (0,1) και στη συνέχεια ευθεία στο (1,1). Αυτό δείχνει ότι το μοντέλο επιτυγχάνει 100% αληθές θετικό ποσοστό με 0% ψευδές θετικό ποσοστό σε κάποιο σημείο κατά τη διάρκεια της διαδικασίας πρόβλεψης. Στα δεξιά της εικόνας βλέπουμε μία καμπύλη ROC με μια διαγώνια γραμμή από το (0,0) στο (1,1). Αυτό αντιπροσωπεύει καμία ικανότητα διάκρισης μεταξύ θετικών και αρνητικών κλάσεων, με το ποσοστό ψευδώς θετικών να αυξάνεται με τον ίδιο ρυθμό όπως το ποσοστό αληθώς θετικών.



Εικόνα 2.4.2.1: AUC Περιοχή κάτω από την καμπύλη ROC (Receiver Operating Characteristic Curve) [48]

- **Αληθώς θετικό ποσοστό**

Ονομάζεται, επίσης, ευαισθησία (sensitivity). Το ποσοστό αληθώς θετικών δεδομένων θεωρείται ως το ποσοστό των θετικών σημείων δεδομένων που θεωρούνται σωστά θετικά, σε σχέση με όλα τα θετικά σημεία δεδομένων.

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

- **Αληθώς αρνητικό ποσοστό**

Ονομάζεται, επίσης, ειδικότητα (specificity). Ψευδώς αρνητικό ποσοστό θεωρείται ως το τμήμα των αρνητικών σημείων δεδομένων που θεωρούνται σωστά ως αρνητικά, σε σχέση με όλα τα σημεία δεδομένων που είναι αρνητικά.

$$TNR = \frac{TN}{TN+FP} \quad (3)$$

- **Ψευδώς θετικό ποσοστό**

Το ποσοστό ψευδώς αρνητικών είναι στην πραγματικότητα το ποσοστό των πραγματικών θετικών που αναγνωρίζονται εσφαλμένα ως αρνητικά

$$FPR = \frac{FP}{FP+TN} \quad (4)$$

Το ποσοστό ψευδώς θετικών και το ποσοστό αληθώς θετικών έχουν και οι δύο τιμές στο εύρος [0, 1]. Έτσι, το AUC είναι μια καμπύλη που σχεδιάζεται μεταξύ False Positive Rate και True Positive Rate σε όλα τα διαφορετικά σημεία δεδομένων με εύρος τιμών [0, 1]. Όσο μεγαλύτερη είναι η τιμή του AUC τόσο καλύτερη είναι η απόδοση του μοντέλου.

- **F1-score**

Είναι ένας αρμονικός μέσος όρος μεταξύ ανάκλησης και ακρίβειας. Το εύρος του είναι [0,1]. Αυτή η μετρική συνήθως λέει πόσο ακριβής (ταξινομεί σωστά πόσες περιπτώσεις) και εύρωστος (δεν χάνει σημαντικό αριθμό περιπτώσεων) είναι ο ταξινομητής.

$$F1 \text{ Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7) \quad [49]$$

- **Ακρίβεια (Precision)**

Υπάρχει μια άλλη μετρική που ονομάζεται Precision. Η ακρίβεια είναι ένα μέτρο της απόδοσης του μοντέλου που λέει πόσες από τις θετικές προβλέψεις που κάνει το μοντέλο είναι πραγματικά σωστές. Υπολογίζεται ως ο αριθμός των αληθώς θετικών προβλέψεων διαιρούμενος με τον αριθμό των αληθώς θετικών και των ψευδώς θετικών προβλέψεων.

$$\text{Ακρίβεια} = \frac{TP}{TP+FP} \quad (5)$$

- **Ανάκληση (Recall) ή Ευαισθησία (Sensitivity)**

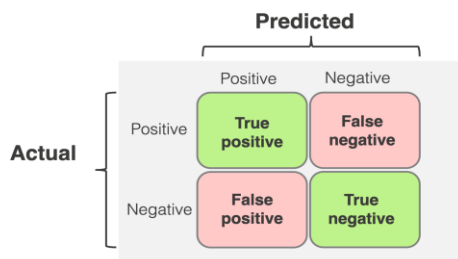
Η ανάκληση επικεντρώνεται στην πληρότητα της ανίχνευσης θετικών κλάσεων. Μια υψηλή ανάκληση δείχνει ότι το μοντέλο καταφέρνει να συλλάβει τις περισσότερες θετικές περιπτώσεις στα δεδομένα. Ωστόσο, η ανάκληση δεν εξετάζει το ποσοστό των ψευδώς θετικών, το οποίο είναι ο ρόλος της ακρίβειας.

$$\text{Ανάκληση} = \frac{TP}{TP+FN} \quad (6)$$

Η χαμηλότερη ανάκληση και η υψηλότερη ακρίβεια δίνουν μεγάλη ακρίβεια, αλλά στη συνέχεια χάνετε μεγάλο αριθμό περιπτώσεων. Όσο μεγαλύτερο είναι το σκορ F1 τόσο καλύτερη θα είναι η απόδοση. Μπορεί να εκφραστεί μαθηματικά με αυτόν τον τρόπο:

- **Πίνακας σύγχυσης (Confusion Matrix)**

Δημιουργεί έναν πίνακα N x N, όπου N είναι ο αριθμός των κλάσεων ή κατηγοριών που πρόκειται να προβλεφθούν. Αν N=2, προκύπτει ένας πίνακας 2x2. Με την υπόθεση ότι υπάρχει ένα πρόβλημα μιας δυαδικής ταξινόμησης, τα δείγματα αυτής της ταξινόμησης ανήκουν είτε στο Ναι είτε στο Όχι. Έτσι, κατασκευάζεται ο ταξινομητής, ο οποίος θα προβλέψει την κατηγορία για το νέο δείγμα εισόδου. Μετά από αυτό, δοκιμάζεται το μοντέλο με δείγματα και προκύπτει το ακόλουθο αποτέλεσμα. Στην εικόνα 2.4.2.2 βλέπουμε την δομή του πίνακα σύγχυσης.



■

Εικόνα 2.4.2.2: Οπτικοποίηση πίνακα σύγχυσης (Confusion Matrix) [50]

- **Τομή Διασταύρωσης και Ένωσης (Intersection over Union – IoU)**

Μια μετρική ακρίβειας, η IoU αξιολογεί την τομή δύο οριοθετημένων κουτιών, το προβλεπόμενο και το κουτί της βασικής αλήθειας (bounding boxes). Η μετρική προέρχεται από τον δείκτη Jaccard (Jaccard Index).

$$\text{Intersection over Union (IoU)} = \frac{A \cap B}{A \cup B} = \frac{TP}{TP+FN+FP} \quad (8) \quad [51]$$



Εικόνα 2.4.2.3: Οπτικοποίηση Intersection over Union – IoU [52]

2.4.3 Συναρτήσεις Απώλειας (Loss Functions)

Η συνάρτηση απώλειας, που αναφέρεται επίσης ως συνάρτηση σφάλματος, είναι ένα κρίσιμο στοιχείο στη μηχανική μάθηση που ποσοτικοποιεί τη διαφορά μεταξύ των προβλεπόμενων εξόδων του αλγορίθμου μηχανικής μάθησης και των πραγματικών τιμών-στόχων. Ο επίσημος όρος για αυτή την αριθμητική ποσοτικοποίηση είναι το σφάλμα πρόβλεψης (**prediction error**). Ο αλγόριθμος μάθησης (learning algorithm) και οι μηχανισμοί σε ένα μοντέλο μηχανικής μάθησης βελτιστοποιούνται για την ελαχιστοποίηση του σφάλματος πρόβλεψης. Αυτό σημαίνει ότι μετά τον υπολογισμό της τιμής για τη συνάρτηση απώλειας, η οποία καθορίζεται από το σφάλμα πρόβλεψης, ο αλγόριθμος μάθησης αξιοποιεί αυτές τις πληροφορίες για να πραγματοποιήσει ενημερώσεις βαρών και παραμέτρων (weight – parameters update), οι οποίες χρησιμοποιούνται κατά τη διάρκεια του επόμενου περάσματος εκπαίδευσης οδηγώντας σε χαμηλότερο σφάλμα πρόβλεψης. Δύο από τις δημοφιλέστερες συναρτήσεις σφάλματος που χρησιμοποιούνται για ταξινόμηση (classification) είναι οι Binary Cross-Entropy Loss (Log Loss) και Categorical Cross-Entropy Loss [53].

- Binary Cross-Entropy Loss (Log Loss):

Η δυαδική απώλεια διασταυρούμενης εντροπίας (ή απώλεια λογαρίθμου) είναι μια ποσοτικοποίηση της διαφοράς μεταξύ της πρόβλεψης του αλγορίθμου μηχανικής μάθησης και της πραγματικής πρόβλεψης του στόχου, η οποία υπολογίζεται από την αρνητική τιμή του αθροίσματος της λογαριθμικής τιμής των πιθανοτήτων των προβλέψεων του αλγορίθμου μηχανικής μάθησης σε σχέση με το συνολικό αριθμό δειγμάτων δεδομένων. Τύπος υπολογισμού BCE:

$$L(y, f(x)) = - [y * \log(f(x)) + (1 - y) * \log(1 - f(x))],$$

Όπου:

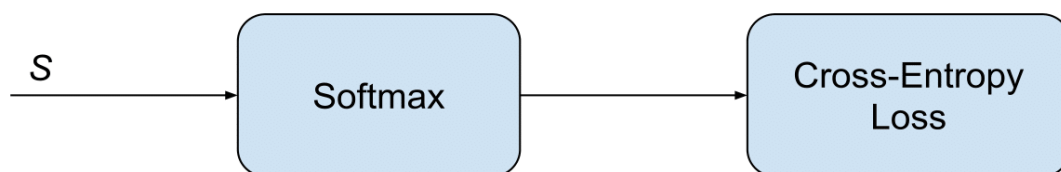
Το L αντιπροσωπεύει τη δυαδική συνάρτηση απώλειας Cross-Entropy,

y είναι η πραγματική δυαδική ετικέτα (0 ή 1),

f(x) είναι η προβλεπόμενη πιθανότητα της θετικής κλάσης (μεταξύ 0 και 1).

- Categorical Cross-Entropy Loss:

Ονομάζεται και Softmax Loss. Είναι μια ενεργοποίηση Softmax συν μια απώλεια Cross-Entropy (βλ. Εικόνα 2.4.3.1). Αν χρησιμοποιήσουμε αυτή την απώλεια, θα εκπαιδεύσουμε ένα CNN να εξάγει μια πιθανότητα πάνω στην C κατηγορίες για κάθε εικόνα. Χρησιμοποιείται για ταξινόμηση πολλαπλών κλάσεων. Στην ειδική (και συνηθισμένη) περίπτωση της ταξινόμησης πολλαπλών κλάσεων οι ετικέτες είναι μονοσήμαντες, οπότε μόνο η θετική κλάση C_p διατηρεί τον όρο όπως στην απώλεια. Υπάρχει μόνο ένα στοιχείο του διανύσματος στόχου t που δεν είναι μηδενικό $t_i=t_p$. Έτσι, απορρίπτοντας τα στοιχεία του αθροίσματος που είναι μηδενικά λόγω των ετικετών-στόχων, μπορούμε να γράψουμε: Όπου S_p είναι το σκορ CNN για τη θετική κλάση.



Εικόνα 2.4.3.1: Οπτικοποίηση SoftMax Loss.

Ο τύπος του SoftMax Activation είναι:

$$f(s)_i = \frac{e^{S_i}}{\sum_j^c e^{S_j}} \quad (1)$$

Ο τύπος του Cross-Entropy Loss είναι:

$$CE = - \sum_i^C t_i * \log(f(s)_i) \quad (2)$$

Στην ειδική περίπτωση της ταξινόμησης πολλαπλών κλάσεων οι ετικέτες είναι one-hot, οπότε μόνο η θετική κλάση C_p διατηρεί τον όρο της στην απώλεια. Υπάρχει μόνο ένα στοιχείο του διανύσματος Target t που δεν είναι μηδενικό $t_i = t_p$.

Μετά από αντικατάσταση της (1) στη (2) και απλοποίηση του τύπου της CE, έχουμε:

$$CE = -\log \left(\frac{e^{S_p}}{\sum_j^c e^{S_j}} \right), \text{όπου } S_p \text{ είναι το Score της θετικής (positive) κλάσης [54].}$$

2.4.4 Αλγόριθμοι βελτιστοποίησης (Optimization Algorithms)

Οι αλγόριθμοι βελτιστοποίησης είναι μαθηματικές μέθοδοι που χρησιμοποιούνται για την προσαρμογή των παραμέτρων του μοντέλου ώστε να ελαχιστοποιηθεί ή να μεγιστοποιηθεί μια αντικειμενική συνάρτηση. Στη μηχανική μάθηση, οι αλγόριθμοι αυτοί διαδραματίζουν καθοριστικό ρόλο στην εκπαίδευση των μοντέλων, βρίσκοντας το βέλτιστο σύνολο παραμέτρων που οδηγεί στην καλύτερη απόδοση. Είναι θεμελιώδεις για την επιτυχία διαφόρων μοντέλων μηχανικής μάθησης, συμπεριλαμβανομένων των νευρωνικών δικτύων [55].

- **Αποτελεσματική εκπαίδευση (Efficient Training):**

Οι αλγόριθμοι βελτιστοποίησης εξορθολογούν τη διαδικασία εκπαίδευσης, επιτρέποντας στα μοντέλα να μαθαίνουν από τα δεδομένα αποτελεσματικά. Βοηθούν στη γρήγορη εύρεση των καλύτερων παραμέτρων, μειώνοντας τον απαιτούμενο χρόνο και τους υπολογιστικούς πόρους.

- **Βελτιωμένη απόδοση (Improved Performance)**

Με τη λεπτομερή ρύθμιση των παραμέτρων του μοντέλου, οι αλγόριθμοι αυτοί βελτιώνουν την ακρίβεια και την ικανότητα γενίκευσης των μοντέλων μηχανικής μάθησης. Η αποτελεσματική βελτιστοποίηση οδηγεί σε μοντέλα που αποδίδουν καλά τόσο σε δεδομένα εκπαίδευσης όσο και σε νέα δεδομένα δοκιμής (unseen test data).

- **Σύγκλιση σε βέλτιστες λύσεις (Convergence to optimal solutions)**

Οι καλοί αλγόριθμοι βελτιστοποίησης διασφαλίζουν ότι η διαδικασία εκπαίδευσης συγκλίνει σε ένα παγκόσμιο ή ικανοποιητικό τοπικό ελάχιστο της συνάρτησης απώλειας, αποφεύγοντας τις κακές λύσεις που θα μπορούσαν να προκύψουν από μη βέλτιστες ρυθμίσεις παραμέτρων.

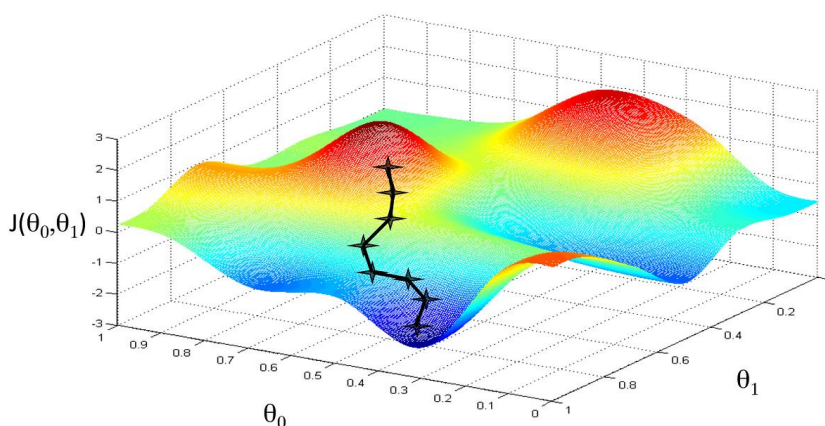
Παρακάτω αναλύονται δύο από τις πιο συχνά χρησιμοποιούμενους αλγορίθμους βελτιστοποίησης στα μοντέλα CNN:

- **Διαβάθμιση Κλίσης (Gradient Descent)**

Είναι ένας θεμελιώδης αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την ελαχιστοποίηση της αντικειμενικής συνάρτησης με επαναληπτική κίνηση προς το

ελάχιστο. Είναι ένας επαναληπτικός αλγόριθμος πρώτης τάξης (first-order) για την εύρεση του τοπικού ελαχίστου μιας διαφορίσιμης συνάρτησης πολλαπλών μεταβλητών. Ο αλγόριθμος λειτουργεί κάνοντας επαναλαμβανόμενα βήματα προς την αντίθετη κατεύθυνση της κλίσης (ή της προσεγγιστικής κλίσης) της συνάρτησης στο τρέχον σημείο, επειδή αυτή είναι η κατεύθυνση της πιο απότομης καθόδου, όπως φαίνεται και στην εικόνα 2.4.4.1.

- **Στοχαστική Διαβάθμιση Κλίσης (SGD):** Αυτή η παραλλαγή προτείνει την ενημέρωση του μοντέλου με τη χρήση ενός μόνο παραδείγματος εκπαίδευσης κάθε φορά, η οποία δεν απαιτεί μεγάλο όγκο υπολογισμών και, επομένως, είναι κατάλληλη για μεγάλα σύνολα δεδομένων. Είναι στοχαστική και μπορεί να παράγει θορυβώδεις ενημερώσεις για αυτό μπορεί να απαιτεί προσεκτική επιλογή των ρυθμών μάθησης (learning rates) [56].



Εικόνα 2.4.4.1: Οπτικοποίηση της Gradient Descent, ξεκινά από ένα σημείο υψηλής απώλειας και μέσω πολλαπλών επαναλήψεων, κάνει βήματα προς την κατεύθυνση της χαμηλότερης απώλειας, με στόχο να βρει τη βέλτιστη διαμόρφωση βάρους [57].

- **Adam (Adaptive Moment Estimation):**

Είναι μια ευρέως χρησιμοποιούμενη τεχνική βελτιστοποίησης. Σε κάθε χρονικό βήμα, ο αλγόριθμος Adam παρακολουθεί τόσο τις κλίσεις όσο και τον κινητό μέσο όρο των δευτέρων ροπών. Χρησιμοποιείται για την τροποποίηση του ρυθμού μάθησης για κάθε παράμετρο της διαδικασίας. Είναι υπολογιστικά αποδοτικός, έχει μικρές απαιτήσεις μνήμης και είναι ιδιαίτερα χρήσιμος για μεγάλα δεδομένα και παραμέτρους.

Το AdaMax είναι μια επέκταση της έκδοσης Adam του Gradient Descent που γενικεύει την προσέγγιση στην infinity norm (το μέγιστο των απόλυτων - max) και μπορεί να οδηγήσει σε αποτελεσματικότερη βελτιστοποίηση σε ορισμένα προβλήματα. Το AdaMax είναι μια επέκταση της έκδοσης Adam διαβάθμισης κλίσης που έχει σχεδιαστεί για να επιταχύνει τη διαδικασία βελτιστοποίησης [58].

ΚΕΦΑΛΑΙΟ 3: Προτεινόμενο Πλαίσιο Διάγνωσης Ασθένειας με χρήση CNN Δικτύων σε Ιατρικές Εικόνες

3.1 Επισκόπηση Πλαισίου

3.1.1 Υπόβαθρο και Κίνητρο

Ο καρκίνος του πνεύμονα θεωρείται ο πιο θανατηφόρος καρκίνος παγκοσμίως. Για το λόγο αυτό, οι χώρες αναπτύσσουν στρατηγικές για την έγκαιρη διάγνωση του καρκίνου του πνεύμονα. Η μελέτη NLST, έδειξε ότι τρεις ετήσιοι κύκλοι προσυμπτωματικού ελέγχου σε άτομα υψηλού κινδύνου με χρήση χαμηλής δόσης υπολογιστικής τομογραφίας (CT) μειώνουν σημαντικά τα ποσοστά θανάτου. Τα μέτρα αυτά σημαίνουν ότι μια συντριπτική ποσότητα εικόνων αξονικής τομογραφίας θα πρέπει να ελεγχθεί από έναν ακτινολόγο. Δεδομένου ότι οι όγκοι είναι πολύ δύσκολο να εντοπιστούν, ακόμη και από έμπειρους γιατρούς, η επιβάρυνση των ακτινολόγων αυξάνεται σημαντικά με τον αριθμό των αξονικών τομογραφιών που πρέπει να αναλυθούν.

Με την αναμενόμενη αύξηση του αριθμού των μέτρων πρόωμης ανίχνευσης, οι επιστήμονες εργάζονται σε ηλεκτρονικές λύσεις που συμβάλλουν στην ελάφρυνση του έργου των γιατρών, στη βελτίωση της ακρίβειας της διάγνωσης με τη μείωση του παράγοντα υποκειμενικότητας, στην επιτάχυνση της ανάλυσης και στη μείωση του ιατρικού κόστους [59].

Για την ανίχνευση κακοήθων όγκων, πρέπει να αναγνωριστούν και να μετρηθούν συγκεκριμένα χαρακτηριστικά. Με βάση τα αναγνωρισμένα χαρακτηριστικά και τον συνδυασμό τους, μπορεί να εκτιμηθεί η πιθανότητα καρκίνου [60]. Στην παρούσα εργασία, παρουσιάζουμε μια ανασκόπηση κάποιων πρόσφατων τεχνικών βαθιάς μάθησης για την ανίχνευση καρκίνου του πνεύμονα και συγκεκριμένα των βαθιών CNN δικτύων. Τα CNN είναι μια κατηγορία νευρωνικών δικτύων που έχουν σχεδιαστεί για να μαθαίνουν, κατά τη διάρκεια της εκπαίδευσης, κάποιες παραμέτρους συνέλιξης από ένα σύνολο διαθέσιμων δεδομένων. Σε γενικές γραμμές, αποτελούνται από διαφορετικά επίπεδα, όπως τα επίπεδα συνελίξεων, τα επίπεδα συγκέντρωσης και άλλα που έχουν ήδη αναφερθεί νωρίτερα. Τα τελευταία χρόνια προτάθηκαν διάφορες αρχιτεκτονικές για τη βελτίωση των επιδόσεων και την υπέρβαση ορισμένων περιορισμών των τυπικών CNN. Ένας από τους παράγοντες που επηρεάζουν την απόδοση των CNN είναι η ύπαρξη θορύβου στα δεδομένα.

3.1.2 Σημασία απαλοιφής θορύβου

Τα CNN δίκτυα χρησιμοποιούνται ευρέως για την ταξινόμηση εικόνων (image classification). Οι θορυβώδεις εικόνες μειώνουν την απόδοση ταξινόμησης των συνελκτικών νευρωνικών δικτύων και αυξάνουν το χρόνο εκπαίδευσης των δικτύων [61]. Ειδικότερα, όσον αφορά την ύπαρξη θορύβου στις χαμηλής δόσης υπολογιστικές τομογραφίες (CT Scans), μία λανθασμένη πρόβλεψη θα μπορούσε να αποδειχθεί καταστροφική για τον ασθενή. Για αυτό το λόγο, στην συγκεκριμένη εργασία, θα δειχθούν τρόποι προεπεξεργασίας δεδομένων και εφαρμογές φίλτρου, με σκοπό την αφαίρεση του θορύβου από

Αξιολόγηση της απόδοσης CNN κατά την ανίχνευση και κατηγοριοποίηση αντικειμένων

στις τομογραφίες και στην συνέχεια θα αναλυθούν τα αποτελέσματα από την εκτέλεση συγκεκριμένων CNN μοντέλων.

3.1.3 Σενάρια Χρήσης του Πλαισίου Διάγνωσης Ασθένειας

Πρωταρχικός στόχος της συγκεκριμένης εργασίας αποτελεί η σύγκριση κάποιων CNN αρχιτεκτονικών με σκοπό την λήψη συμπερασμάτων για την καταλληλότητα χρησιμοποίησής τους για προβλέψεις καλοθών ή κακοθών όγκων καρκίνου του πνεύμονα. Αναλυτικότερα, θα συγκριθούν:

- Οι σύγχρονες αρχιτεκτονικές VGG και EfficientNetB3 στα αρχικά δεδομένα (original data)
- Η επιρροή του γκαουσιανού και κοκκώδη θορύβου στην αρχιτεκτονική VGG.
- Η απόδοση του VGG σε αποθορυβοποιημένα δεδομένα με την εφαρμογή φίλτρου Median και Moving Average
- Η απόδοση του VGG σε αποθορυβοποιημένα δεδομένα με διαδοχική εφαρμογή φίλτρου Median και Moving Average

Μέσα από αυτήν την μελέτη, διερευνώνται οι επιπτώσεις του θορύβου και των τεχνικών απαλοιφής θορύβου στην αποδοτικότητα των συνελκτικών μοντέλων.

3.2 Περιγραφή Πλαισίου

Η ροή εργασίας ξεκινά με την εισαγωγή των απαραίτητων βιβλιοθηκών, ακολουθούμενη από τη φόρτωση και την προεπεξεργασία του συνόλου δεδομένων, τη διασφάλιση της ισορροπίας και τον διαχωρισμό του σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής.

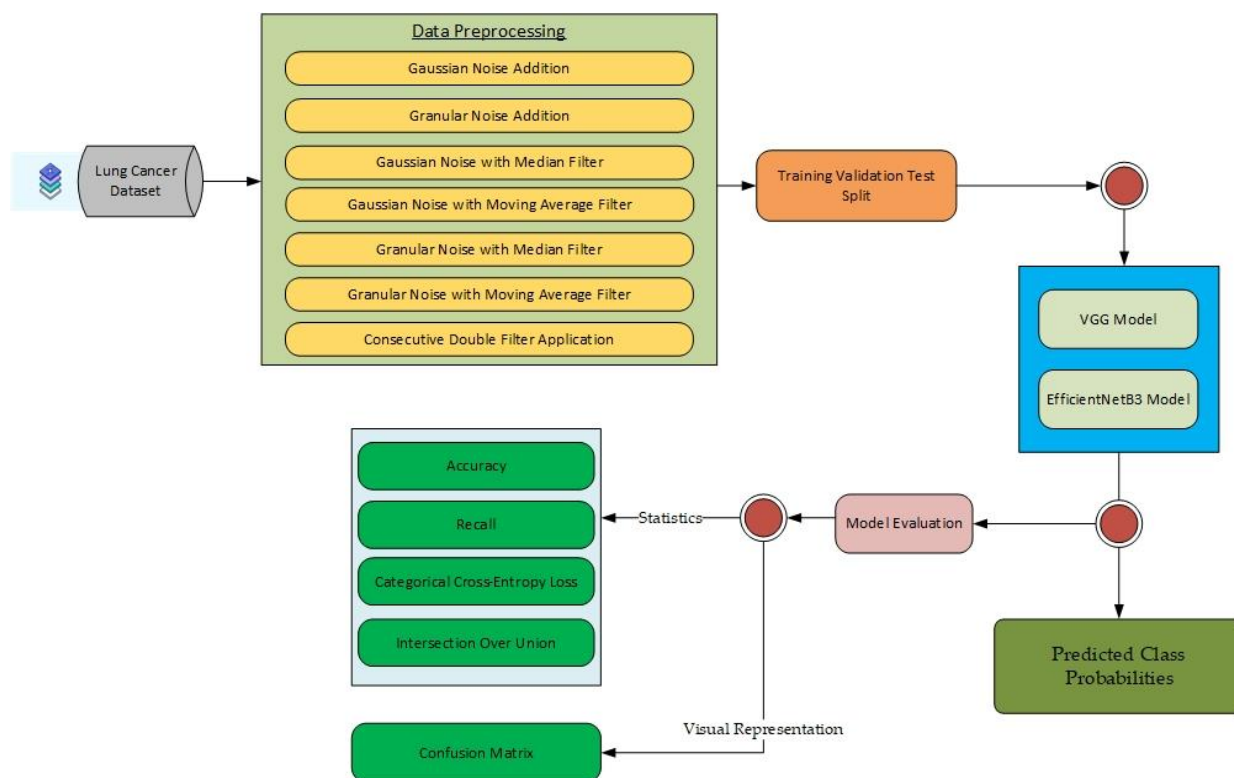
Μέρος της προεπεξεργασίας αποτελούν:

- Προσθήκη Gaussian και Granular Noise
- Απαλοιφή Gaussian Noise με Median Filtering
- Απαλοιφή Gaussian Noise με Moving Average Filtering
- Απαλοιφή Granular Noise με Median Filtering
- Απαλοιφή Granular Noise με Moving Average Filtering
- Διαδοχική απαλοιφή Gaussian Noise με Moving Average Filtering και Granular Noise με Median Filtering

Ο θόρυβος που προστίθεται σε κάθε περίπτωση είναι ίδιου μεγέθους 0.3% (Noise Scale 0.3). Οι γεννήτριες δεδομένων εικόνας μετατρέπουν τις εικόνες σε πίνακες για την εκπαίδευση. Ένα μοντέλο CNN κατασκευάζεται με τη χρήση συνελκτικών και πυκνών μπλοκ, εκπαιδεύεται και αξιολογείται

για ακρίβεια, ανάκληση και IoU. Επιπλέον, ένα προ-εκπαιδευμένο μοντέλο EfficientNetB3 τελειοποιείται.

Και τα δύο μοντέλα αξιολογούνται μέσω διαγραμμάτων επιδόσεων, πινάκων σύγκρισης και μετρικών αξιολόγησης για τη μέτρηση της απόδοσης ταξινόμησης σε εικόνες πνευμόνων. Η ροή του πλαισίου διάγνωσης ασθενείας παρουσιάζεται στην Εικόνα 3.2.1.



Εικόνα 3.2.1: Διάγραμμα Ροής Εργασιών

3.2.1 Δεδομένα και Προεπεξεργασία

Περιγραφή του συνόλου δεδομένων που χρησιμοποιήθηκε

Αυτό το σύνολο δεδομένων [62] περιέχει 25.000 ιστοπαθολογικές εικόνες με 3 κλάσεις. Οι εικόνες έχουν μέγεθος 768 x 768 και είναι σε μορφή αρχείου JPEG. Οι εικόνες δημιουργήθηκαν από ένα αρχικό δείγμα συμβατών με το HIPAA και επικυρωμένων πηγών, το οποίο αποτελείται από 750 συνολικές εικόνες πνευμονικού ιστού (250 καλοήθεις πνευμονικοί ιστοί, 250 αδενοκαρκινώματα πνεύμονα και 250 πλακώδη καρκινώματα πνεύμονα) και αυξήθηκαν σε 15.000 χρησιμοποιώντας ενίσχυση του συνόλου δεδομένων, μέσω ενίσχυσης θέσης (Position Augmentation), χρώματος (Color Augmentation) και θολώματος (Blurring). Υπάρχουν τρεις κλάσεις στο σύνολο δεδομένων, καθεμία με 5.000 εικόνες, οι οποίες είναι οι εξής:

- Καλοήθης ιστός πνεύμονα (Benign Tissue)
- Αδενοκαρκίνωμα πνεύμονα (Adenocarcinoma)
- Πλακώδες καρκίνωμα του πνεύμονα (Squamous Cell Carcinoma)

Προεπεξεργασία δεδομένων πραγματοποιείται με την πρόσθεση γκαουσιανού θορύβου και την απαλοιφή του, με χρήση φίλτρου Median της python βιβλιοθήκης Scikit-learn. Επίσης, σαν μέρος της βελτιστοποίησης των μοντέλων, οι εικόνες των τομογραφιών, δέχονται διαφόρων ειδών αλλαγής μεγέθους (rescaling) και κανονικοποίησης (normalization).

3.2.2 Μοντέλα και Αρχιτεκτονικές

VGG CNN: Το βασικό μοντέλο που θα χρησιμοποιηθεί για τον εντοπισμό (ανίχνευση) του καρκίνου του πνεύμονα από εικόνες υπολογιστικών τομογραφιών (CT-Scans), είναι μία απλουστευμένη εκδοχή του VGG-16 που αναλύθηκε νωρίτερα στην εργασία.

Δομή μοντέλου (Model Structure)

- **Συνελικτικά μπλοκ (Convolutional blocks):** Κάθε συνελικτικό μπλοκ αποτελείται από δύο συνελικτικά στρώματα με φίλτρα 3×3 , ακολουθούμενα από ένα στρώμα ομαδικής κανονικοποίησης (Batch Normalization) και ένα στρώμα μέγιστης συγκέντρωσης (Max Pooling Layer). Αυτή η δομή επαναλαμβάνεται για πολλά μπλοκ (5) με αυξανόμενο αριθμό φίλτρων (16, 32, 64, 128, 256).
- **Πυκνά μπλοκ (Dense Blocks):** Κάθε πυκνό μπλοκ αποτελείται από ένα πυκνό στρώμα που ακολουθείται από Batch Normalization και Dropout για κανονικοποίηση. Τα πυκνά στρώματα μειώνουν προοδευτικά τον αριθμό των μονάδων, με στρώματα Dropout για τον μετριασμό της υπερπροσαρμογής (overfitting).
- **Στρώμα εξόδου (Output Layer):** Το τελικό στρώμα είναι ένα πυκνό στρώμα με συνάρτηση ενεργοποίησης Softmax, η οποία είναι τυπική για εργασίες ταξινόμησης. Ο αριθμός των μονάδων σε αυτό το στρώμα αντιστοιχεί στον αριθμό των κλάσεων (Class Counts).
- **Μεταγλώττιση και εκπαίδευση (Compilation and Training):** Το μοντέλο μεταγλωττίζεται με τον βελτιστοποιητή (optimizer) Adamax και εκπαιδεύεται χρησιμοποιώντας τη συνάρτηση απώλειας categorical cross-entropy, με πρόσθετες μετρικές όπως η ακρίβεια (accuracy) και η ανάκληση (recall).

Το μοντέλο μοιάζει περισσότερο με την αρχιτεκτονική VGG-16 με ορισμένες σύγχρονες βελτιώσεις:

Σύγκριση με VGG-16

- **Συνελικτικά μπλοκ:** Αυτό το μοντέλο χρησιμοποιεί πολλαπλά στρώματα συνελικτικής ανάλυσης με μικρά φίλτρα 3×3 , ακολουθούμενα από max pooling. Η χρήση δύο στρωμάτων Conv2D σε κάθε μπλοκ είναι ευθέως σύμφωνη με την προσέγγιση VGG-16, όπου χρησιμοποιούνται μπλοκ από 2 ή 3 στρώματα συνελίξεων.
- **Διπλασιασμός φίλτρων:** Όπως και το VGG-16, το μοντέλο διπλασιάζει τον αριθμό των φίλτρων σε κάθε διαδοχικό μπλοκ συνελικτικής ανάλυσης ($16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256$).

- **Πυκνά στρώματα:** Το μοντέλο περιλαμβάνει πολλαπλά πυκνά στρώματα μετά τα μπλοκ συνελίξεων, παρόμοια με τα πλήρως συνδεδεμένα στρώματα (Fully-Connected Layers) στο VGG-16.

Βελτιώσεις σε σχέση με το VGG-16

- **Κανονικοποίηση δέσμης:** Η ομαδοποίηση δέσμης εφαρμόζεται μετά τα συνελικτικά στρώματα, η οποία δεν περιλαμβανόταν στον αρχικό σχεδιασμό του VGG-16, αλλά αποτελεί κοινή προσθήκη στις σύγχρονες αρχιτεκτονικές για τη σταθεροποίηση της εκπαίδευσης.
- **Dropout:** Το Dropout περιλαμβάνεται στα πυκνά μπλοκ για κανονικοποίηση, το οποίο αποτελεί άλλη μια βελτίωση σε σχέση με το τυπικό VGG-16.
- **Μετρική IoU:** Η συμπερίληψη μιας προσαρμοσμένης μετρικής IoU είναι μια προσθήκη για συγκεκριμένες εργασίες.

Βασικές διαφορές από την VGG-16

Απλούστευση: Το παρεχόμενο μοντέλο χρησιμοποιεί λιγότερα φίλτρα στα πρώτα στρώματα και έναν απλούστερο συνολικό σχεδιασμό. Τα πλήρως συνδεδεμένα στρώματα στο VGG-16 έχουν πολύ περισσότερες μονάδες (4096 μονάδες) σε σύγκριση με αυτό το μοντέλο.

Dropout και κανονικοποίηση δέσμης: Η συμπερίληψή τους εδώ βελτιώνει την κανονικοποίηση του μοντέλου και την απόδοση.

Λόγοι επιλογής του μοντέλου: Αυτό το μοντέλο περιγράφεται ως μια αρχιτεκτονική που μοιάζει με την VGG με σύγχρονες βελτιώσεις, όπως η ομαδική κανονικοποίηση και το Dropout. Ακολουθεί τις βασικές αρχές των δικτύων VGG, δηλαδή τη στοίβαξη μικρών φίλτρων συνελίξεων και τον διπλασιασμό του αριθμού των φίλτρων σε κάθε μπλοκ, αλλά απλοποιεί τη σχεδίαση και προσθέτει τεχνικές κανονικοποίησης, καθιστώντας το πιο ανθεκτικό και ενδεχομένως πιο κατάλληλο για συγκεκριμένες εργασίες.

EfficientNetB3: Το μοντέλο που αναπτύχθηκε βασίζεται στην αρχιτεκτονική EfficientNetB3 (βλ. Εικόνα 3.2.2.1), ένα σύγχρονο συνελικτικό νευρωνικό δίκτυο (CNN), γνωστό για την ισορροπία μεταξύ αποτελεσματικότητας και ακρίβειας. Η δομή του μοντέλου είναι ένας υβριδικός σχεδιασμός που αξιοποιεί προ-εκπαιδευμένα στρώματα από το EfficientNetB3 και ενσωματώνει στρώματα, προσαρμοσμένα στη συγκεκριμένη εργασία ταξινόμησης. Ακολουθεί λεπτομερής περιγραφή της δομής του μοντέλου:

- **Βασικό μοντέλο (Basic Model): EfficientNetB3**

Προ-εκπαιδευμένο (Pre-Trained) Backbone: Η βάση του μοντέλου είναι η αρχιτεκτονική EfficientNetB3, η οποία έχει προ-εκπαιδευτεί στο σύνολο δεδομένων ImageNet. Αυτό το Backbone αποτελείται από μια σειρά επιπέδων συνελίξεως που έχουν σχεδιαστεί για την αποτελεσματική εξαγωγή

χαρακτηριστικών (feature extraction) από εικόνες με ταυτόχρονη ελαχιστοποίηση του αριθμού των παραμέτρων. Εξαιρώντας τα κορυφαία (πλήρως συνδεδεμένα) στρώματα (`include_top = False`), το μοντέλο διατηρεί μόνο τα στρώματα εξαγωγής χαρακτηριστικών.

Στρώμα εισόδου (Input Layer): Το μοντέλο έχει σχεδιαστεί για να δέχεται εικόνες εισόδου συγκεκριμένου μεγέθους (π.χ. 224x224x3 για εικόνες RGB), εξασφαλίζοντας τη συμβατότητα με την αρχιτεκτονική EfficientNetB3.

- **Προσαρμοσμένα επίπεδα για λεπτομερή ρύθμιση (Custom Layers for Fine-Tuning):**

Συγκέντρωση μέσου όρου (Average Pooling): Ένα επίπεδο `GlobalAveragePooling2D` εφαρμόζεται στην έξοδο του βασικού μοντέλου EfficientNetB3. Αυτό το στρώμα συγκεντρώνει τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών (feature maps) σε ένα ενιαίο διάνυσμα με μέσο όρο των τιμών σε κάθε κανάλι. Η λειτουργία αυτή μειώνει τον αριθμό των παραμέτρων και προετοιμάζει τους χάρτες χαρακτηριστικών για τα επόμενα πυκνά στρώματα.

Κανονικοποίηση δέσμης (Batch Normalization): Μετά το στρώμα συγκέντρωσης, εισάγεται ένα στρώμα `Batch Normalization`. Αυτό το στρώμα κανονικοποιεί τις ενεργοποιήσεις (activations), γεγονός που βοηθά στη σταθεροποίηση και την επιτάχυνση της διαδικασίας εκπαίδευσης μειώνοντας τις εσωτερικές μετατοπίσεις των συνδιακυμάνσεων (internal covariate shifts).

- **Πυκνά μπλοκ (Dense Blocks):**

Πρώτο πυκνό μπλοκ: Το πρώτο προσαρμοσμένο πυκνό μπλοκ αποτελείται από ένα πλήρως συνδεδεμένο στρώμα (FC Layer) με 128 μονάδες, ακολουθούμενο από `Batch Normalization` και ένα στρώμα εγκατάλειψης (Dropout Layer) με ποσοστό εγκατάλειψης 50%. Αυτό το μπλοκ έχει σχεδιαστεί για την εκμάθηση χαρακτηριστικών υψηλού επιπέδου, μειώνοντας παράλληλα τον κίνδυνο υπερπροσαρμογής μέσω της κανονικοποίησης Dropout.

Δεύτερο πυκνό μπλοκ: Το δεύτερο πυκνό μπλοκ έχει παρόμοια δομή, αλλά περιέχει μόνο 32 μονάδες και ποσοστό εγκατάλειψης 20%. Αυτό το μπλοκ τελειοποιεί περαιτέρω τα χαρακτηριστικά που μαθαίνονται και προετοιμάζει το δίκτυο για την τελική ταξινόμηση.

- **Στρώμα Εξόδου (Output Layer):**

Ταξινομητής Softmax (Softmax Classifier): Το τελευταίο στρώμα του μοντέλου είναι ένα πυκνό στρώμα με αριθμό μονάδων ίσο με τον αριθμό των κλάσεων στην εργασία ταξινόμησης. Αυτό το στρώμα χρησιμοποιεί μια συνάρτηση ενεργοποίησης `Softmax`, η οποία εξάγει μια κατανομή πιθανοτήτων σε όλες τις κλάσεις. Η κλάση με την υψηλότερη πιθανότητα επιλέγεται ως πρόβλεψη του μοντέλου.

- **Μεταγλώττιση και Βελτιστοποίηση (Compilation and Optimization):**

Βελτιστοποιητής: Το μοντέλο συντάσσεται με τη χρήση του βελτιστοποιητή `Adamax`, ενός αλγορίθμου βελτιστοποίησης προσαρμοστικού ρυθμού μάθησης που συνδυάζει τα πλεονεκτήματα τόσο του `AdaGrad` όσο και του `RMSProp`.

Συνάρτηση απώλειας: Η συνάρτηση απώλειας που χρησιμοποιείται είναι η κατηγορική διασταυρούμενη εντροπία, η οποία είναι κατάλληλη για εργασίες ταξινόμησης πολλαπλών κλάσεων. Μετράει τη διαφορά μεταξύ της προβλεπόμενης κατανομής πιθανοτήτων και της πραγματικής (one-hot encoded).

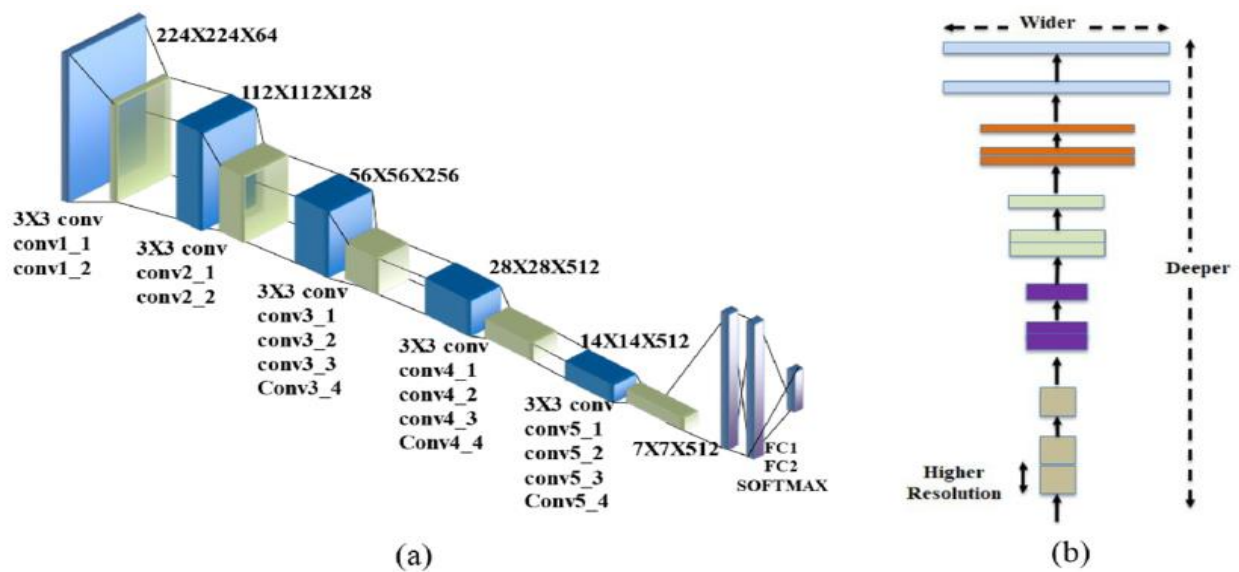
Μετρικές αξιολόγησης: Το μοντέλο αξιολογείται με τη χρήση μετρικών όπως η ακρίβεια, η ανάκληση και η IoU, παρέχοντας μια ολοκληρωμένη αξιολόγηση της απόδοσής του σε διάφορες διαστάσεις.

- **Ροή εργασίας εκπαίδευσης και αξιολόγησης μοντέλων (Training and Evaluation Workflow)**

Διαδικασία εκπαίδευσης: Το μοντέλο εκπαιδεύεται σε πολλαπλές εποχές (epochs), με παρακολούθηση της απόδοσης τόσο στα σύνολα δεδομένων εκπαίδευσης όσο και στα σύνολα δεδομένων επικύρωσης (Validation). Αυτό επιτρέπει προσαρμογές σε πραγματικό χρόνο και έγκαιρη διακοπή, εάν είναι απαραίτητο, για την αποφυγή υπερβολικής προσαρμογής.

Αξιολόγηση και πρόβλεψη: Μετά την εκπαίδευση, το μοντέλο αξιολογείται στο σύνολο δεδομένων δοκιμής (Test Data). Δημιουργούνται προβλέψεις και σχεδιάζεται ο πίνακας σύγχυσης (Confusion Matrix) για την οπτικοποίηση της απόδοσης ταξινόμησης του μοντέλου και της κατανομής των σφαλμάτων.

Λόγοι επιλογής του μοντέλου: Το μοντέλο συνδυάζει τα πλεονεκτήματα του Backbone EfficientNetB3, όπως είναι το Compound Scaling (Εικόνα 3.2.2.1), με προσαρμοσμένα πυκνά στρώματα για να δημιουργήσει μια στιβαρή και αποδοτική αρχιτεκτονική για την ταξινόμηση εικόνων. Η χρήση της συγκέντρωσης μέσου όρου, της κανονικοποίησης δέσμης (Batch Normalization) και των στρωμάτων εγκατάλειψης (Dropout Layers) διασφαλίζει ότι το μοντέλο είναι τόσο ισχυρό όσο και ανθεκτικό στην υπερπροσαρμογή.



Εικόνα 3.2.2.1: [63] a) Αρχιτεκτονική EfficientNetB3 b) Compound Scaling.

3.3 Μετρικές αξιολόγησης

Στη συγκεκριμένη μελέτη, χρησιμοποιήθηκαν διάφορες μετρικές αξιολόγησης για να αξιολογηθεί η απόδοση των μοντέλων, συμπεριλαμβανομένης της ακρίβειας, της ανάκλησης, της τομής πάνω από την ένωση (IoU) και της κατηγορικής απώλειας διασταυρούμενης εντροπίας.

- Ακρίβεια (Accuracy)

Πλεονεκτήματα

Απλότητα: Η ακρίβεια είναι απλή και εύκολη στην ερμηνεία, καθώς αντιπροσωπεύει το ποσοστό των σωστά ταξινομημένων περιπτώσεων επί του συνόλου των περιπτώσεων.

Συνολική απόδοση: Παρέχει γρήγορη επισκόπηση της απόδοσης του μοντέλου σε όλες τις κλάσεις.

Μειονεκτήματα:

Έλλειψη εξειδίκευσης: Η ακρίβεια δεν λαμβάνει υπόψη πόσο καλά το μοντέλο αποδίδει σε μεμονωμένες κλάσεις, ιδίως σε περιπτώσεις όπου ορισμένες κλάσεις είναι πιο δύσκολο να προβλεφθούν από άλλες.

- Ανάκληση (Recall)

Πλεονεκτήματα: Εστίαση στην ευαισθησία: Η ανάκληση, επίσης γνωστή ως ευαισθησία (sensitivity), μετρά την ικανότητα του μοντέλου να εντοπίζει σωστά όλες τις θετικές περιπτώσεις. Αυτό είναι ιδιαίτερα σημαντικό σε ιατρικές εφαρμογές όπου η παράλειψη μιας θετικής περίπτωσης (όπως μια καρκινική βλάβη) μπορεί να έχει σοβαρές συνέπειες.

Μειονεκτήματα: Δεν αποτελεί πλήρη μετρική: Η ανάκληση θα πρέπει ιδανικά να εξετάζεται μαζί με την ακρίβεια (ή μια μετρική που εξισορροπεί και τα δύο, όπως το σκορ F1) για να υπάρχει μια πλήρης εικόνα της απόδοσης του μοντέλου.

- Απώλεια κατηγορικής διασταυρούμενης εντροπίας (Categorical Cross-Entropy Loss)

Πλεονεκτήματα: Τιμωρεί την εσφαλμένη ταξινόμηση: Αυτή η συνάρτηση απώλειας είναι ευαίσθητη στις διαφορές μεταξύ της προβλεπόμενης κατανομής πιθανοτήτων και της πραγματικής κατανομής, πράγμα που σημαίνει ότι τιμωρεί το μοντέλο για σίγουρες αλλά εσφαλμένες προβλέψεις περισσότερο από ό,τι για αβέβαιες αλλά εσφαλμένες.

Αδυναμίες: Σε αντίθεση με την ακρίβεια ή την ανάκληση, η απώλεια κατηγορικής διασταυρούμενης εντροπίας δεν είναι τόσο εύκολα ερμηνεύσιμη ως προς το ποσοστό των σωστά ταξινομημένων περιπτώσεων. Απαιτεί μεγαλύτερη κατανόηση των πιθανοτήτων εξόδου του μοντέλου.

- Διατομή έναντι Ένωσης (IoU)

Δυνατά σημεία: Η IoU (Intersection Over Union) είναι η πιο αποτελεσματική μέθοδος για την αντιμετώπιση των προβλημάτων που προκύπτουν από την ανάλυση των δεδομένων. Η IoU είναι η γεωμετρική ακρίβεια και είναι ιδιαίτερα πολύτιμη σε εργασίες που περιλαμβάνουν χωρικές προβλέψεις, όπως η τμηματοποίηση εικόνων, όπου μετρά την επικάλυψη μεταξύ των προβλεπόμενων περιοχών και των περιοχών βασικής αλήθειας. Στην ιατρική απεικόνιση, η IoU παρέχει μια σαφή κατανόηση του πόσο καλά το μοντέλο περιγράφει τις περιοχές ενδιαφέροντος, όπως οι καρκινικοί

ιστοί. Σε αντίθεση με την ακρίβεια, η IoU παρέχει μια πιο λεπτομερή άποψη της απόδοσης (διακριτική ικανότητα), λαμβάνοντας υπόψη τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα σε σχέση με τα πραγματικά θετικά αποτελέσματα, κάτι που είναι κρίσιμο σε ιατρικές εφαρμογές όπου είναι απαραίτητος ο ακριβής εντοπισμός.

Αδυναμίες: Η IoU είναι υπολογιστικά πιο πολύπλοκη από την ακρίβεια ή την ανάκληση και απαιτεί ακριβείς υπολογισμούς της επικάλυψης της περιοχής.

Η σημασία της IoU: Η IoU ξεχωρίζει ως ιδιαίτερα σημαντική μετρική στο πλαίσιο της ανάλυσης ιατρικών εικόνων. Στην ανίχνευση καρκίνου, όπου ο ακριβής εντοπισμός και η οριοθέτηση του όγκου είναι ζωτικής σημασίας, η IoU παρέχει μια πιο λεπτομερή αξιολόγηση από μετρικές όπως η ακρίβεια ή η ανάκληση. Συγκεκριμένα:

- Εστίαση στον εντοπισμό (Localization)

Η έμφαση που δίνει η IoU στη χωρική επικάλυψη μεταξύ των προβλεπόμενων και των πραγματικών περιοχών διασφαλίζει ότι τα μοντέλα όχι μόνο αναγνωρίζουν την παρουσία καρκίνου αλλά και εντοπίζουν με ακρίβεια την έκτασή του εντός της εικόνας. Αυτό είναι κρίσιμο σε εφαρμογές όπως η κατάτμηση όγκων, όπου το μέγεθος και το σχήμα του όγκου μπορεί να επηρεάσει άμεσα τις αποφάσεις θεραπείας.

- Χειρισμός επικάλυψης κλάσεων

Σε σενάρια όπου διαφορετικοί τύποι ιστών ή στάδια καρκίνου έχουν επικαλυπτόμενα χαρακτηριστικά στις εικόνες, η IoU βοηθά στη διάκριση του πόσο καλά το μοντέλο μπορεί να διαχωρίσει αυτές τις κλάσεις με βάση τις χωρικές τους ιδιότητες, καθιστώντας την ανώτερη μετρική.

- Εξισορρόπηση της ακρίβειας και της ανάκλησης

Ενώ η ανάκληση επικεντρώνεται στη σύλληψη όλων των αληθώς θετικών αποτελεσμάτων, η IoU εξισορροπεί εγγενώς αυτό με την ακρίβεια του εντοπισμού. Μια υψηλή βαθμολογία IoU υποδεικνύει ότι το μοντέλο όχι μόνο ανιχνεύει τις περισσότερες θετικές περιπτώσεις αλλά και αποφεύγει τα ψευδώς θετικά αποτελέσματα, οδηγώντας σε πιο αξιόπιστες και κλινικά χρήσιμες προβλέψεις.

3.4 Βιβλιοθήκες και Χαρακτηριστικά υπολογιστικού συστήματος

1. Χρησιμοποιούμενες βιβλιοθήκες

- Tensorflow/Keras: Οι βιβλιοθήκες αυτές χρησιμοποιήθηκαν για τη δημιουργία, την εκπαίδευση και την αξιολόγηση του μοντέλου CNN (αρχιτεκτονική VGG).

Βασικές λειτουργίες: Επίπεδα μοντέλου, βελτιστοποιητές, συναρτήσεις απώλειας και callbacks εκπαίδευσης για την παρακολούθηση της απόδοσης σε όλες τις εποχές.

- NumPy: Απαραίτητη για αριθμητικές πράξεις και χειρισμούς συστοιχιών, ιδίως κατά την επεξεργασία εικόνων και το χειρισμό μεγάλων συνόλων δεδομένων. Το NumPy επέτρεψε αποδοτικούς υπολογισμούς κατά την προεπεξεργασία και τον μετασχηματισμό δεδομένων.

- OpenCV: Χρησιμοποιείται για εργασίες επεξεργασίας εικόνων, όπως η εφαρμογή θορύβου (Gaussian και Granular) και η αφαίρεση θορύβου μέσω φίλτρων.
- Matplotlib: Αυτή η βιβλιοθήκη οπτικοποίησης χρησιμοποιήθηκε για την απεικόνιση γραφημάτων ακρίβειας/απώλειας, πινάκων σύγχυσης και άλλων μετρήσεων απόδοσης.
- Scikit-learn: Αυτή η βιβλιοθήκη παρείχε βοηθητικά προγράμματα για τον υπολογισμό μετρήσεων όπως η ακρίβεια, και η ανάκληση. Χρησιμοποιήθηκε για τη δημιουργία των πινάκων σύγχυσης.
- Pandas Χρησιμοποιείται για την οργάνωση και την ανάλυση μετρήσεων απόδοσης σε διαφορετικές πειραματικές περιπτώσεις. Το Pandas διευκόλυνε τη σύγκριση διαφόρων συνδυασμών θορύβου και φίλτρων.

2. Χαρακτηριστικά υπολογιστικού συστήματος

- Επεξεργαστής (CPU): Τα πειράματα διεξήχθησαν σε επεξεργαστή πολλαπλών πυρήνων. Η CPU χειριζόταν τις γενικές εργασίες προεπεξεργασίας δεδομένων, τις λειτουργίες εισόδου/εξόδου αρχείων και τις εργασίες αρχικής φόρτωσης δεδομένων.
- Μονάδα επεξεργασίας γραφικών (GPU): Η εκπαίδευση του μοντέλου CNN επιταχύνθηκε με τη χρήση μιας GPU Nvidia GTX 1050ti. Οι δυνατότητες παράλληλης επεξεργασίας της GPU μείωσαν σημαντικά τον χρόνο εκπαίδευσης επιταχύνοντας τις λειτουργίες πινάκων και την αντίστροφη διάδοση.
- RAM: Το σύστημα ήταν εξοπλισμένο με μνήμη RAM 8 GB, επαρκής για τη φόρτωση και την επεξεργασία παρτίδων εικόνων αξονικής τομογραφίας στη μνήμη χωρίς συμφόρηση επιδόσεων κατά τη διάρκεια της εκπαίδευσης.

3. Περιβάλλον ανάπτυξης

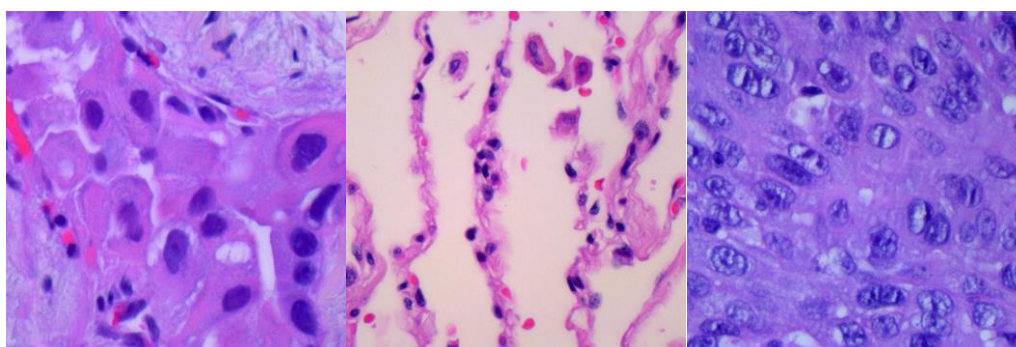
Η κωδικοποίηση και ο πειραματισμός πραγματοποιήθηκαν στο PyCharm. Αυτό το περιβάλλον επέτρεψε την εύκολη παρακολούθηση της απόδοσης του μοντέλου, την οπτικοποίηση των μετρήσεων και την αποσφαλμάτωση κατά τη διάρκεια της πειραματικής φάσης.

Κεφάλαιο 4: Αποτελέσματα Αξιολόγησης Πλαισίου

4.1 Σύγκριση VGG και EfficientNetB3 στα αρχικά δεδομένα (Raw Unprocessed Data):

Μοντέλο VGG

Σε αυτό το σενάριο, το μοντέλο VGG εφαρμόζεται στο αρχικό, αμετάβλητο σύνολο δεδομένων για τον καρκίνο του πνεύμονα. Ο στόχος είναι να καθοριστεί μια βασική απόδοση για την αρχιτεκτονική VGG χωρίς πρόσθετο θόρυβο ή φιλτράρισμα. Αυτό παρέχει εικόνα για το πόσο καλά το μοντέλο μπορεί να ταξινομήσει τους τρεις τύπους πνευμονικού ιστού (αδενοκαρκίνωμα, πλακώδες καρκίνωμα και καλοήθης ιστός) υπό βέλτιστες συνθήκες, όπου η ποιότητα των δεδομένων είναι υψηλή. Αναλύοντας μετρικές όπως η ακρίβεια, η ανάκληση και η απώλεια κατηγορικής διασταυρούμενης εντροπίας, μπορούμε να αξιολογήσουμε τα εγγενή πλεονεκτήματα και τους περιορισμούς του μοντέλου VGG όταν εργάζεται με καθαρές ιατρικές εικόνες.



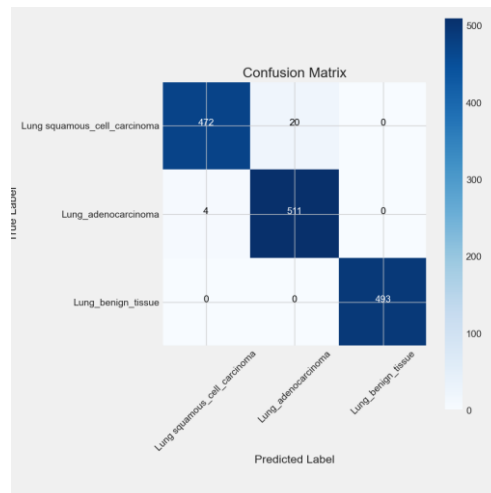
Εικόνα 4.1.1: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell

Πίνακας 4.1.1: Πίνακας απόδοσης VGG για 20 εποχές.

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.994	0.994	0.990	0.020
Validation	0.986	0.986	0.975	0.046
Test	0.984	0.984	0.971	0.045



Εικόνα 4.1.2: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell

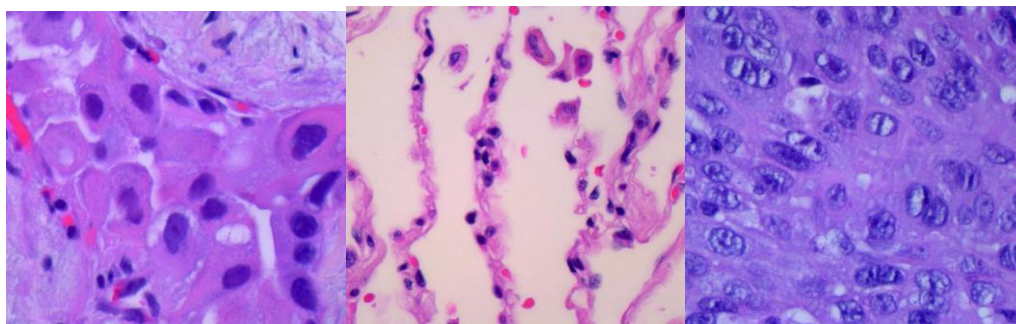


Εικόνα 4.1.3: Πίνακας Σύγχυσης VGG με Original Dataset

Από τις επιδόσεις ανά κατηγορία καρκίνου (καλοήθης, αδenoκαρκίνωμα, πλακώδες καρκίνωμα), παρατηρείται ότι το μοντέλο προέβλεψε με 100% ακρίβεια την κλάση καλοήθους ιστού, ενώ το μεγαλύτερο ποσοστό λαθών προέκυψε από λανθασμένη ταξινόμηση του Squamous Cell ως Adenocarcinoma. Προφανώς το γενικό ποσοστό ακριβούς ταξινόμησης (98%) είναι πολύ υψηλό, αλλά ακόμα σημαντικότερο είναι το γεγονός ότι δεν συνέβη διάγνωση υγιούς ατόμου ως ασθενούς.

Μοντέλο EfficientNetB3

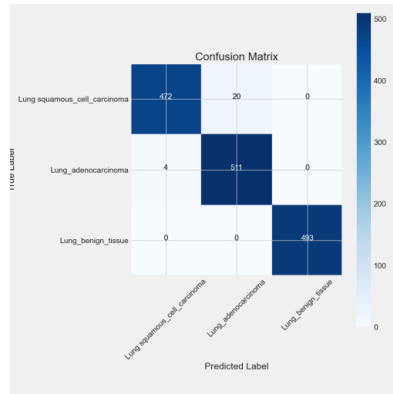
Σε αυτό το σενάριο, το μοντέλο EfficientNetB3 δοκιμάζεται στο αρχικό, αμετάβλητο σύνολο δεδομένων για τον καρκίνο του πνεύμονα για να παρέχει μια βασική αξιολόγηση των επιδόσεων. Το EfficientNetB3, γνωστό για την επεκτασιμότητα και την αποτελεσματικότητά του, αναμένεται να έχει καλές επιδόσεις υπό βέλτιστες συνθήκες με δεδομένα υψηλής ποιότητας. Αυτό το πείραμα αποσκοπεί στην αξιολόγηση της ακρίβειας ταξινόμησης και της ευρωστίας του σε σύγκριση με το μοντέλο VGG όταν δεν εφαρμόζεται θόρυβος ή φιλτράρισμα. Θα αναλυθούν βασικές μετρικές όπως η ακρίβεια, η ανάκληση, η IoU και η απώλεια διασταυρούμενης εντροπίας για να προσδιοριστεί πόσο καλά το EfficientNetB3 μπορεί να ταξινομήσει τύπους πνευμονικού ιστού σε ένα ιδανικό σενάριο.



Εικόνα 4.1.4: Original Dataset Samples: Adenocarcinoma - Benign Tissue - Squamous Cell

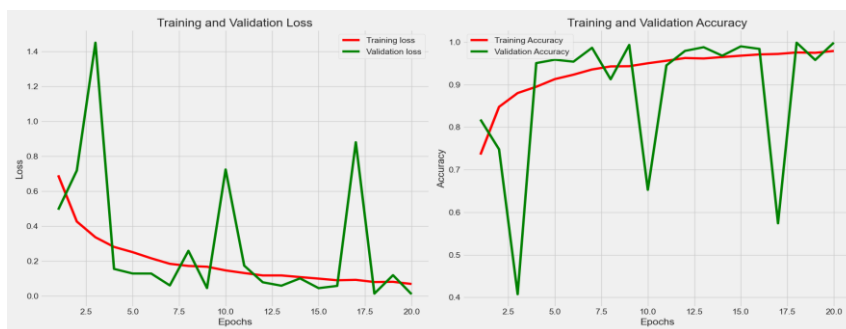
Πίνακας 4.1.2: Πίνακας απόδοσης EfficientNetB3 για 20 εποχές.

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.999	0.999	0.999	0.005
Validation	0.998	0.998	0.997	0.009
Test	0.998	0.998	0.997	0.009



Εικόνα 4.1.5: Πίνακας Σύγκρισης EfficientNetB3 με Original Dataset

Από τις επιδόσεις ανά κατηγορία καρκίνου (καλοήθης, αδenoκαρκίνωμα, πλακώδες καρκίνωμα), παρατηρείται ότι το μοντέλο προέβλεψε με 100% ακρίβεια την κλάση καλοήθους ιστού, ενώ το μεγαλύτερο ποσοστό λαθών προέκυψε ξανά από λανθασμένη ταξινόμηση του Squamous Cell ως Adenocarcinoma. Το ποσοστό ακριβούς ταξινόμησης (99%) είναι πολύ υψηλό, ξεπερνώντας το μοντέλο VGG.



Εικόνα 4.1.6: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές

- **Σύγκριση των VGG και EfficientNetB3**

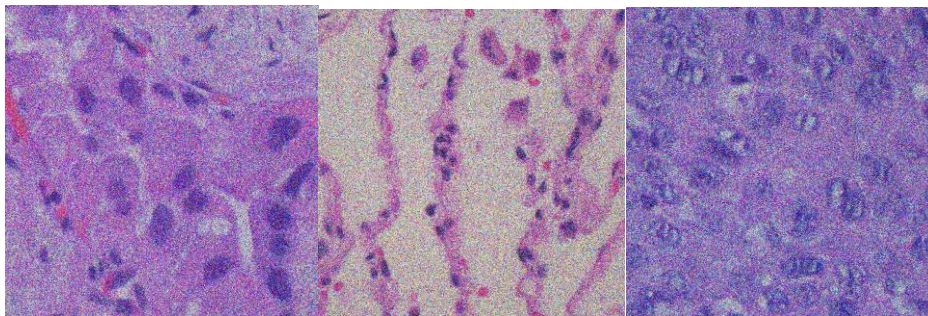
Συγκρίνοντας τα αποτελέσματα των βασικών μετρικών (ακρίβεια, ανάκληση, απώλεια, IoU) για τα δύο μοντέλα, φαίνεται πως το EfficientNetB3 ξεπερνά το VGG μοντέλο κατά μικρά ποσοστά σε όλα τα στατιστικά. Η μεγαλύτερη διαφορά στην αποδοτικότητα γίνεται εμφανής στην τιμή της Cross-Entropy Loss, όπου πλέον η απώλεια είναι σχεδόν μηδενική. Επίσης, παρατηρείται και στα δύο μοντέλα μεγαλύτερο ποσοστό misclassification μεταξύ αδenoκαρκινώματος (actual label) και πλακώδους ιστού (predicted label). Το πλεονέκτημα του pretrained μοντέλου EfficientNetB3 βασίζεται κυρίως στο compound scaling, όπου το μοντέλο αλλάζει το μέγεθος, το βάθος και την ανάλυση της εικόνας με συγκεκριμένη αναλογία για την βελτιστοποίηση της ταξινόμησης.

4.2 Αποτελέσματα με προσθήκη γκαουσιανού (Gaussian) και κοκκώδη (Granular) θορύβου

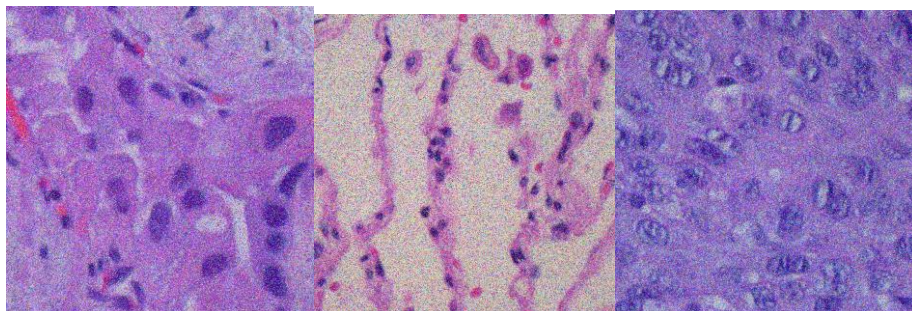
Στόχος είναι να συγκριθεί ο τρόπος με τον οποίο το μοντέλο VGG χειρίζεται κάθε τύπο παραμόρφωσης. Εφαρμόζοντας αυτούς τους τύπους θορύβου ξεχωριστά, στοχεύουμε να αξιολογήσουμε πώς αλλάζει η απόδοση του μοντέλου VGG με διαφορετικά είδη παραμορφώσεων της εικόνας. Συγκεκριμένα, θα αξιολογήσουμε μετρικές όπως η ακρίβεια, η ανάκληση, η IoU και η απώλεια κατηγορικής διασταυρούμενης εντροπίας για να καθορίσουμε πώς κάθε τύπος θορύβου επηρεάζει τις δυνατότητες ταξινόμησης του μοντέλου και την εξαγωγή χαρακτηριστικών.

Οπτική Σύγκριση

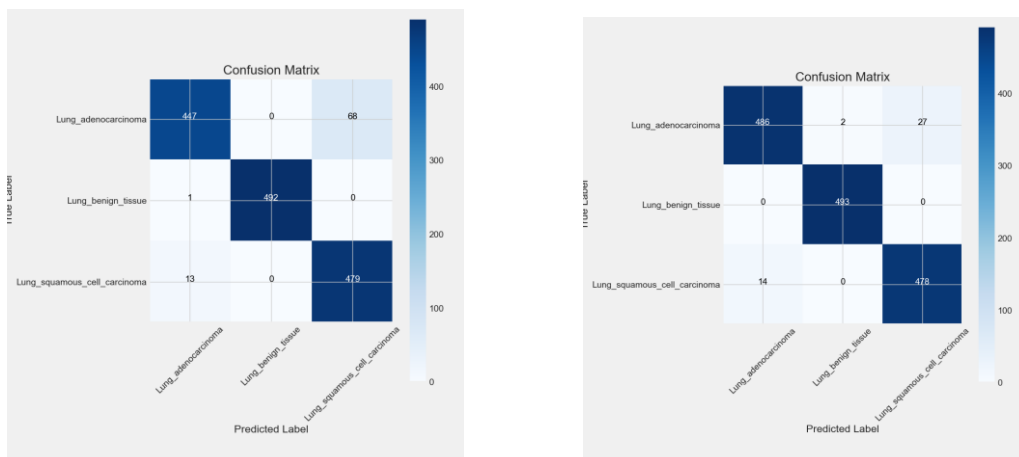
- Μοντέλο VGG με προσθήκη Gaussian και Granular θορύβου (Noise Scale = 0.3)



Εικόνα 4.2.1: Gaussian Noised Samples: Adenocarcinoma - Benign Tissue - Squamous Cell



Εικόνα 4.2.2: Granular Noised Samples: Adenocarcinoma - Benign Tissue - Squamous Cell



Εικόνα 4.2.3: Πίνακας Σύγκρισης Gaussian – Granular θορύβου

Αξιολόγηση της απόδοσης CNN κατά την ανίχνευση και κατηγοριοποίηση αντικειμένων



Εικόνα 4.2.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές για Gaussian noise



Εικόνα 4.2.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές για Granular noise

Σύγκριση Απόδοσης

Πίνακας 4.2.1: Αποτελέσματα απόδοσης για 20 εποχές Gaussian Noise.

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.994	0.994	0.989	0.016
Validation	0.949	0.949	0.907	0.164
Test	0.945	0.945	0.898	0.164

Πίνακας 4.2.2: Αποτελέσματα απόδοσης για 20 εποχές Granular Noise.

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.990	0.990	0.984	0.041
Validation	0.970	0.970	0.947	0.080
Test	0.971	0.970	0.949	0.081

Συγολιασμός

Η υψηλότερη απόδοση του μοντέλου VGG στον κοκκώδη θόρυβο σε σύγκριση με τον γκαουσιανό θόρυβο μπορεί να εξηγηθεί από τα διαφορετικά χαρακτηριστικά αυτών των δύο τύπων θορύβου και τον τρόπο με τον οποίο αλληλοεπιδρούν με τη δομή των εικόνων αξονικής τομογραφίας και το μοντέλο βαθιάς μάθησης.

1. Τύπος θορύβου

• Γκαουσιανός θόρυβος

Αυτός ο τύπος θορύβου είναι προσθετικός και κατανέμεται σε όλα τα εικονοστοιχεία της εικόνας σύμφωνα με μια γκαουσιανή κατανομή (μέση τιμή 0, διακύμανση που εξαρτάται από την κλίμακα). Αυτή η τυχαιότητα επηρεάζει ομοιόμορφα ολόκληρη την εικόνα, διαταράσσοντας τις ακμές και τις λεπτότερες λεπτομέρειες των μοτίβων του πνευμονικού ιστού που είναι ζωτικής σημασίας για τη διάκριση μεταξύ καρκινικών και καλοηθών περιοχών.

Επίδραση στο VGG: Ο γκαουσιανός θόρυβος τείνει να θολώνει την υφή και να παραμορφώνει σημαντικά δομικά χαρακτηριστικά στις αξονικές τομογραφίες, τα οποία είναι βασικά για το μοντέλο VGG προκειμένου να κάνει ακριβείς προβλέψεις. Δεδομένου ότι το VGG έχει σχεδιαστεί για να καταγράφει ιεραρχικά χαρακτηριστικά μέσω των συνελκτικών του στρωμάτων, αυτή η θόλωση δυσχεραίνει την ικανότητα του μοντέλου να ανιχνεύει λεπτής κλίμακας χωρικές λεπτομέρειες.

• Κοκκώδης θόρυβος

Ο κοκκώδης θόρυβος, από την άλλη πλευρά, τείνει να δημιουργεί συστάδες τοπικού θορύβου, που συχνά εμφανίζονται ως φαινόμενα «αλατιού και πιπεριού» ή αποσπασματικές παραμορφώσεις.

Επίδραση στο VGG: Ενώ ο κοκκώδης θόρυβος μπορεί να είναι οπτικά ενοχλητικός, συνήθως επηρεάζει μόνο συγκεκριμένες περιοχές της εικόνας και αφήνει άλλες περιοχές σχετικά ανέπαφες. Το μοντέλο VGG, με τη βαθιά αρχιτεκτονική του και τα στρώματα συγκέντρωσης, είναι ικανό να φιλτράρει μέρος αυτού του τοπικού θορύβου εστιάζοντας σε μεγαλύτερα, πιο ανθεκτικά χαρακτηριστικά. Επιπλέον, τα στρώματα συνέλιξης μπορούν να αγνοήσουν μικρότερα, απομονωμένα σημεία θορύβου πιο εύκολα από ό,τι μπορούν να αντιμετωπίσουν εκτεταμένο κατανεμημένο θόρυβο.

2. Ανθεκτικότητα των μοντέλων βαθιάς μάθησης στο θόρυβο

• Τοπική έναντι ολικής διαταραχής

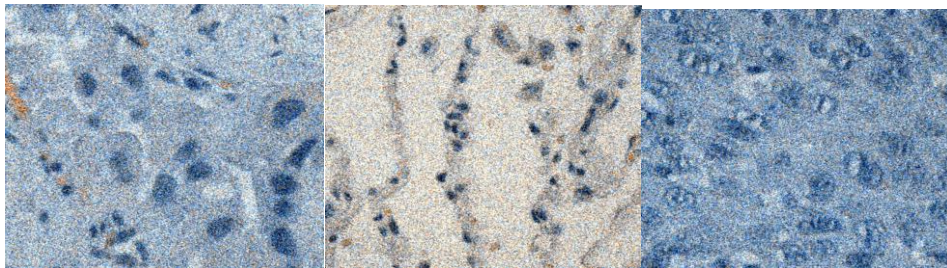
Τα μοντέλα βαθιάς μάθησης, ειδικά εκείνα με πολλά επίπεδα όπως το VGG, μπορούν να διαχειριστούν κάποιο επίπεδο τοπικού θορύβου, επειδή το βάθος τους τους επιτρέπει να ανακατασκευάζουν ουσιαστικά μοτίβα ακόμη και παρουσία διαταραχών. Ο κοκκώδης θόρυβος, ο οποίος είναι πιο τοπικός, επιτρέπει στο μοντέλο να «προσπεράσει» τις θορυβώδεις περιοχές και να εξακολουθεί να συλλαμβάνει ολικά μοτίβα. Ο γκαουσιανός θόρυβος, που είναι πιο ομοιόμορφα κατανεμημένος, διαταράσσει την ικανότητα του μοντέλου να ανακτά σαφή σήματα από τα δεδομένα.

4.3 Αποτελέσματα με απαλοιφή Gaussian θορύβου μέσω Median και Moving Average φίλτρου

Σε αυτό το σενάριο, αρχικά προστίθεται γκαουσιανός θόρυβος στο σύνολο δεδομένων για τον καρκίνο του πνεύμονα και εφαρμόζονται δύο διαφορετικές τεχνικές φιλτραρίσματος - το φιλτράρισμα του μέσου όρου και το φιλτράρισμα του κινητού μέσου όρου - για την αφαίρεση του θορύβου. Ο στόχος είναι να αξιολογηθεί πόσο αποτελεσματικά κάθε μέθοδος φιλτραρίσματος μπορεί να αποκαταστήσει την ποιότητα της εικόνας και να βελτιώσει την απόδοση του μοντέλου VGG στην ταξινόμηση των τύπων ιστού του πνεύμονα. Το πείραμα συγκρίνει την αποτελεσματικότητα αυτών των δύο τεχνικών φιλτραρίσματος στην αποκατάσταση της ποιότητας της εικόνας και στη βελτίωση της απόδοσης του μοντέλου VGG.

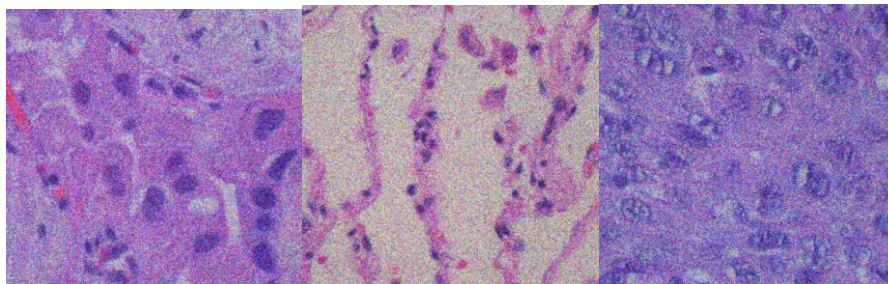
Οπτική Σύγκριση

- Μοντέλο VGG με προσθήκη Gaussian θορύβου και εφαρμογή Median φίλτρου

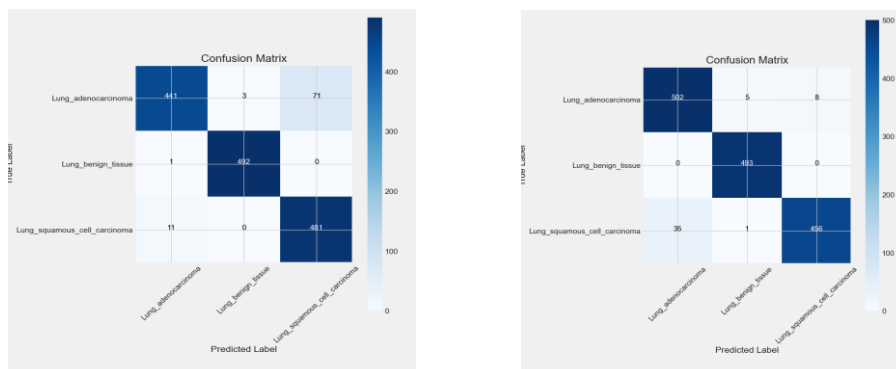


Εικόνα 4.3.1: Median Filtered Gaussian Noised Samples
Adenocarcinoma - Benign Tissue - Squamous Cell

- Μοντέλο VGG με προσθήκη Gaussian θορύβου και εφαρμογή Moving Average φίλτρου



Εικόνα 4.3.2: Moving Average Filtered Gaussian Noised Samples
Adenocarcinoma - Benign Tissue - Squamous Cell



Εικόνα 4.3.3: Πίνακας Σύγκρισης απαλοιφής Gaussian θορύβου Median - Moving Average Filter

Αξιολόγηση της απόδοσης CNN κατά την ανίχνευση και κατηγοριοποίηση αντικειμένων



Εικόνα 4.3.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Median Filter



Εικόνα 4.3.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter

Σύγκριση Απόδοσης

Πίνακας 4.3.1: Αποτελέσματα απόδοσης για 20 εποχές Median Filter σε Gaussian noise

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.974	0.974	0.951	0.063
Validation	0.941	0.941	0.891	0.169
Test	0.942	0.942	0.893	0.170

Πίνακας 4.3.2: Αποτελέσματα απόδοσης για 20 εποχές Moving Average Filter σε Gaussian noise

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.986	0.986	0.976	0.041
Validation	0.975	0.975	0.957	0.077
Test	0.967	0.967	0.944	0.089

Σχολιασμός

Η ανώτερη απόδοση του φίλτρου κινητού μέσου έναντι του φίλτρου διάμεσου στον Γκαουσιανό θόρυβο μπορεί να εξηγηθεί από τη φύση του ίδιου του Γκαουσιανού θορύβου και τον τρόπο με τον οποίο τα δύο φίλτρα λειτουργούν διαφορετικά όσον αφορά την αφαίρεση αυτού του θορύβου και τη διατήρηση σημαντικών χαρακτηριστικών της εικόνας.

- **Φίλτρο κινητού μέσου όρου (Moving Average Filter)**

Αυτό το φίλτρο εξομαλύνει μια εικόνα αντικαθιστώντας την τιμή κάθε εικονοστοιχείου με το μέσο όρο των τιμών των εικονοστοιχείων που το περιβάλλουν εντός καθορισμένου παραθύρου. Το φίλτρο υποθέτει ότι τα γειτονικά εικονοστοιχεία είναι πιθανό να έχουν παρόμοιες τιμές και ο μέσος όρος βοηθά στη μείωση της επίδρασης κάθε μεμονωμένου θορυβώδους εικονοστοιχείου με την ανάμειξή του με τα γειτονικά .

Επίδραση στον Γκαουσιανό θόρυβο: Δεδομένου ότι ο γκαουσιανός θόρυβος εξαπλώνεται στην εικόνα με σχετικά ομαλό και συνεχή τρόπο, το φίλτρο κινούμενου μέσου όρου μπορεί να μειώσει αποτελεσματικά τον θόρυβο με τον υπολογισμό του μέσου όρου των διακυμάνσεων της έντασης των εικονοστοιχείων. Το φίλτρο δεν αντιμετωπίζει τον θόρυβο ως ακραίες τιμές, αλλά μειώνει τη συνολική του ένταση, οδηγώντας σε μια πιο ομαλή, πιο φυσική εικόνα.

- **Φίλτρο διάμεσου**

Αυτό το φίλτρο αντικαθιστά την τιμή κάθε εικονοστοιχείου με τη διάμεση τιμή των γύρω εικονοστοιχείων εντός ενός καθορισμένου παραθύρου. Το φίλτρο διάμεσου είναι ιδιαίτερα αποτελεσματικό στην αφαίρεση του παλμικού θορύβου, όπως ο κοκκώδης θόρυβος, επειδή εστιάζει στην εξάλειψη των ακραίων τιμών, ενώ αφήνει την υπόλοιπη εικόνα σε μεγάλο βαθμό ανέπαφη.

Επίδραση στον Γκαουσιανό θόρυβο: Δεν υπάρχουν απότομες, απομονωμένες ακραίες τιμές στην εικόνα. Το διάμεσο φίλτρο μπορεί να μην είναι τόσο αποτελεσματικό στην αντιμετώπιση του γκαουσιανού θορύβου, επειδή έχει σχεδιαστεί για να χειρίζεται διακριτές, ακραίες τιμές θορύβου και όχι συνεχείς, μικρές αποκλίσεις. Ως αποτέλεσμα, ενδέχεται να μην εξομαλύνει την εικόνα τόσο αποτελεσματικά και να οδηγήσει σε λιγότερο συνεπή μείωση του θορύβου σε σύγκριση με το φίλτρο κινητού μέσου όρου.

Ο γκαουσιανός θόρυβος είναι μια συνεχής κατανομή: Το φίλτρο κινητού μέσου όρου είναι εγγενώς κατάλληλο για τη μείωση του συνεχούς θορύβου όπως ο γκαουσιανός θόρυβος. Με τον μέσο όρο των τιμών των εικονοστοιχείων, μειώνει τη διακύμανση που προκαλεί ο θόρυβος, με αποτέλεσμα μια πιο ομαλή εικόνα. Το φίλτρο συμβάλλει στην μείωση της επίδρασης του γκαουσιανού θορύβου, διατηρώντας παράλληλα τη συνολική δομή της εικόνας.

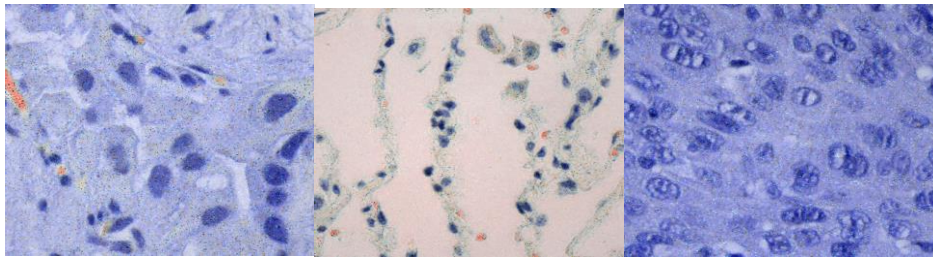
Επίδραση εξομάλυνσης του φίλτρου κινητού μέσου όρου: Το φίλτρο κινητού μέσου όρου εξομαλύνει τις διακυμάνσεις που προκαλούνται από το θόρυβο πιο αποτελεσματικά από το φίλτρο διάμεσου. Ενώ το φίλτρο διάμεσου απορρίπτει τις ακραίες τιμές, δεν χειρίζεται καλά τις μικρές, συνεχείς διακυμάνσεις, που είναι η ουσία του γκαουσιανού θορύβου. Το φίλτρο κινητού μέσου όρου, αντίθετα, μειώνει αυτές τις μικρές διακυμάνσεις και δημιουργεί μια πιο ομοιόμορφη εικόνα.

4.4 Αποτελέσματα με απαλοιφή Granular θορύβου μέσω Median και Moving Average φίλτρου

Σε αυτό το σενάριο, προστίθεται κοκκώδης θόρυβος στο σύνολο δεδομένων και εφαρμόζονται δύο τεχνικές απομάκρυνσης του θορύβου, το φιλτράρισμα του μέσου όρου και το φιλτράρισμα του κινητού μέσου όρου, για να βελτιωθεί η ποιότητα της εικόνας. Στόχος είναι να αξιολογηθεί πόσο καλά κάθε μέθοδος φιλτραρίσματος μετριάζει την επίδραση του κοκκώδους θορύβου και βελτιώνει την ικανότητα του μοντέλου VGG να ταξινομεί τύπους πνευμονικού ιστού. Το πείραμα συγκρίνει την απόδοση του VGG μετά την εφαρμογή και των δύο φίλτρων, με έμφαση σε μετρήσεις όπως η ακρίβεια, η ανάκληση και η IoU. Αυτή η σύγκριση θα βοηθήσει να προσδιοριστεί ποια μέθοδος φιλτραρίσματος χειρίζεται καλύτερα τον κοκκώδη θόρυβο και βοηθά στη βελτίωση των δυνατοτήτων ταξινόμησης του μοντέλου.

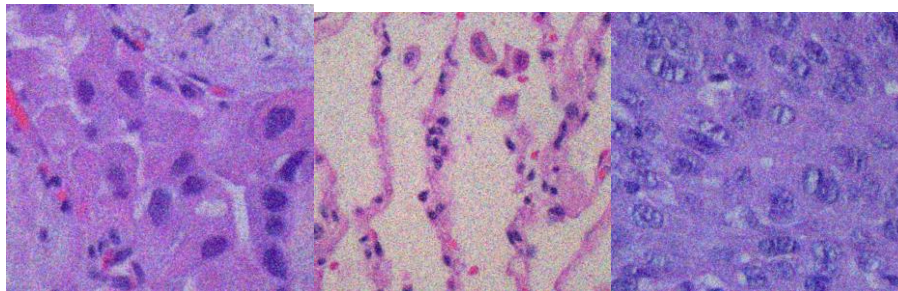
Οπτική Σύγκριση

- Μοντέλο VGG με προσθήκη Granular θορύβου και εφαρμογή Median φίλτρου:

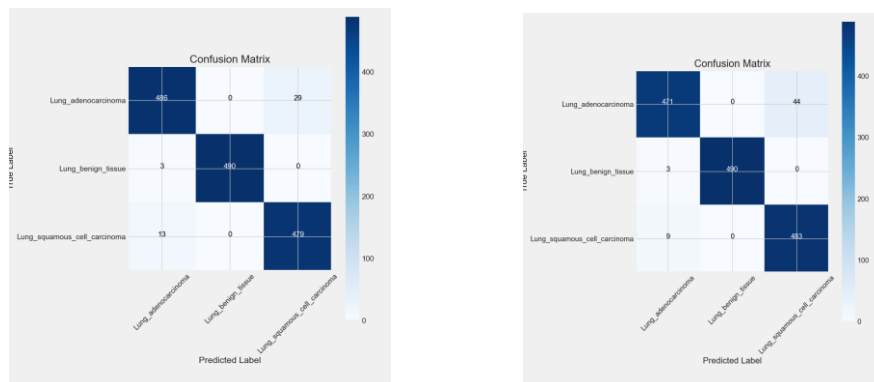


Εικόνα 4.4.1: Median Filtered Granular Noised Samples Adenocarcinoma - Benign Tissue - Squamous Cell

- Μοντέλο VGG με προσθήκη Gaussian θορύβου και εφαρμογή Moving Average φίλτρου

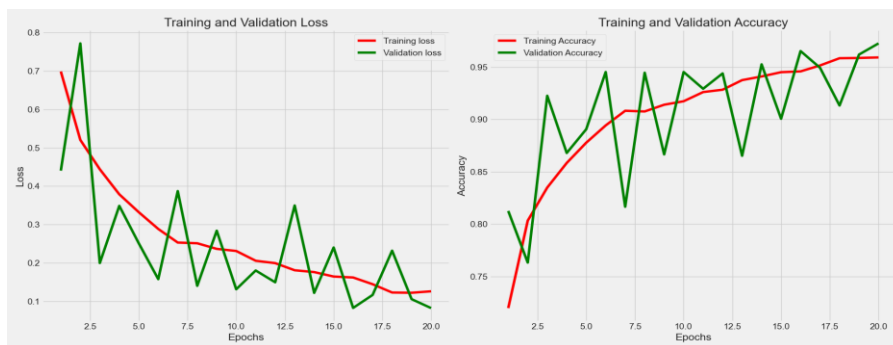


Εικόνα 4.4.2: Moving Average Filtered Granular Noised Samples Adenocarcinoma - Benign Tissue - Squamous Cell

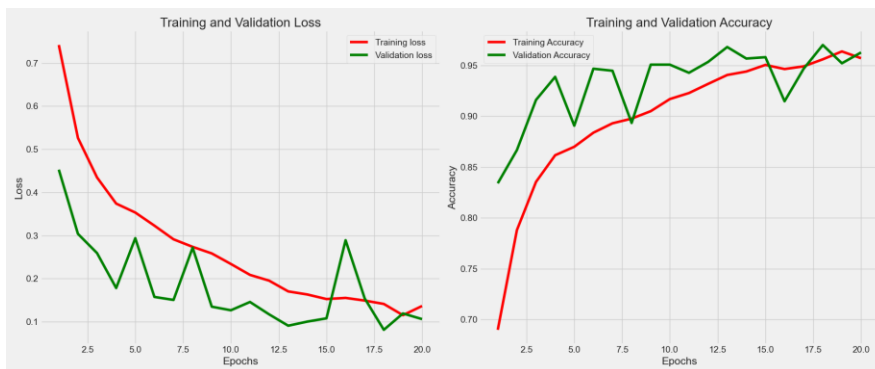


Εικόνα 4.4.3: Πίνακας Σύγκρισης απαλοιφής Granular θορύβου Median - Moving Average Filtered

Αξιολόγηση της απόδοσης CNN κατά την ανίχνευση και κατηγοριοποίηση αντικειμένων



Εικόνα 4.4.4: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Median Filter



Εικόνα 4.4.5: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter

Σύγκριση Απόδοσης

Πίνακας 4.4.1: Αποτελέσματα για 20 εποχές Median Filter σε Granular noise:

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.991	0.991	0.985	0.039
Validation	0.972	0.972	0.952	0.082
Test	0.970	0.970	0.947	0.076

Πίνακας 4.4.2: Αποτελέσματα για 20 εποχές Moving Average Filter σε Granular noise:

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.982	0.982	0.969	0.066
Validation	0.962	0.962	0.935	0.010
Test	0.962	0.962	0.935	0.114

Σχολιασμός

Η ανώτερη αποτελεσματικότητα του φίλτρου διάμεσου στον κοκκώδη θόρυβο σε σύγκριση με το φίλτρο κινητού μέσου όρου μπορεί να αποδοθεί στις θεμελιώδεις διαφορές στον τρόπο λειτουργίας των δύο φίλτρων και στον τρόπο με τον οποίο συμπεριφέρεται ο κοκκώδης θόρυβος στις εικόνες.

- Τύπος του κοκκώδους θορύβου (θόρυβος αλατιού και πιπεριού)
- Κοκκώδης θόρυβος (Salt-and-Pepper Noise):

Αυτός ο τύπος θορύβου εμφανίζεται ως τυχαία, υψηλής έντασης (αλάτι) ή χαμηλής έντασης (πιπέρι) εικονοστοιχεία διάσπαρτα σε μια εικόνα. Σε αντίθεση με τον γκαουσιανό θόρυβο, ο κοκκώδης θόρυβος είναι διακριτός και εκδηλώνεται ως αιχμηρές, απομονωμένες ακραίες τιμές στην εικόνα, διαταράσσοντας μόνο μερικά εικονοστοιχεία αντί να επηρεάζει ομοιόμορφα ολόκληρη την εικόνα.

- Ο θόρυβος εμφανίζεται συχνά ως μαύρα ή λευκά στίγματα (υψηλή αντίθεση) έναντι του φόντου της εικόνας, καθιστώντας τον πολύ διακριτό αλλά εντοπισμένο.
- Μηχανισμός Λειτουργίας Φίλτρων:
- **Φίλτρο διάμεσου:**

Αυτό το φίλτρο λειτουργεί αντικαθιστώντας την τιμή κάθε εικονοστοιχείου με τη διάμεση τιμή των εικονοστοιχείων μέσα σε ένα καθορισμένο παράθυρο (γειτονιά) γύρω από αυτό. Το διάμεσο φίλτρο είναι ειδικά σχεδιασμένο για να εξαλείφει τις ακραίες τιμές χωρίς να μεταβάλλει σημαντικά την υπόλοιπη εικόνα. Δεδομένου ότι ο κοκκώδης θόρυβος αποτελείται από αιχμηρές, ακραίες τιμές, το φίλτρο διάμεσου προσδιορίζει αποτελεσματικά αυτά τα θορυβώδη εικονοστοιχεία και τα αντικαθιστά με πιο αντιπροσωπευτικές τιμές από τα γειτονικά τους.

Επίδραση στον κοκκώδη θόρυβο: Το διάμεσο φίλτρο είναι ιδιαίτερα αποτελεσματικό στην αφαίρεση του θορύβου αλατιού και πιπεριού, επειδή στοχεύει ειδικά στις ακραίες τιμές. Διατηρεί τις ακμές και τις λεπτές λεπτομέρειες, ενώ εξαλείφει τα απομονωμένα θορυβώδη εικονοστοιχεία, παράγοντας μια καθαρότερη εικόνα που διατηρεί μεγάλο μέρος της αρχικής της δομής.

- **Φίλτρο κινητού μέσου όρου**

Το φίλτρο κινούμενου μέσου όρου εξομαλύνει μια εικόνα αντικαθιστώντας κάθε εικονοστοιχείο με το μέσο όρο των γειτονικών του εικονοστοιχείων. Ενώ αυτό μπορεί να μειώσει το θόρυβο, το κάνει με θόλωση της εικόνας. Το φίλτρο μέσου όρου αντιμετωπίζει εξίσου όλες τις μεταβολές των εικονοστοιχείων, συμπεριλαμβανομένου του κοκκώδους θορύβου, και τις υπολογίζει κατά μέσο όρο.

Επίδραση στον κοκκώδη θόρυβο: Δεδομένου ότι ο κοκκώδης θόρυβος είναι έντονος και εντοπισμένος, το φίλτρο κινητού μέσου δεν στοχεύει σε αυτά τα θορυβώδη εικονοστοιχεία τόσο αποτελεσματικά όσο το φίλτρο διάμεσου. Αντί να αφαιρεί τις ακραίες τιμές, θολώνει το θόρυβο μαζί με τα γύρω εικονοστοιχεία, οδηγώντας σε απώλεια λεπτομέρειας και λιγότερο αποτελεσματική αφαίρεση του θορύβου. Αυτό έχει ως αποτέλεσμα μια λιγότερο ευκρινή εικόνα, όπου ο κοκκώδης θόρυβος μειώνεται κάπως αλλά δεν εξαλείφεται

- Διατήρηση άκρων (edge preservation):
- Φίλτρο διάμεσου:

Διατήρηση αιχμηρών άκρων και σημαντικών χαρακτηριστικών, ζωτικής σημασίας για τη διάκριση μεταξύ διαφορετικών τύπων ιστών σε αξονικές τομογραφίες.

- Φίλτρο κινητού μέσου:

Θολώνει τις ακμές και τις λεπτομέρειες, μειώνοντας τη σαφήνεια της εικόνας και ενδεχομένως εμποδίζοντας την απόδοση του μοντέλου.

- Τοπική αφαίρεση θορύβου έναντι σφαιρικής εξομάλυνσης:
- Φίλτρο διάμεσου:

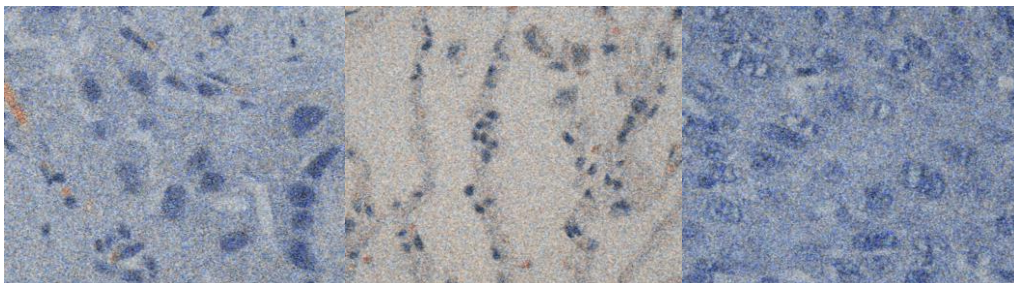
Υπερέχει στην αφαίρεση εντοπισμένου, απομονωμένου κοκκώδους θορύβου.

- Φίλτρο κινούμενου μέσου όρου:

Εφαρμόζει σφαιρική εξομάλυνση, η οποία είναι λιγότερο αποτελεσματική για απομονωμένο θόρυβο όπως ο κοκκώδης θόρυβος.

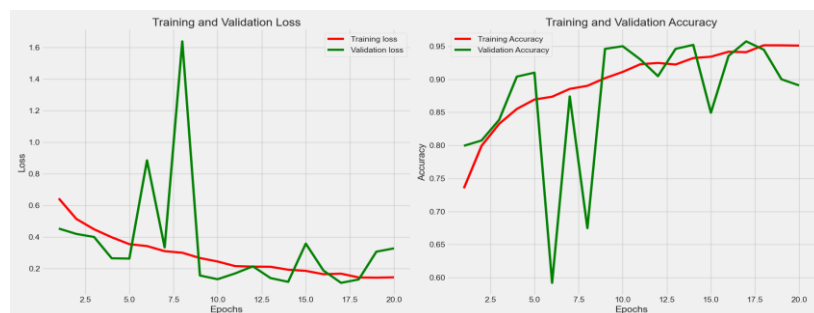
4.5 Αποτελέσματα με διαδοχική απαλοιφή Gaussian θορύβου μέσω Moving Average φίλτρου και Granular θορύβου μέσω Median Φίλτρου

Σε αυτό το σενάριο, τόσο το φιλτράρισμα της διάμεσης όσο και το φιλτράρισμα του κινητού μέσου εφαρμόζονται διαδοχικά για την αφαίρεση του κοκκώδους θορύβου από το σύνολο δεδομένων για τον καρκίνο του πνεύμονα. Ο στόχος είναι να διερευνηθεί κατά πόσον ο συνδυασμός αυτών των δύο μεθόδων φιλτραρίσματος μπορεί να ενισχύσει την αποτελεσματικότητα της αφαίρεσης του θορύβου και να βελτιώσει την απόδοση του μοντέλου VGG πέρα από αυτό που μπορεί να επιτευχθεί με ένα μόνο φίλτρο. Εφαρμόζοντας αυτά τα δύο φίλτρα διαδοχικά, το πείραμα επιδιώκει να καθορίσει εάν το διπλό φιλτράρισμα βελτιώνει την απόδοση ταξινόμησης του μοντέλου VGG. Θα αναλυθούν βασικές μετρικές αξιολόγησης, όπως η ακρίβεια, η ανάκληση και η IoU, για να εκτιμηθεί η αποτελεσματικότητα αυτής της προσέγγισης σε σύγκριση με τη χρήση κάθε φίλτρου ξεχωριστά.

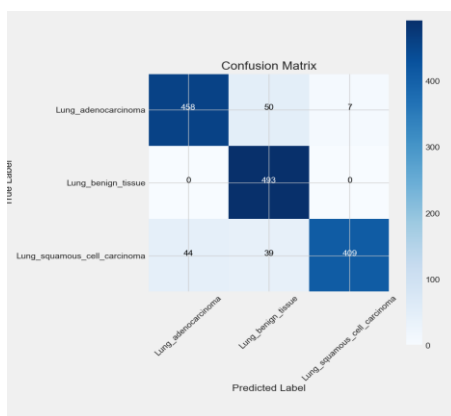


Εικόνα 4.5.1: Διαδοχική απαλοιφή γκαουσιανού θορύβου με φίλτρο κινητού μέσου όρου - απαλοιφή κοκκώδους θορύβου με φίλτρο διαμέσου.

Adenocarcinoma - Benign Tissue - Squamous Cell



Εικόνα 4.5.2: Διάγραμμα απώλειας και ακρίβειας σε σχέση με τις εποχές Moving Average Filter



Εικόνα 4.5.3: Πίνακας Σύγκρισης απαλοιφής Granular θορύβου Median - Moving Average Filtered

Σύγκριση Απόδοσης

Πίνακας 4.5.1: Αποτελέσματα για 20 εποχές Double Filter:

Data / Metric	Accuracy	Recall	IoU	Categorical Cross-Entropy Loss
Train	0.920	0.919	0.866	0.263
Validation	0.890	0.890	0.820	0.327
Test	0.906	0.906	0.842	0.309

Κατά την εφαρμογή διπλού φίλτρου (πρώτα προσθέτοντας γκαουσιανό θόρυβο και αφαιρώντας τον με το φίλτρο κινητού μέσου όρου και στη συνέχεια εφαρμόζοντας κοκκώδη θόρυβο και αφαιρώντας τον με το φίλτρο διάμεσου), η απόδοση του μοντέλου μειώνεται σημαντικά. Αυτό μπορεί να εξηγηθεί από διάφορους βασικούς παράγοντες:

- **Διαδοχική επίδραση του θορύβου και του φιλτραρίσματος**

Όταν εισάγεται γκαουσιανός θόρυβος και στη συνέχεια αφαιρείται με το φίλτρο κινητού μέσου όρου, η διαδικασία φιλτραρίσματος εξομαλύνει την εικόνα, μειώνοντας έτσι τον γκαουσιανό θόρυβο αλλά εισάγοντας επίσης θόλωση. Αυτό οδηγεί σε απώλεια σημαντικών λεπτών λεπτομερειών και ακμών που είναι κρίσιμες για την ικανότητα του μοντέλου να ταξινομεί με ακρίβεια τις εικόνες.

Μετά την εξομάλυνση της εικόνας, η εφαρμογή κοκκώδους θορύβου (θόρυβος αλατιού και πιπεριού) σε μια ήδη υποβαθμισμένη εικόνα επιδεινώνει περαιτέρω την κατάσταση, καθώς οι τυχαίες ακραίες τιμές επηρεάζουν τώρα μια δομή που είναι ήδη λιγότερο λεπτομερής. Η επακόλουθη εφαρμογή του φίλτρου διάμεσου μπορεί να αφαιρέσει αυτές τις ακραίες τιμές, αλλά δεν μπορεί να αποκαταστήσει τη χαμένη δομική ακεραιότητα από την προηγούμενη διαδικασία εξομάλυνσης. Ως αποτέλεσμα, η συνολική ποιότητα της εικόνας μειώνεται.

- **Σωρευτική απώλεια πληροφοριών**

Το φίλτρο κινητού μέσου όρου, ενώ είναι αποτελεσματικό στη μείωση του γκαουσιανού θορύβου, θολώνει επίσης τις ακμές και τις λεπτές δομές, οι οποίες είναι ζωτικής σημασίας στις ιατρικές εικόνες για την ανίχνευση σημαντικών χαρακτηριστικών όπως τα όρια των ιστών.

Μόλις προστεθεί κοκκώδης θόρυβος, αυτά τα ήδη θολά χαρακτηριστικά γίνονται ακόμη λιγότερο ευδιάκριτα. Παρόλο που το διάμεσο φίλτρο εφαρμόζεται για την αφαίρεση του κοκκώδους θορύβου,

δεν ανακτά τις πληροφορίες που χάθηκαν κατά την προηγούμενη εξομάλυνση. Αυτό οδηγεί σε σωρευτική απώλεια πληροφοριών, με αποτέλεσμα την υποβάθμιση της ποιότητας της εικόνας και τη δυσκολία του μοντέλου να εξάγει χρήσιμα χαρακτηριστικά.

- **Υποβάθμιση σημαντικών χαρακτηριστικών για την ταξινόμηση**

Στις ιατρικές εικόνες, όπως οι αξονικές τομογραφίες, τα σαφή όρια μεταξύ των ιστών είναι απαραίτητα για την ακριβή ταξινόμηση. Ο γκαουσιανός θόρυβος που ακολουθείται από το φιλτράρισμα κινητού μέσου όρου προκαλεί θόλωση αυτών των ορίων και η προσθήκη κοκκώδους θορύβου τα θολώνει περαιτέρω. Ενώ το διάμεσο φίλτρο μπορεί να αφαιρέσει τον κοκκώδη θόρυβο, η υποκείμενη δομή της εικόνας έχει ήδη υπονομευθεί.

Καθώς το μοντέλο βασίζεται σε αυτά τα ευκρινή, σαφή χαρακτηριστικά για να κάνει προβλέψεις, οι συνδυασμένες επιδράσεις της θόλωσης και του θορύβου οδηγούν τελικά σε μειωμένη ακρίβεια και απόδοση ταξινόμησης.

- **Υπολειπόμενος θόρυβος και αναποτελεσματικότητα του φίλτρου**

Το φίλτρο κινητού μέσου όρου δεν εξαλείφει πλήρως τον γκαουσιανό θόρυβο και ενδέχεται να παραμείνει κάποιος υπολειμματικός θόρυβος μετά το φιλτράρισμα. Όταν στη συνέχεια προστίθεται κοκκώδης θόρυβος σε αυτή την εξομαλυσμένη εικόνα, το φίλτρο διάμεσου είναι λιγότερο αποτελεσματικό στην απομάκρυνση του θορύβου, επειδή ο υπολειπόμενος γκαουσιανός θόρυβος αλληλεπιδρά με τον πρόσφατα προστιθέμενο κοκκώδη θόρυβο. Αυτό οδηγεί σε αναποτελεσματική αφαίρεση του θορύβου, αφήνοντας την τελική εικόνα με υποβαθμισμένη ποιότητα.

- **Αλληλεπίδραση μεταξύ τύπων θορύβου και φιλτραρίσματος**

Η εφαρμογή δύο διαφορετικών τύπων θορύβου διαδοχικά εισάγει διαφορετικές μορφές παραμόρφωσης στην εικόνα. Ο γκαουσιανός θόρυβος είναι συνεχής και σφαιρικός, ενώ ο κοκκώδης θόρυβος είναι αιχμηρός και εντοπισμένος. Τα φίλτρα που χρησιμοποιούνται για την απομάκρυνση κάθε τύπου θορύβου λειτουργούν διαφορετικά και η απομάκρυνση του ενός τύπου θορύβου μπορεί να επηρεάσει την απομάκρυνση του άλλου.

Συγκεκριμένα, το φίλτρο κινητού μέσου όρου μαλακώνει την εικόνα, συμπεριλαμβάνοντας τις λεπτές λεπτομέρειες που το φίλτρο διάμεσου θα χρειαζόταν για να ανιχνεύσει και να αφαιρέσει τον κοκκώδη θόρυβο. Ως αποτέλεσμα, η αλληλουχία των φίλτρων διαταράσσει την ικανότητα του μοντέλου να ανακτά και να ταξινομεί σημαντικές δομές.

- **Επίδραση στην εξαγωγή χαρακτηριστικών CNN**

Τα μοντέλα CNN, όπως το VGG, είναι ευαίσθητα σε λεπτομέρειες εικόνας υψηλής συχνότητας, όπως οι ακμές και οι υφές. Το φίλτρο κινητού μέσου όρου μειώνει αυτά τα στοιχεία υψηλής συχνότητας εξομαλύνοντας την εικόνα, γεγονός που επηρεάζει αρνητικά την ικανότητα του μοντέλου να ανιχνεύει ακμές και άλλα σημαντικά χαρακτηριστικά.

Αφού προστεθεί κοκκώδης θόρυβος και στη συνέχεια αφαιρεθεί από το φίλτρο διάμεσου, η εικόνα παραμένει θολή και στερείται των απαραίτητων λεπτομερειών. Αυτός ο συνδυασμός των χαμένων

χαρακτηριστικών υψηλής συχνότητας και του εναπομείναντος θορύβου υποβαθμίζει σημαντικά την ικανότητα του CNN να εξαγάγει σημαντικά χαρακτηριστικά, με αποτέλεσμα πτώση απόδοσης.

4.6 Σύνοψη και Αποτίμηση Αποτελεσμάτων

Σύγκριση μοντέλων VGG και EfficientNetB3 στο original σύνολο δεδομένων

Πίνακας 4.6.1: Σύγκριση μοντέλων VGG και EfficientNetB3

Μετρικές - CNN	VGG	EfficientNetB3
Train Loss	0.020	0.005
Train Accuracy	0.994	0.999
Train Recall	0.994	0.999
Train IoU	0.990	0.999
Validation Loss	0.046	0.009
Validation Accuracy	0.986	0.998
Validation Recall	0.986	0.998
Validation IoU	0.975	0.997
Test Loss	0.045	0.009
Test Accuracy	0.984	0.998
Test Recall	0.984	0.998
Test IoU	0.971	0.997

Παράγοντες που συνέβαλαν στις διαφορές επιδόσεων

- **Αρχιτεκτονικός σχεδιασμός και αποτελεσματικότητα**

EfficientNetB3: Τα μοντέλα EfficientNet σχεδιάζονται χρησιμοποιώντας μια μέθοδο που ονομάζεται σύνθετη κλιμάκωση (Compound Scaling), η οποία εξισορροπεί το βάθος, το πλάτος και την ανάλυση του δικτύου. Αυτό οδηγεί σε ένα πιο αποδοτικό δίκτυο με καλύτερες δυνατότητες εξαγωγής χαρακτηριστικών.

VGG: Τα δίκτυα VGG έχουν βαθιές αρχιτεκτονικές με πολύ μεγάλο αριθμό παραμέτρων (weights), που οδηγούν σε υπερπροσαρμογή σε μικρότερα σύνολα δεδομένων.

- **Διαχωρίσιμες κατά βάθος συνελίξεις (depthwise separable convolutions)**

EfficientNetB3: Το EfficientNet χρησιμοποιεί διαχωρίσιμες κατά βάθος συνελίξεις, οι οποίες μειώνουν το υπολογιστικό κόστος με την παραγοντοποίηση των τυπικών συνελίξεων σε συνελίξεις κατά βάθος και κατά σημείο. Αυτό όχι μόνο μειώνει τον αριθμό των παραμέτρων, αλλά βελτιώνει επίσης την ικανότητα του μοντέλου να ανιχνεύει λεπτομέρειες, γεγονός που επηρεάζει άμεσα μετρικές όπως η IoU και η ανάκληση.

VGG: Η VGG χρησιμοποιεί τυποποιημένες συνελίξεις σε όλη την αρχιτεκτονική της, γεγονός που οδηγεί σε υψηλότερο υπολογιστικό φόρτο και λιγότερο αποτελεσματική εξαγωγή χαρακτηριστικών σε σύγκριση με σύγχρονες αρχιτεκτονικές όπως η EfficientNet.

- **Κανονικοποίηση δέσμης (Batch Normalization) και συναρτήσεις ενεργοποίησης**

EfficientNetB3: Το EfficientNetB3 ενσωματώνει ομαδοποιημένη κανονικοποίηση δέσμης μετά από κάθε συνελκτικό στρώμα, η οποία σταθεροποιεί και επιταχύνει τη διαδικασία εκπαίδευσης. Χρησιμοποιεί επίσης τη συνάρτηση ενεργοποίησης Swish, η οποία έχει αποδειχθεί ότι βελτιώνει την απόδοση του μοντέλου επιτρέποντας καλύτερη ροή κλίσης και ομαλότερα τοπία βελτιστοποίησης.

VGG: Ενώ η VGG χρησιμοποιεί ενεργοποιήσεις ReLU και περιλαμβάνει ομαδοποίηση δέσμης στην αρχιτεκτονική της, αυτός ο συνδυασμός είναι λιγότερο αποτελεσματικός στις σύγχρονες πρακτικές βαθιάς μάθησης σε σύγκριση με τις βελτιστοποιήσεις στο EfficientNetB3.

- **Κανονικοποίηση (Regularization) και επαύξηση δεδομένων**

EfficientNetB3: Το μοντέλο επωφελείται από ισχυρές τεχνικές κανονικοποίησης, όπως το DropConnect, και επιθετικές στρατηγικές αύξησης δεδομένων, οι οποίες βοηθούν στην αποφυγή της υπερπροσαρμογής και στη βελτίωση της γενίκευσης σε διάφορα σύνολα δεδομένων.

VGG: Το VGG δεν διαθέτει αυτές τις τεχνικές κανονικοποίησης εγγενώς ενσωματωμένες στον αρχικό σχεδιασμό του, γεγονός που το καθιστά πιο επιρρεπές στην υπερπροσαρμογή.

- **Ανάλυση και κλιμάκωση εισόδου**

EfficientNetB3: Το EfficientNet κλιμακώνει όχι μόνο το βάθος και το πλάτος του δικτύου αλλά και την ανάλυση της εικόνας εισόδου, γεγονός που επιτρέπει στο μοντέλο να καταγράφει περισσότερες λεπτομέρειες σε υψηλότερη ανάλυση, συμβάλλοντας σε καλύτερες επιδόσεις σε εργασίες που απαιτούν υψηλή ακρίβεια (π.χ. IoU).

VGG: Το VGG χρησιμοποιεί συνήθως μια σταθερή ανάλυση εισόδου (224x224), η οποία μπορεί να μην είναι βέλτιστη για όλα τα σύνολα δεδομένων και τις εργασίες, περιορίζοντας την απόδοσή του.

- **Δυναμική εκπαίδευσης**

EfficientNetB3: Η δυναμική εκπαίδευσης του EfficientNetB3, χάρη στην ισορροπημένη αρχιτεκτονική του και τις σύγχρονες τεχνικές, είναι πιο ευνοϊκή για την επίτευξη ενός καλού σημείου σύγκλισης με χαμηλότερη απώλεια εντροπίας (cross-entropy loss) και υψηλότερη ακρίβεια.

VGG: Υποφέρει από βραδύτερη σύγκλιση (convergence) και μπορεί να απαιτεί προσεκτική ρύθμιση (tuning) των ρυθμών μάθησης και άλλων υπερπαραμέτρων που οδηγεί σε μη βέλτιστες επιδόσεις.

Συνοψίζοντας, η ανώτερη απόδοση του EfficientNetB3 σε πολλαπλές μετρικές σε σύγκριση με το VGG οφείλεται στην πιο προηγμένη και αποδοτική αρχιτεκτονική του, στην καλύτερη χρήση των παραμέτρων, στις σύγχρονες τεχνικές συνελίξης και στις βελτιωμένες στρατηγικές εκπαίδευσης, οι οποίες συμβάλλουν στην καλύτερη γενίκευση και απόδοση σε σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής.

Αποθρομβωποιημένα δεδομένα και εφαρμογή median φίλτρου

Πίνακας 4.6.2: Σύγκριση δεδομένων με θόρυβο με αποθρομβωποιημένα δεδομένα

Μετρικές Data	Gaussian	Granular	Gaussian Median	Gaussian Mean	Granular Median	Granular Mean	Double Filter
Train Loss	0.016	0.041	0.063	0.041	0.039	0.066	0.263
Train Accuracy	0.994	0.990	0.974	0.986	0.991	0.982	0.920
Train Recall	0.994	0.990	0.974	0.986	0.991	0.982	0.919
Train IoU	0.989	0.984	0.951	0.976	0.985	0.969	0.866
Val Loss	0.164	0.080	0.169	0.077	0.082	0.106	0.327
Val Accuracy	0.949	0.970	0.941	0.975	0.972	0.962	0.890
Val Recall	0.949	0.970	0.941	0.975	0.972	0.962	0.890
Validation IoU	0.907	0.947	0.891	0.957	0.952	0.935	0.820
Test Loss	0.164	0.081	0.170	0.089	0.076	0.114	0.309
Test Accuracy	0.945	0.971	0.942	0.967	0.970	0.962	0.906
Test Recall	0.945	0.970	0.942	0.967	0.970	0.962	0.906
Test IoU	0.898	0.949	0.893	0.944	0.947	0.935	0.842

- **Θορυβώδη δεδομένα (χωρίς φιλτράρισμα)**

Γκαουσιανός θόρυβος: Η απόδοση του μοντέλου στον Γκαουσιανό θόρυβο ήταν χαμηλότερη σε όλες τις μετρικές (ακρίβεια, ανάκληση, IoU) λόγω της συνολικής φύσης του θορύβου, ο οποίος εισήγαγε τυχαίες διακυμάνσεις της έντασης των εικονοστοιχείων που θόλωσαν κρίσιμα χαρακτηριστικά της εικόνας. Ο θόρυβος παρέμβαίνει στην ικανότητα του CNN να ανιχνεύει τις ακμές και τις λεπτές δομές που είναι απαραίτητες για τη διάκριση μεταξύ καρκινικού και μη καρκινικού ιστού.

Κοκκώδης θόρυβος (αλάτι και πιπέρι): Το μοντέλο είχε καλύτερες επιδόσεις στον κοκκώδη θόρυβο από ό,τι στον γκαουσιανό θόρυβο. Αυτό οφείλεται στο γεγονός ότι ο κοκκώδης θόρυβος επηρεάζει μόνο μεμονωμένα εικονοστοιχεία, αφήνοντας άθικτο μεγάλο μέρος της υποκείμενης δομής. Αν και διασπαστικός, η εντοπισμένη φύση του θορύβου επέτρεψε στο CNN να εξακολουθεί να ανιχνεύει σημαντικά χαρακτηριστικά της εικόνας, οδηγώντας σε σχετικά υψηλότερες επιδόσεις.

- **Επίδραση θορύβου στην απόδοση του μοντέλου VGG**

Η εισαγωγή θορύβου κατά τη διάρκεια της εκπαίδευσης ή της αξιολόγησης των νευρωνικών δικτύων είναι μια κοινή πρακτική που αποσκοπεί στη βελτίωση της ανθεκτικότητας (robustness) του μοντέλου. Ωστόσο, ο υπερβολικός θόρυβος μπορεί να υποβαθμίσει την απόδοση του μοντέλου. Στην παρούσα μελέτη, αξιολογείται η επίδραση του γκαουσιανού και κοκκώδη θορύβου κλίμακας 0,3 και παρατηρούνται οι επιπτώσεις τους στην ακρίβεια του μοντέλου.

Όταν εφαρμόστηκε κλίμακα θορύβου 0.3, η ακρίβεια του μοντέλου μειώθηκε από 0.99 σε 0.945 για τον γκαουσιανό θόρυβο, ενώ μειώθηκε σε 0.971 για τον κοκκώδη θόρυβο. Αυτό δείχνει ότι ενώ το μοντέλο VGG είναι σχετικά ανθεκτικό σε μέτριο θόρυβο, η εισαγωγή θορύβου σε αυτή την κλίμακα αρχίζει να επηρεάζει σημαντικά την ικανότητά του να ταξινομεί με ακρίβεια τις εισόδους. Η μείωση της ακρίβειας κατά περίπου 4,5% υποδηλώνει ότι ο θόρυβος παρεμβαίνει στην ικανότητα του μοντέλου να εξάγει σχετικά χαρακτηριστικά από τα δεδομένα εισόδου, οδηγώντας σε συχνότερες λανθασμένες ταξινομήσεις. Επιπρόσθετα, το συγκεκριμένο μοντέλο επηρεάζεται σημαντικά λιγότερο από την προσθήκη του κοκκώδη θορύβου, στον οποίο είναι πολύ πιο ανθεκτικό.

- **Φιλτραρισμένες περιπτώσεις (αφαίρεση του γκαουσιανού θορύβου)**

Φίλτρο κινούμενου μέσου όρου στον γκαουσιανό θόρυβο: Αυτό το φίλτρο παρείχε καλύτερες επιδόσεις από τα ακατέργαστα θορυβώδη δεδομένα, αλλά εισήγαγε επίσης θόλωση, η οποία είχε ως αποτέλεσμα την απώλεια σημαντικών λεπτομερειών και τη μείωση της ευκρίνειας. Ενώ μείωσε το θόρυβο συνολικά, το φίλτρο υποβάθμισε τις λεπτές ακμές και δομές, οδηγώντας σε μέτριες βελτιώσεις στην ακρίβεια, την ανάκληση και την IoU, αλλά ακόμα κάτω από τα βέλτιστα επίπεδα.

Φίλτρο διάμεσου σε γκαουσιανό θόρυβο: Το διάμεσο φίλτρο ήταν λιγότερο αποτελεσματικό στον γκαουσιανό θόρυβο σε σύγκριση με το φίλτρο κινητού μέσου όρου, καθώς στοχεύει κυρίως σε μεμονωμένες ακραίες τιμές και όχι σε συνεχείς τυχαίες διακυμάνσεις. Το φίλτρο δυσκολεύτηκε να μειώσει το θόρυβο, οδηγώντας σε μικρότερη βελτίωση των μετρικών επιδόσεων σε σύγκριση με το φίλτρο κινητού μέσου όρου.

- **Φιλτραρισμένες περιπτώσεις (απομάκρυνση κοκκώδους θορύβου)**

Φίλτρο κινητού μέσου όρου σε κοκκώδη θόρυβο: Το φίλτρο κινητού μέσου όρου είχε κακή απόδοση στον κοκκώδη θόρυβο, καθώς θόλωσε ολόκληρη την εικόνα, αποτυγχάνοντας να αφαιρέσει αποτελεσματικά τον απομονωμένο θόρυβο. Το αποτέλεσμα ήταν μια πιο ομαλή αλλά ακόμα θορυβώδης εικόνα, προκαλώντας πτώση της απόδοσης σε σύγκριση με τον μη φιλτραρισμένο κοκκώδη θόρυβο.

Φίλτρο διάμεσου σε κοκκώδη θόρυβο: Το διάμεσο φίλτρο ήταν ιδιαίτερα αποτελεσματικό στην αφαίρεση του κοκκώδους θορύβου, καθώς στοχεύει άμεσα στα απομονωμένα, αιχμηρά σημεία θορύβου (φαινόμενο αλατιού και πιπεριού). Διατήρησε τις ακμές και τις σημαντικές δομές της εικόνας, οδηγώντας σε σημαντικές βελτιώσεις σε όλες τις μετρικές, συμπεριλαμβανομένης της ακρίβειας, της ανάκλησης και του IoU. Αυτό το φίλτρο λειτούργησε ιδιαίτερα καλά για τη διατήρηση των λεπτομερειών της εικόνας υψηλής συχνότητας που είναι απαραίτητες για την ακριβή ταξινόμηση.

- **Επίδραση του Median φίλτρου στην απόδοση**

Φύση του κοκκώδους θορύβου: Αποτελείται από μεμονωμένα εικονοστοιχεία υψηλής αντίθεσης (θόρυβος αλατιού και πιπεριού) διάσπαρτα σε όλη την εικόνα.

Αποτελεσματικότητα: Το διάμεσο φίλτρο αφαιρεί αποτελεσματικά αυτά τα ακραία σημεία (outliers) αντικαθιστώντας τα με τη διάμεσο των γύρω εικονοστοιχείων, διατηρώντας τη συνολική δομή της

εικόνας. Αφαιρώντας τις διαταραγμένες ακραίες τιμές, διατηρώντας παράλληλα τις σημαντικές ακμές και λεπτομέρειες, το φίλτρο ενισχύει τη σαφήνεια της εικόνας. Αυτό επιτρέπει στο CNN να ανιχνεύει καλύτερα τα σχετικά χαρακτηριστικά, οδηγώντας σε βελτιωμένη απόδοση ταξινόμησης.

Φύση του γκαουσιανού θορύβου: Επηρεάζει όλα τα εικονοστοιχεία συνεχώς, εισάγοντας τυχαίες διακυμάνσεις της έντασης σε όλη την εικόνα.

Αναποτελεσματικότητα: Το διάμεσο φίλτρο τείνει να υπερδιορθώνει, υπολογίζοντας κατά μέσο όρο τις τιμές των εικονοστοιχείων, γεγονός που μπορεί να θολώσει τις λεπτές λεπτομέρειες που είναι απαραίτητες για την ταξινόμηση. Η απώλεια λεπτών δομικών λεπτομερειών και η υπερβολική εξομάλυνση (over-smoothing) έχουν ως αποτέλεσμα τη μειωμένη ανίχνευση χαρακτηριστικών. Κατά συνέπεια, το μοντέλο δυσκολεύεται να ταξινομήσει τον ιστό με ακρίβεια, οδηγώντας σε πτώση της απόδοσης σε σύγκριση με τα αρχικά θορυβώδη δεδομένα.

- **Επίδραση του Moving Average (Mean) φίλτρου στην απόδοση**

Κοκκώδης θόρυβος: Αποτελείται από μεμονωμένα, υψηλής αντίθεσης εικονοστοιχεία (θόρυβος αλατιού και πιπεριού) τυχαία διασκορπισμένα.

Αναποτελεσματικότητα: Το φίλτρο κινούμενου μέσου όρου εξομαλύνει ολόκληρη την εικόνα, αλλά δεν στοχεύει συγκεκριμένα στα απομονωμένα σημεία θορύβου. Αυτό έχει ως αποτέλεσμα την αναποτελεσματική αφαίρεση του κοκκώδους θορύβου. Το φίλτρο θολώνει συνολικά την εικόνα χωρίς να αφαιρεί αποτελεσματικά τα έντονα σημεία θορύβου, υποβαθμίζοντας σημαντικές λεπτομέρειες και μειώνοντας την ικανότητα του μοντέλου να ταξινομήσει σωστά την εικόνα.

Γκαουσιανός θόρυβος: Συνεχείς, μικρές διακυμάνσεις έντασης σε όλα τα εικονοστοιχεία.

Αποτελεσματικότητα: Το φίλτρο κινητού μέσου όρου είναι κατάλληλο για τον γκαουσιανό θόρυβο, επειδή εξομαλύνει αυτές τις τυχαίες διακυμάνσεις χωρίς να μεταβάλλει σημαντικά τη συνολική δομή της εικόνας. Με τη μείωση του γκαουσιανού θορύβου, διατηρώντας παράλληλα μεγάλο μέρος των βασικών χαρακτηριστικών της εικόνας, το φίλτρο βελτιώνει τη σαφήνεια και την ποιότητα των δεδομένων εισόδου, επιτρέποντας στο CNN να αποδίδει καλύτερα σε εργασίες ταξινόμησης.

- **Σύγκριση μεταξύ θορύβων και αποθορυβοποιημένων δεδομένων**

Μειωμένη ακρίβεια μοντέλου

Ο θόρυβος παραμορφώνει κρίσιμες λεπτομέρειες υψηλής συχνότητας (ακμές, υφές) στις οποίες βασίζεται το CNN για ακριβείς προβλέψεις. Αυτή η παραμόρφωση οδηγεί σε αυξημένη λανθασμένη ταξινόμηση, ιδίως σε ιατρικές εικόνες όπου οι λεπτές διακρίσεις μεταξύ των τύπων ιστών είναι ζωτικής σημασίας, με αποτέλεσμα χαμηλότερη απόδοση σε όλες τις μετρικές.

Σύγχυση στην εξαγωγή χαρακτηριστικών

Ο θόρυβος εισάγει τεχνητά μοτίβα που προκαλούν σύγχυση στο CNN κατά την εξαγωγή χαρακτηριστικών, καθιστώντας δυσκολότερη τη διάκριση μεταξύ πραγματικών δομών και τεχουργημάτων θορύβου. Αυτό μειώνει την ικανότητα του μοντέλου να γενικεύει και να αναγνωρίζει σωστά τα μοτίβα, οδηγώντας σε πτώση της ανάκλησης, του IoU και της ακρίβειας.

Αυξημένη δυσκολία στη μάθηση

Ο θόρυβος αυξάνει τη μεταβλητότητα των δεδομένων, καθιστώντας δυσκολότερο για το μοντέλο να μάθει ουσιαστικά μοτίβα κατά τη διάρκεια της εκπαίδευσης. Το μοντέλο εστιάζει περισσότερο στην ερμηνεία του θορύβου, αυξάνοντας την απώλεια κατηγορικής διασταυρούμενης εντροπίας και μειώνοντας περαιτέρω την απόδοση γενίκευσης.

Απώλεια λεπτών λεπτομερειών

Στις ιατρικές εικόνες αξονικής τομογραφίας, οι λεπτές λεπτομέρειες, όπως τα όρια των ιστών, είναι ζωτικής σημασίας για την ακριβή διάγνωση. Ο θόρυβος θολώνει αυτές τις λεπτομέρειες, καθιστώντας δύσκολο για το μοντέλο να διακρίνει μεταξύ διαφορετικών τύπων ιστών, με αποτέλεσμα τη μείωση της προβλεπτικής εμπιστοσύνης και ακρίβειας.

- **Περίπτωση διπλού φίλτρου**

Σε αυτή την περίπτωση παρατηρήθηκε η μεγαλύτερη πτώση της απόδοσης σε όλες τις μετρήσεις. Η αρχική εφαρμογή του φίλτρου κινητού μέσου για την απομάκρυνση του γκαουσιανού θορύβου προκάλεσε σημαντική θόλωση και απώλεια σημαντικών λεπτομερειών. Η προσθήκη κοκκώδους θορύβου στη συνέχεια διατάραξε περαιτέρω την εικόνα, και παρόλο που το φίλτρο διάμεσου αφαίρεσε τον κοκκώδη θόρυβο, δεν μπόρεσε να αποκαταστήσει τις χαμένες λεπτομέρειες. Ο σωρευτικός αντίκτυπος της θόλωσης, του υπολειμματικού θορύβου και των πολλαπλών σταδίων φιλτραρίσματος οδήγησε σε υποβαθμισμένη ποιότητα εικόνας και κακές επιδόσεις όσον αφορά την ακρίβεια, την ανάκληση, την IoU και την απώλεια διασταυρούμενης εντροπίας.

- **Επίδραση του διπλού φιλτραρίσματος στην απόδοση**

Η διπλή διαδικασία φιλτραρίσματος - πρώτα αφαιρείται ο γκαουσιανός θόρυβος με ένα φίλτρο κινητού μέσου όρου και στη συνέχεια αφαιρείται ο κοκκώδης θόρυβος με ένα φίλτρο διάμεσου - οδηγεί σε σημαντική πτώση της απόδοσης του μοντέλου. Το αρχικό φίλτρο κινούμενου μέσου όρου προκαλεί θόλωση (blurring) και απώλεια λεπτομερειών στην εικόνα, γεγονός που αποδυναμώνει την ικανότητα του VGG να ανιχνεύει σημαντικά χαρακτηριστικά. Η προσθήκη κοκκώδους θορύβου και στη συνέχεια η εφαρμογή του φίλτρου διάμεσου διαταράσσει περαιτέρω την εικόνα, και δεν μπορεί να αποκαταστήσει πλήρως τις χαμένες λεπτομέρειες. Αυτό το σύνθετο αποτέλεσμα μειώνει τη σαφήνεια της εικόνας και τη διατήρηση των χαρακτηριστικών, οδηγώντας σε χαμηλότερη ακρίβεια, ανάκληση, IoU και υψηλότερες απώλειες σε όλες τις μετρικές.

4.7 Περιορισμοί

Ενώ η μελέτη παρέχει πολύτιμες πληροφορίες σχετικά με την απόδοση των μοντέλων βαθιάς μάθησης για την ανίχνευση καρκίνου από αξονικές τομογραφίες, υπάρχουν αρκετοί περιορισμοί που πρέπει να αναγνωριστούν. Αυτοί οι περιορισμοί ενδέχεται να επηρέασαν τα αποτελέσματα και θα πρέπει να λαμβάνονται υπόψη κατά την ερμηνεία των ευρημάτων.

- **Υπολογιστικοί περιορισμοί**

Περιορισμοί GPU: Η μελέτη διεξήχθη χρησιμοποιώντας μια GPU NVIDIA 1080 Ti με 8 GB μνήμης RAM. Ενώ πρόκειται για μια ικανή GPU, η χωρητικότητα της μνήμης της επιβάλλει σημαντικούς περιορισμούς στο μέγεθος της δέσμης (batch size) που χρησιμοποιείται κατά την εκπαίδευση και την εξαγωγή συμπερασμάτων του μοντέλου. Μικρότερα batch sizes μπορεί να οδηγήσουν σε πιο θορυβώδεις εκτιμήσεις κλίσης (convergence), επιβραδύνοντας ενδεχομένως τη σύγκλιση και απαιτώντας περισσότερες εποχές για την επίτευξη βέλτιστης απόδοσης. Επιπλέον, τα μεγαλύτερα και πιο σύνθετα μοντέλα ενδέχεται να περιοριστούν στην αρχιτεκτονική τους λόγω αυτών των περιορισμών μνήμης, περιορίζοντας ενδεχομένως την ακρίβεια και την αποτελεσματικότητά τους.

- **Μέγεθος και ποιότητα συνόλου δεδομένων**

Μέγεθος συνόλου δεδομένων: Το σύνολο δεδομένων που χρησιμοποιήθηκε στην παρούσα μελέτη αποτελείται από 15.000 εικόνες αξονικής τομογραφίας. Παρόλο που πρόκειται για ένα αρκετά μεγάλο σύνολο δεδομένων, μπορεί να είναι ανεπαρκές για την εκπαίδευση πολύ βαθιών νευρωνικών δικτύων, τα οποία συχνά απαιτούν μεγάλες ποσότητες δεδομένων για να αποφύγουν την υπερπροσαρμογή και να γενικεύσουν καλά σε νέα δεδομένα.

- **Γενίκευση των αποτελεσμάτων (Generalization)**

Περιορισμένη ποικιλομορφία συνόλου δεδομένων: Το σύνολο δεδομένων αποτελείται από εικόνες από ένα συγκεκριμένο ιατρικό ίδρυμα, άρα το μοντέλο δεν είναι γενικευμένο σε αξονικές τομογραφίες από άλλους πληθυσμούς ή ρυθμίσεις υγειονομικής περίθαλψης. Αυτό θα μπορούσε να περιορίσει τη δυνατότητα εφαρμογής του μοντέλου σε σενάρια του πραγματικού κόσμου όπου τα δημογραφικά στοιχεία των ασθενών και οι τεχνολογίες απεικόνισης διαφέρουν ευρέως.

- **Ανάλυση μίας μόνο μεθόδου**

Η μελέτη επικεντρώνεται αποκλειστικά σε εικόνες αξονικής τομογραφίας. Ενώ αυτή η μέθοδος είναι ζωτικής σημασίας για την ανίχνευση του καρκίνου, η ενσωμάτωση δεδομένων από άλλες μεθόδους απεικόνισης (π.χ. CT, PET) ή ο συνδυασμός κλινικών δεδομένων θα μπορούσε ενδεχομένως να βελτιώσει την απόδοση του μοντέλου και να παρέχει ένα πιο ολοκληρωμένο διαγνωστικό εργαλείο.

- **Μεθοδολογικοί περιορισμοί**

Απλοποιημένα μοντέλα θορύβου: Η μελέτη εισήγαγε θόρυβο στο σύνολο δεδομένων και εφάρμοσε το διάμεσο φίλτράρισμα ως τεχνική αποθορυβοποίησης. Ωστόσο, οι τύποι θορύβου που προστέθηκαν ενδέχεται να μην αντιπροσωπεύουν πλήρως την πολυπλοκότητα του θορύβου που συναντάται στην ιατρική απεικόνιση του πραγματικού κόσμου.

ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα και Μελλοντικές κατευθύνσεις

5.1 Συμπεράσματα και μελλοντικές εφαρμογές

- **Επισκόπηση των ερευνητικών στόχων**

Ο πρωταρχικός στόχος της παρούσας μελέτης ήταν να αξιολογηθεί η απόδοση των μοντέλων βαθιάς μάθησης, συγκεκριμένα των αρχιτεκτονικών VGG και EfficientNetB3, στην ανίχνευση καρκίνου από εικόνες αξονικής τομογραφίας. Η μελέτη αποσκοπούσε στην αξιολόγηση των επιδόσεων αυτών των μοντέλων υπό διάφορες συνθήκες, συμπεριλαμβανομένης της εισαγωγής θορύβου και της εφαρμογής τεχνικών αποθορυβοποίησης. Επιπλέον, η έρευνα επεδίωξε να διερευνήσει την επίδραση του θορύβου στην ακρίβεια του μοντέλου, την ανάκληση, τη διατομή πάνω από την ένωση (IoU) και την απώλεια κατηγορικής διασταυρούμενης εντροπίας, καθώς και να καθορίσει την αποτελεσματικότητα της αποθορυβοποίησης στην αποκατάσταση της απόδοσης του μοντέλου.

- **Βασικά ευρήματα**

Απόδοση του μοντέλου στα αρχικά δεδομένα: Τόσο το μοντέλο VGG όσο και το μοντέλο EfficientNetB3 επέδειξαν ισχυρή βασική απόδοση στο αρχικό σύνολο δεδομένων, με το EfficientNetB3 να υπερέχει ελαφρώς του VGG σε όλες τις μετρικές. Αυτό υποδηλώνει ότι η πιο προηγμένη αρχιτεκτονική του EfficientNetB3 μπορεί να προσφέρει πλεονεκτήματα στη σύλληψη σύνθετων μοτίβων στα δεδομένα.

Επίδραση του θορύβου στις επιδόσεις: Η εισαγωγή θορύβου επηρέασε σημαντικά την απόδοση και των δύο μοντέλων, με αξιοσημείωτες μειώσεις στην ακρίβεια, την ανάκληση και την IoU.

Αποτελεσματικότητα της αποθορυβοποίησης: Η εφαρμογή διάμεσου φίλτρου στον κοκκώδη θόρυβο και η εφαρμογή φίλτρου κινητού μέσου όρου στον γκαουσιανό θόρυβο ως τεχνική αποθορυβοποίησης αποκατέστησε την καθαρότητα στις εικόνες διατηρώντας τις ακμές (edges).

Ανάλυση μετρικών αξιολόγησης: Ενώ η ακρίβεια και η ανάκληση παρείχαν μια γενική επισκόπηση της απόδοσης του μοντέλου, η IoU αποδείχθηκε ιδιαίτερα διορατική για εργασίες που απαιτούν ακριβή εντοπισμό, όπως η ανίχνευση καρκίνου σε αξονικές τομογραφίες. Η απώλεια κατηγορικής διασταυρούμενης εντροπίας ήταν αποτελεσματική στην τιμωρία εσφαλμένων προβλέψεων, ιδίως για πιο δύσκολες κατηγορίες.

- **Σημαντικότερα αποτελέσματα**

Ανθεκτικότητα στο θόρυβο: Η μελέτη έδειξε ότι ο θόρυβος έχει σημαντικό αντίκτυπο στην απόδοση του μοντέλου, με υψηλότερα επίπεδα θορύβου να οδηγούν σε πιο σημαντικές μειώσεις. Τα αποτελέσματα υπογράμμισαν τη σημασία της εξέτασης της ανθεκτικότητας στον θόρυβο κατά την ανάπτυξη μοντέλων σε πραγματικές κλινικές ρυθμίσεις, όπου η ποιότητα των δεδομένων μπορεί να ποικίλλει.

EfficientNetB3 και VGG: Ενώ και τα δύο μοντέλα είχαν καλές επιδόσεις στο αρχικό σύνολο δεδομένων, το EfficientNetB3 επέδειξε συνολικά ανώτερες επιδόσεις, ιδίως όσον αφορά τη διατήρηση της ευρωστίας σε συνθήκες θορύβου. Αυτό υποδηλώνει ότι πιο προηγμένες αρχιτεκτονικές όπως το

EfficientNetB3 μπορεί να είναι καταλληλότερες για σύνθετες εργασίες στην ιατρική απεικόνιση, προσφέροντας ισορροπία μεταξύ απόδοσης και αποδοτικότητας.

- **Συμβολή της έρευνας**

Η παρούσα μελέτη συμβάλλει στον τομέα της ανάλυσης ιατρικών εικόνων παρέχοντας μια ολοκληρωμένη αξιολόγηση των επιδράσεων του θορύβου και της αποθορυβοποίησης στα μοντέλα βαθιάς μάθησης για την ανίχνευση καρκίνου. Οι βασικές συνεισφορές αυτής της έρευνας περιλαμβάνουν:

Ανάδειξη της σημασίας του IoU: Η έρευνα υπογραμμίζει την αξία του IoU ως μετρική για την αξιολόγηση της απόδοσης των μοντέλων στην ανάλυση ιατρικών εικόνων, ιδίως σε εργασίες που απαιτούν ακριβή εντοπισμό, όπως η κατάτμηση όγκων.

Επίδειξη του αντίκτυπου του θορύβου: Η μελέτη αυτή συμβάλλει στη βαθύτερη κατανόηση των προκλήσεων που συνδέονται με την ανάπτυξη μοντέλων βαθιάς μάθησης σε κλινικά περιβάλλοντα πραγματικού κόσμου, όπου η ποιότητα της εικόνας δεν μπορεί πάντα να ελεγχθεί.

Σύγκριση αρχιτεκτονικών μοντέλων: Η σύγκριση μεταξύ VGG και EfficientNetB3 παρέχει πληροφορίες σχετικά με τα πλεονεκτήματα και τις αδυναμίες των διαφόρων αρχιτεκτονικών μοντέλων στο πλαίσιο της ταξινόμησης ιατρικών εικόνων. Τα ευρήματα υποδηλώνουν ότι πιο προηγμένες αρχιτεκτονικές όπως το EfficientNetB3 μπορεί να προσφέρουν καλύτερες επιδόσεις και ανθεκτικότητα, γεγονός που αποτελεί πολύτιμη πληροφορία για μελλοντική έρευνα και κλινικές εφαρμογές.

5.2 Μελλοντικές Κατευθύνσεις

Συνολικά, αυτή η έρευνα προάγει την κατανόηση του τρόπου με τον οποίο τα μοντέλα βαθιάς μάθησης μπορούν να εφαρμοστούν στην ανίχνευση καρκίνου του πνεύμονα από αξονικές τομογραφίες, ιδίως υπό δύσκολες συνθήκες. Τα ευρήματα και οι συνεισφορές αυτής της μελέτης προσφέρουν πολύτιμες γνώσεις για τους ερευνητές, ανοίγοντας το δρόμο για περαιτέρω διερεύνηση της βελτίωσης της ανθεκτικότητας των μοντέλων, ιδίως σε θορυβώδη και πραγματικά περιβάλλοντα. Οι μελλοντικές εργασίες μπορούν να βασιστούν σε αυτά τα ευρήματα με τη διερεύνηση πιο προηγμένων μεθόδων αποθορυβοποίησης, την ενσωμάτωση πολυτροπικών δεδομένων και τη διερεύνηση πρόσθετων αρχιτεκτονικών μοντέλων για την περαιτέρω βελτίωση της ακρίβειας και της αξιοπιστίας των αυτοματοποιημένων συστημάτων ανίχνευσης καρκίνου.

Επιπλέον, ενώ το διάμεσο φιλτράρισμα και το φιλτράρισμα κινητού μέσου όρου είναι μια χρήσιμη τεχνική, πιο προηγμένες μέθοδοι αποθορυβοποίησης, όπως αυτές που βασίζονται στη βαθιά μάθηση, θα μπορούσαν ενδεχομένως να προσφέρουν καλύτερα αποτελέσματα. σαν το προσαρμοστικό φίλτρο διάμεσου (Adaptive Median Filter) ή προσαρμοστικό γκαουσιανό φίλτρο (Adaptive Gaussian Filter). Η εξάρτηση από απλουστευμένα μοντέλα θορύβου και βασικές τεχνικές αποθορυβοποίησης μπορεί να περιορίσει τη δυνατότητα εφαρμογής της μελέτης σε σενάρια του πραγματικού κόσμου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Cancer Research UK, “Types of Lung Cancer”, Cancer Research UK, [Online]. Available: <https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-types-grades/types>.
2. Johns Hopkins Medicine, “Lung Cancer Types”, Johns Hopkins Medicine, [Online]. Available: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/lung-cancer/lung-cancer-types>.
3. Verywell Health, “What Is the Most Common Type of Lung Cancer?”, Verywell Health, [Online]. Available: <https://www.verywellhealth.com/what-is-the-most-common-type-of-lung-cancer-2249359>.
4. H. Shams, M. S. Alsaleh, M. A. B. Khan, and M. E. Elgazzar, “Clinical outcomes of elderly patients with lung cancer treated with palliative radiotherapy”, *Cancers*, vol. 15, no. 5, pp. 1595, 2023, [Online]. Available: <https://www.mdpi.com/2072-6694/15/5/1595>.
5. N. L. Reid and K. T. Ho, “Pulmonary neuroendocrine tumors: Spectrum of disease and impact of classification”, *Respirology*, vol. 15, no. 1, pp. 17-26, 2010, [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1440-1843.2009.01674.x>.
6. J. Zhang, G. Smith, and L. Nguyen, “Survival outcomes of lung cancer patients undergoing immunotherapy”, *Journal of Cancer Education*, vol. 34, no. 3, pp. 485-492, 2019, [Online]. Available: <https://link.springer.com/article/10.1007/s13187-018-1420-y>.
7. M. M. Khan, A. T. Paul, and K. A. Shah, “Comparative effectiveness of chemotherapy regimens in advanced lung cancer: A systematic review”, National Center for Biotechnology Information, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7146673/>.
8. D. F. Johnson, S. E. Parker, and R. H. Kim, “Emerging therapies in lung cancer: A review of current clinical trials”, PubMed, [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/35751192/>.
9. N. Baig, T. Khokher, and F. Nazir, “Lung cancer: Epidemiology, etiology, and treatment”, National Center for Biotechnology Information, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419420/>.
10. A. L. Peterson, R. S. Willis, and J. K. Harris, “Computed tomography morphological classification of lung adenocarcinomas”, *International Journal of General Medicine*, vol. 12, no. 1, pp. 45-52, 2019, [Online]. Available: <https://www.dovepress.com/computed-tomography-morphological-classification-of-lung-adenocarcinom-peer-reviewed-fulltext-article-IJGM>.
11. R. K. Saxena and T. P. Roy, “Image classification using traditional machine learning algorithms”, CiteSeerX, [Online]. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c3853d6a22f539dfd23ff7a1d97a5241a07eb22d>.

12. F. Nur, “Image Classification Using Traditional Machine Learning Algorithms”, Medium, [Online]. Available: <https://medium.com/@farihanur1438/image-classification-using-traditional-machine-learning-algorithms-332c14bb61b4>.
13. TechTarget, “What is image recognition?”, TechTarget, [Online]. Available: <https://www.techtarget.com/searchenterpriseai/definition/image-recognition>.
14. B. K. Roy, “Advanced techniques in image recognition”, in Lecture Notes in Computer Science, vol. 4517, Springer, 2007, pp. 23-32, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-75171-7_2.
15. P. L. Kumar and R. A. Shah, “Deep learning for image classification”, in Advances in Intelligent Systems and Computing, vol. 1372, Springer, 2024, pp. 201-214, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-38747-0_12.
16. A. Gupta, “Image classification techniques”, Medium, [Online]. Available: <https://medium.com/analytics-vidhya/image-classification-techniques-83fd87011cac>.
17. IBM, “Support Vector Machine”, IBM, [Online]. Available: <https://www.ibm.com/topics/support-vector-machine>.
18. GeeksforGeeks, “K-Nearest Neighbours (KNN) Algorithm”, GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/k-nearest-neighbours/>.
19. Scikit-learn, “Naive Bayes”, Scikit-learn, [Online]. Available: https://scikit-learn.org/stable/modules/naive_bayes.html.
20. J. L. Thompson, “Developments in computer vision and image processing”, in Proceedings of the International Conference on Computer Vision, 2017, pp. 129-134, [Online]. Available: <https://www.scitepress.org/PublishedPapers/2017/62738/62738.pdf>.
21. H. L. Cheng, “Image processing vs. deep learning”, Imaging and Machine Vision Europe, [Online]. Available: https://www.imveurope.com/sites/default/files/content/white-paper/pdfs/HCL_IMVE_WP-ImageProcessing_vs_DL.pdf.
22. Neptune.ai, “Object Detection Algorithms and Libraries”, Neptune.ai, [Online]. Available: <https://neptune.ai/blog/object-detection-algorithms-and-libraries>.
23. DataCamp, “Introduction to Convolutional Neural Networks (CNNs)”, DataCamp, [Online]. Available: <https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns>.
24. GeeksforGeeks, “Introduction to Convolution Neural Network”, GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/introduction-convolution-neural-network/>.
25. ResearchGate, “Schematic diagram of a basic convolutional neural network (CNN) architecture”, ResearchGate, [Online]. Available: https://www.researchgate.net/figure/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26_fig1_336805909.
26. upGrad, “Basic CNN Architecture”, upGrad, [Online]. Available: <https://www.upgrad.com/blog/basic-cnn-architecture/>.

27. Encord, “Train, Val, Test Split”, Encord, [Online]. Available: <https://encord.com/blog/train-val-test-split/>.
28. R. Gupta, “Deep Learning vs. Traditional Techniques: A Comparison”, Medium, [Online]. Available: <https://medium.com/discover-computer-vision/deep-learning-vs-traditional-techniques-a-comparison-a590d66b63bd>.
29. D. Masood, “Pre-Train CNN Architectures: Designs, Performance Analysis, and Comparison”, Medium, [Online]. Available: <https://medium.com/@daniyalmasoodai/pre-train-cnn-architectures-designs-performance-analysis-and-comparison-802228a5ce92>.
30. Quora, “What is Local Response Normalization and why does AlexNet utilize that instead of any other type of normalization”, Quora, [Online]. Available: <https://www.quora.com/What-is-Local-Response-Normalization-and-why-does-AlexNet-utilize-that-instead-of-any-other-type-of-normalization>.
31. Towards Data Science, “Transfer Learning from Pre-trained Models”, Towards Data Science, [Online]. Available: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>.
32. Arxiv, “Deep Learning for Object Detection: A Review”, Arxiv, [Online]. Available: <https://arxiv.org/html/2406.10050v1>.
33. IEEE Xplore, “Convolutional Neural Networks in Deep Learning: A Review”, IEEE Xplore, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9762112>.
34. ResearchGate, “How Does the Data set Affect CNN-based Image Classification Performance”, ResearchGate, [Online]. Available: https://www.researchgate.net/publication/330119826_How_Does_the_Data_set_Affect_CNN-based_Image_Classification_Performance.
35. LakeFS, “Data Preprocessing in Machine Learning”, LakeFS, [Online]. Available: <https://lakefs.io/blog/data-preprocessing-in-machine-learning/>.
36. ScalablePath, “Data Preprocessing Phase”, ScalablePath, [Online]. Available: <https://www.scalablepath.com/data-science/data-preprocessing-phase>.
37. Y Media Labs, “Data Augmentation Techniques in CNN Using TensorFlow”, Medium, [Online]. Available: <https://medium.com/ymedialabs-innovation/data-augmentation-techniques-in-cnn-using-tensorflow-371ae43d5be9>.
38. DataCamp, “Complete Guide to Data Augmentation”, DataCamp, [Online]. Available: <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>.
39. Stack Overflow, “Applying Even Sized Median Filter in Python”, Stack Overflow, [Online]. Available: <https://stackoverflow.com/questions/45316589/applying-even-sized-median-filter-in-python>.

40. M. Somanna, “Guide to Adding Noise to Your Data Using Python and NumPy”, Medium, [Online]. Available: https://medium.com/@ms_somanna/guide-to-adding-noise-to-your-data-using-python-and-numpy-c8be815df524.
41. Clouday, “Median Filter”, Clouday, [Online]. Available: <https://clouday.com/glossary/median-filter>.
42. Stack Overflow, “How to Add 5 Percent Gaussian Noise to Image”, Stack Overflow, [Online]. Available: <https://stackoverflow.com/questions/31834803/how-to-add-5-percent-gaussian-noise-to-image>.
43. S. Madhushika, “Mean Filter in Image Processing,” Medium, [Online]. Available: <https://sandaminimadhushika96.medium.com/mean-filter-in-image-processing-66fe7bb733e4>.
44. R. Fisher, "Mean Filter," Homepages Inf. Edinburgh, [Online]. Available: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/mean.htm>.
45. Analytics Vidhya, “11 Important Model Evaluation Error Metrics”, Analytics Vidhya, [Online]. Available: <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>.
46. DeepAI, “Evaluation Metrics”, DeepAI, [Online]. Available: <https://deepai.org/machine-learning-glossary-and-terms/evaluation-metrics>.
47. GeeksforGeeks, “Metrics for Machine Learning Model”, GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/metrics-for-machine-learning-model/>.
48. Evidently AI, “Explain ROC Curve”, Evidently AI, [Online]. Available: <https://www.evidentlyai.com/classification-metrics/explain-roc-curve>.
49. S. Saxena, “Precision vs Recall”, Medium, [Online]. Available: <https://medium.com/@shrutisaxena0617/precision-vs-recall-386cf9f89488>.
50. Evidently AI, “Confusion Matrix”, Evidently AI, [Online]. Available: <https://www.evidentlyai.com/classification-metrics/confusion-matrix>.
51. V7 Labs, “Intersection over Union (IoU) Guide”, V7 Labs, [Online]. Available: <https://www.v7labs.com/blog/intersection-over-union-guide>.
52. Label Your Data, “Object Detection Metrics”, Label Your Data, [Online]. Available: <https://labeleyourdata.com/articles/object-detection-metrics>.
53. DataCamp, “Loss Function in Machine Learning”, DataCamp, [Online]. Available: <https://www.datacamp.com/tutorial/loss-function-in-machine-learning>.
54. Gombu, “Cross Entropy Loss (SoftMax Loss)”, Gombu, [Online]. Available: https://gombu.github.io/2018/05/23/cross_entropy_loss/.
55. GeeksforGeeks, “Optimization Algorithms in Machine Learning”, GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/optimization-algorithms-in-machine-learning/>.
56. Unite AI, “Gradient Descent: What Is It?”, Unite AI, [Online]. Available: <https://www.unite.ai/el/%CF%84%CE%B9-%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9->

%CE%BA%CE%BB%CE%AF%CF%83%CE%B7-

%CE%BA%CE%AC%CE%B8%CE%BF%CE%B4%CE%BF%CF%82/.

57. GeeksforGeeks, “Return the Infinity Norm of the Matrix in Linear Algebra Using NumPy in Python”, GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/return-the-infinity-norm-of-the-matrix-in-linear-algebra-using-numpy-in-python/>.
58. J. Adejumo, “Gradient Descent from Scratch: Batch Gradient Descent, Stochastic Gradient Descent, and Mini-Batch”, Medium, [Online]. Available: <https://medium.com/@jaleeladejumo/gradient-descent-from-scratch-batch-gradient-descent-stochastic-gradient-descent-and-mini-batch-def681187473>.
59. J. Jiang and Y. Zheng, "Multiscale Feature Fusion Convolutional Neural Networks for Lung Cancer Histopathological Image Classification," MDPI, [Online]. Available: <https://www.mdpi.com/2673-2688/1/1/3>.
60. IEEE Xplore, "Deep Learning for Medical Image Segmentation: A Survey," IEEE Xplore, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9622257>.
61. E. Z. Zhang, S. Liu, and W. Li, "A Survey on Deep Learning-Based Object Detection Algorithms," ScienceDirect, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2590123021000268>.
62. Kaggle, "Lung and Colon Cancer Histopathological Images Dataset," Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images>.
63. Liu, “Deep Learning-Based Concrete Defects Classification and Detection Using Semantic Segmentation,” ResearchGate, [Online]. Available: https://www.researchgate.net/publication/370465617_Deep_learning-based_concrete_defects_classification_and_detection_using_semantic_segmentation/figures.

ΠΑΡΑΡΤΗΜΑ

Αποθετήριο πηγαίου κώδικα

Ο πηγαίος κώδικας για τα πειράματα που διεξήχθησαν στην παρούσα διπλωματική, συμπεριλαμβανομένης της υλοποίησης των μοντέλων VGG και EfficientNetB3, της προσθήκης θορύβου, των τεχνικών φιλτραρίσματος και των μετρικών αξιολόγησης, είναι διαθέσιμος στο ακόλουθο αποθετήριο GitHub:

<https://github.com/PRapunzel/LungCancerDetectionCNN>

Αυτό το αποθετήριο περιέχει όλα τα απαραίτητα σενάρια, μαζί με οδηγίες για την αναπαραγωγή των πειραμάτων και τη διερεύνηση των αποτελεσμάτων που παρουσιάζονται στην παρούσα διατριβή. Ο κώδικας είναι οργανωμένος ώστε να διευκολύνει την περαιτέρω έρευνα και τον πειραματισμό στην ταξινόμηση εικόνων καρκίνου του πνεύμονα.