



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΜΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΞΙΟΠΟΙΗΣΗ ΑΡΧΙΤΕΚΤΟΝΙΚΩΝ ΣΥΝΕΛΙΚΤΙΚΩΝ
ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΤΟΝ ΕΝΤΟΠΙΣΜΟ ΕΥΡΗΜΑΤΩΝ
ΣΕ ΒΙΟΨΙΕΣ ΤΟΥ ΠΑΧΕΟΣ ΕΝΤΕΡΟΥ**

Τσιλιγιάννη Θεοδώρα

Επιβλέπων: Τζάλλας Αλέξανδρος

Αναπληρωτής Καθηγητής

Άρτα, Απρίλιος, 2026

**LEVERAGING CONVOLUTIONAL NEURAL NETWORK
ARCHITECTURES FOR FINDINGS DETECTION IN COLON
BIOPSIES**

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, 01/04/2026

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Τζάλλας Αλέξανδρος,

ΔΕΠ, Αναπληρωτής καθηγητής

2. Μέλος επιτροπής

Τσούλος Ιωάννης,

ΔΕΠ, Καθηγητής

3. Μέλος επιτροπής

Γιαννακέας Νικόλαος,

ΔΕΠ, Αναπληρωτής Καθηγητής

© Τσιλιγιάννη, Θεοδώρα, 2026.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα μεταπτυχιακή εργασία είναι εκ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Τσιλιγιάννη, Θεοδώρα

Υπογραφή

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Αλέξανδρο Τζάλλα για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου το συγκεκριμένο θέμα, καθώς επίσης και για τη συνεχή στήριξή του κατά τη διάρκεια των σπουδών μου. Ευχαριστώ επίσης και τον καθηγητή μου κ. Νικόλαο Γιαννακέα, για την πολύτιμη καθοδήγηση και την υποστήριξη που μου παρείχε, κατά τη διάρκεια εκπόνησης της παρούσας μεταπτυχιακής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την στήριξή της καθ' όλη τη διάρκεια των σπουδών μου.

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία εξετάζει την αξιοποίηση αρχιτεκτονικών Συνελκτικών Νευρωνικών Δικτύων (CNNs) για τον εντοπισμό και την ανάλυση ευρημάτων σε βιοψίες του παχέος εντέρου. Αρχικά, παρουσιάζεται το θεωρητικό υπόβαθρο σχετικά με τον καρκίνο του παχέος εντέρου, την Τεχνητή Νοημοσύνη, τα Τεχνητά Νευρωνικά Δίκτυα και ιδιαίτερα τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) όπως και τις εφαρμογές τους στην ιατρική απεικόνιση.

Βασικός στόχος της εργασίας είναι η διερεύνηση, από υπολογιστικής προσέγγισης, της δυνατότητας αυτόματης τμηματοποίησης ιστολογικών περιοχών και η ποσοτική ανάλυση χαρακτηριστικών του ιστού, όπως είναι το ποσοστό κολλαγόνου. Στο πειραματικό τμήμα χρησιμοποιήθηκε το μοντέλο Mask R-CNN για την τμηματοποίηση τριών ιστολογικών περιοχών, mucosa, muscularis και submucosa, σε εικόνες βιοψίας του παχέος εντέρου. Έπειτα, με βάση τις προβλεπόμενες μάσκες που προέκυψαν από την τμηματοποίηση, εφαρμόστηκε ο αλγόριθμος $k - \text{means}$ για χρωματική ανάλυση των περιοχών, και ο υπολογισμός του δείκτη CPA, ο οποίος υποδεικνύει ποσοτικό δείκτη ύπαρξης κολλαγόνου.

Τα αποτελέσματα έδειξαν υψηλή απόδοση του μοντέλου, με Συνολική Ακρίβεια 0,961 στο επισημασμένο υποσύνολο δεδομένων και καλή αναγνώριση των τριών κλάσεων, ειδικά της κλάσης mucosa. Επιπλέον, η ποσοτική ανάλυση του δείκτη CPA έδειξε διαφοροποίηση στην ύπαρξη κολλαγόνου στις ιστολογικές περιοχές, με τις υψηλότερες τιμές ποσοστών να εμφανίζονται στις κλάσεις muscularis και submucosa.

Συνολικά, η προτεινόμενη μεθοδολογία προσφέρει έναν αξιόπιστο τρόπο ποσοτικής αξιολόγησης ιστολογικών εικόνων και θα μπορούσε να συμβάλλει θετικά σε συστήματα υποστήριξης διάγνωσης.

Λέξεις-κλειδιά: Βιοψίες Παχέος Εντέρου, Ιστοπαθολογικές Εικόνες, Συνελκτικά Νευρωνικά Δίκτυα, Mask R-CNN, Ανάλυση Κολλαγόνου.

ABSTRACT

The present thesis explores the application of Convolutional neural Network (CNN) architectures for the detection and analysis of findings in colon biopsies. Initially, the theoretical background presents the fundamental concepts related to colorectal cancer, Artificial Intelligence, Artificial Neural Networks and more specifically Convolutional Neural Networks, while also reviewing their contemporary applications in medical imaging and histopathology.

The primary aim of this study is to investigate a computational framework capable of supporting the automatic segmentation of histological regions and the quantitative assessment of tissue morphological characteristics, such as the percentage of collagen. In the experimental part, a Mask R-CNN model was used to segment three histological regions, mucosa, muscularis and submucosa, in colon biopsy images. Subsequently, the predicted segmentation masks were utilized in combination with the k-means clustering algorithm to perform color – based analysis and calculate the Collagen Proportionate Area (CPA) which served as a quantitative indicator of collagen distribution within the tissue.

The results showed high level of model performance, with an Overall Accuracy of 0.961 on the annotated subset of data and particularly strong performance in the identification of the mucosa class. Furthermore, CPA analysis revealed distinct differences among the histological regions examined, with the higher values observed in the muscularis and submucosa classes.

In conclusion, the proposed methodology provides a reliable approach to the quantitative analysis of histopathological images and may serve as a foundation for the development of future computer-based diagnostic systems.

Keywords: Colon biopsies, Histopathological Images, Convolutional Neural Networks, Mask R-CNN, Collagen Analysis.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1.	Θεωρητικό Υπόβαθρο	1
1.1.	Η ασθένεια του καρκίνου του παχέος εντέρου	2
1.2.	Αίτια ανάπτυξης της νόσου	3
1.2.1.	Οικογενειακό Ιατρικό Ιστορικό και Γενετική	3
1.2.2.	Προσωπικό Ιατρικό Ιστορικό	4
1.2.3.	Τρόπος ζωής – Διατροφικοί παράγοντες	5
1.2.4.	Άλλοι Παράγοντες	6
1.3.	Συμπτώματα του καρκίνου του παχέος εντέρου	7
1.4.	Στάδια του καρκίνου του παχέος εντέρου	7
1.5.	Διάγνωση	10
1.6.	Θεραπεία	11
2.	Τεχνητή Νοημοσύνη	16
2.1.	Εισαγωγή	16
2.2.	Τι είναι η Τεχνητή Νοημοσύνη	17
2.3.	Πεδία εφαρμογής της Τεχνητής Νοημοσύνης	19
2.4.	Η Τεχνητή Νοημοσύνη στην Ιατρική	20
2.5.	Εφαρμογή Τεχνητής Νοημοσύνης στην ιατρική	23
2.6.	Η Τεχνητή Νοημοσύνη στον Καρκίνο του Παχέος Εντέρου	25
2.7.	Μηχανική Μάθηση	26
3.	Τεχνητά Νευρωνικά Δίκτυα	28
3.1.	Εισαγωγή	28
3.2.	Νευρωνικά Δίκτυα – Βασική Δομή και Λειτουργία	29
3.2.1.	Το μοντέλο του Τεχνητού Νευρώνα	29
3.2.2.	Συνάρτηση Ενεργοποίησης	31
3.2.3.	Το Perceptron	32
3.2.4.	Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων	33
3.2.5.	Διαδικασία Μάθησης	36
3.2.6.	Ο Αλγόριθμος Οπισθοδιάδοσης του λάθους (Error Back Propagation Algorithm)	37
3.2.7.	Ενημέρωση συναπτικών βαρών με Γενικευμένο Κανόνα Δέλτα	39

3.2.8.	Αρχικοποίηση Συναπτικών Βαρών	40
3.2.9.	Προεπεξεργασία του διανύσματος εισόδου	40
3.2.10.	Εύρεση βέλτιστης Αρχιτεκτονικής Δικτύου	41
4.	Συνελικτικά Νευρωνικά Δίκτυα	43
4.1.	Η Θέση των Συνελικτικών Νευρωνικών Δικτύων στο Πεδίο της Τεχνητής Νοημοσύνης 43	
4.2.	Η Σημασία των CNNs στην Ιατρική Απεικόνιση και την Ιστοπαθολογία.....	44
4.3.	Στόχοι και Δομή Κεφαλαίου	46
4.4.	Ιστορική αναδρομή και εξέλιξη των CNN.....	47
4.4.1.	Από το Perceptron στο Neocognitron.....	47
4.4.2.	LeNet-5: Η Γέννηση του Σύγχρονου CNN.....	48
4.4.3.	Η Επανάσταση της Βαθιάς Μάθησης: AlexNet, Big Data και GPUs.....	49
4.4.4.	Η πορεία προς το Βάθος και την Αποδοτικότητα.....	50
4.5.	Θεμελιώδη στοιχεία των Συνελικτικών Νευρωνικών Δικτύων	52
4.5.1.	Δομή ενός CNN.....	52
4.5.2.	Επίπεδο Εισόδου (Input Layer).....	53
4.5.3.	Συνελικτικό Επίπεδο (Convolutional Layer).....	53
4.5.4.	Συναρτήσεις Ενεργοποίησης (Activation Functions)	55
4.5.5.	Επίπεδο Συγκέντρωσης (Pooling Layer)	56
4.5.6.	Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer - Dense).....	57
4.5.7.	Επίπεδο Εξόδου (Output Layer).....	58
4.6.	Υπολογιστική Λογική.....	59
4.6.1.	Εξίσωση Συνελίξεων (Mathematical Definition)	59
4.6.2.	Παράδειγμα 2D Συνέλιξης	60
4.6.3.	Υπολογιστική Πολυπλοκότητα και Μείωση Παραμέτρων	62
4.7.	Τεχνικές Βελτίωσης.....	64
4.7.1.	Batch Normalization (Ομαδική Κανονικοποίηση).....	64
4.7.2.	Dropout (Απόσυρση)	66
4.7.3.	Data Augmentation (Επαύξηση Δεδομένων).....	67
4.7.4.	Early Stopping (Πρώιμη Διακοπή).....	68
4.8.	Εξειδικευμένες Αρχιτεκτονικές	69
4.8.1.	VGGNet.....	69
4.9.	ResNet (Residual Networks).....	72

4.9.1.	Θεωρητική Βάση: Το Πρόβλημα της Εξαφανιζόμενης Κλίσης (Vanishing Gradient Problem).....	72
4.9.2.	Η Καινοτομία: Παρακαμπτήριες Συνδέσεις (Skip Connections / Identity Shortcuts).....	73
4.9.3.	Εκδοχές και Πλεονεκτήματα	73
4.9.4.	Εφαρμογές στην Ιστοπαθολογία (Ιατρικές Εικόνες)	74
4.10.	Transfer Learning και Fine-tuning (Μάθηση Μεταφοράς και Λεπτομερής Ρύθμιση) .	75
4.10.1.	Προεκπαίδευση σε ImageNet και Επαναχρησιμοποίηση Βαρών	76
4.10.2.	Fine-tuning (Λεπτομερής Ρύθμιση) σε Ιατρικά Datasets.....	76
5.	Εφαρμογές CNN Στη Βιοιατρική Εικόνα	79
5.1.	Η Μεταμόρφωση της Ιατρικής Διάγνωσης μέσω της Υπολογιστικής Όρασης.....	79
5.2.	Επισκόπηση Εφαρμογών CNN στην Ιστοπαθολογία.....	80
5.2.1.	Καρκίνος του Παχέος Εντέρου (CRC).....	81
5.2.2.	Καρκίνος του Μαστού (Breast Cancer).....	81
5.2.3.	Άλλοι Τύποι Ιστών (Πνεύμονας, Δέρμα, Εγκέφαλος).....	82
5.3.	ResNet: Η Αρχιτεκτονική Επιλογής για Ταξινόμηση και Πρόγνωση.....	82
5.3.1.	Εφαρμογή: Ταξινόμηση Τύπων Ιστού σε Βιοψίες Παχέος Εντέρου.....	83
5.3.2.	Εφαρμογή: Πρόγνωση Μοριακών Βιοδεικτών (π.χ. MSI).....	84
5.4.	R-CNN: Ανίχνευση και Εντοπισμός Καρκινικών Περιοχών.....	85
5.4.1.	Η Εξέλιξη της Οικογένειας R-CNN.....	85
5.4.2.	Εφαρμογή: Εντοπισμός Ευρυμάτων σε Βιοψίες Παχέος Εντέρου.....	86
6.	Ανάλυση ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου με χρήση Mask R-CNN για μέτρηση ποσοστού κολλαγόνου	87
6.1.	Ανάλυση Ιστοπαθολογικών Εικόνων	87
6.2.	Ιστολογικές Περιοχές ως Κλάσεις Segmentation	88
6.3.	Δεδομένα και Επισημάνσεις Αναφοράς	90
6.4.	Τμηματοποίηση Ιστολογικών Περιοχών με Mask R-CNN.....	92
6.5.	Mask R-CNN: Ροή Επεξεργασίας και παραγωγή τελικών μασκών.....	93
6.6.	Αξιολόγηση Τμηματοποίησης	95
6.6.1.	Πίνακας σύγχυσης	95
6.6.2.	Μετρικές ανά κλάση (Sensitivity, Specificity, F1 Score) και Συνολικές Μετρικές (Total Accuracy).....	96
6.7.	Αποτελέσματα	98
6.7.1.	Αποτελέσματα Τμηματοποίησης	98

6.7.2.	Αποτελέσματα Ποσοστού Κολλαγόνου (CPA) ανά ιστολογική περιοχή.....	101
6.8.	Συμπεράσματα.....	104
6.9.	Προτάσεις για μελλοντική έρευνα.....	106
ΒΙΒΛΙΟΓΡΑΦΙΑ.....		107

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Κανονικοποιημένος πίνακας σε ποσοστά (Confussion Matrix Percentages).....	98
Πίνακας 2 Μετρικές ανά κλάση - Μέση τιμή \pm SD τυπική απόκλιση.....	99
Πίνακας 3 Συνολική Ακρίβεια (Total Accuracy)	99
Πίνακας 4 CPA Testing % στο σύνολο των εικόνων N = 500	102
Πίνακας 5 CPA Testing % για Testing Area > 0.....	102
Πίνακας 6 CPA μεταξύ масκών αναφοράς και προβλεπόμενων масκών.....	102

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ

Εικόνα 1 Μοντέλο Τεχνητού Νευρώνα.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
Σχήμα 1 Μοντέλο Τεχνητού Νευρώνα.....	30
Σχήμα 2 Συνάρτηση Κατωφλίου	31
Σχήμα 3 Σιγμοειδής Συνάρτηση	32
Σχήμα 4 Το όριο και οι περιοχές απόφασης	33
Σχήμα 5 Δίκτυο ενός επιπέδου πρόσθιας τροφοδότησης.....	34
Σχήμα 6 Πολυεπίπεδο Δίκτυο Πρόσθιας Τροφοδότησης	35
Σχήμα 7 Αναδρομικό Νευρωνικό Δίκτυ	36
Σχήμα 8 Παραδείγματα αποτελεσμάτων Τμηματοποίησης στο σύνολο Testing.....	100
Σχήμα 9 Ενδεικτικές επισημάνσεις σε εικόνες του συνόλου Training	101

ΑΠΟΔΟΣΗ ΟΡΩΝ / ΓΛΩΣΣΑΡΙΟ

Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network – CNN): Είδος Τεχνητού Νευρωνικού Δικτύου Βαθιάς Μάθησης, ειδικά σχεδιασμένο για επεξεργασία και ανάλυση οπτικών δεδομένων. Χρησιμοποιεί συνελικτικά στρώματα για να αναγνωρίζει αυτόματα χαρακτηριστικά, όπως σχήματα και υφές, και χρησιμοποιείται ευρέως στην ιατρική απεικόνιση.

Mask R-CNN: Προηγμένο μοντέλο Βαθιάς Μάθησης που επεκτείνει το Faster R-CNN. Χρησιμοποιείται για την κατάτμηση στιγμιότυπων (instance segmentation), επιτρέποντας στον αλγόριθμο όχι μόνο να εντοπίζει και να ταξινομεί αντικείμενα σε μια εικόνα, αλλά και να σχεδιάζει το ακριβές περίγραμμά τους σε επίπεδο pixel.

Resnet: Τύπος Συνελικτικού Νευρωνικού Δικτύου που εισήγαγε την έννοια “συνδέσεις παράκαμψης” (skip connections). Αυτές, επιτρέπουν την εκπαίδευση πολύ βαθιών δικτύων, επιλύοντας το πρόβλημα της εξαφάνισης της κλίσης, βελτιώνοντας έτσι την ακρίβεια στην αναγνώριση προτύπων.

K-means Clustering (Ομαδοποίηση k – μέσων): Αλγόριθμος μη επιβλεπόμενης μάθησης που ομαδοποιεί δεδομένα σε αριθμό ομάδων με βάση την ομοιότητά τους. Στην επεξεργασία εικόνας χρησιμοποιείται για τη μείωση χρωμάτων ή τον αυτόματο διαχωρισμό περιοχών με παρόμοια οπτικά χαρακτηριστικά.

Ιστοπαθολογική εικόνα (Histopathological Image): Ψηφιακή εικόνα δείγματος βιολογικού ιστού, που έχει υποστεί ειδική επεξεργασία. Χρησιμοποιείται για την μελέτη της μορφολογίας, της δομής και πιθανών αλλοιώσεων του ιστού.

Τμηματοποίηση Εικόνας (Image Segmentation): Η διαδικασία διαχωρισμού μιας ψηφιακής εικόνας σε επιμέρους περιοχές με κοινά χαρακτηριστικά.

Στόχος είναι η απλούστευση της αναπαράστασης της εικόνας για να γίνεται ευκολότερα η ανάλυση αντικειμένων ή περιοχών ενδιαφέροντος.

Μάσκα Τμηματοποίησης (Segmentation Mask): Δυαδική ή πολυκλασική αναπαράσταση εικόνας στην οποία κάθε *pixel* αντιστοιχεί σε μια κατηγορία ή περιοχή ενδιαφέροντος.

Επισημάνσεις Αναφοράς (Ground Truth Annotations): Χειροκίνητες επισημάνσεις ή ετικέτες από ειδικούς που θεωρούνται ως το σωστό αποτέλεσμα. Χρησιμοποιούνται για την εκπαίδευση και την αξιολόγηση ενός μοντέλου μηχανικής μάθησης, ώστε να συγκρίνεται η πρόβλεψή του με την πραγματική κλάση.

CPA (Collagen Proportionate Area): Ποσοτικός δείκτης που μετράει την αναλογία της περιοχής που καταλαμβάνει το κολλαγόνο, ως προς τη συνολική περιοχή ενός δείγματος ιστού που εξετάζεται.

Mucosa (Βλεννογόνος): Η εσωτερική στοιβάδα του τοιχώματος του παχέος εντέρου η οποία έρχεται σε επαφή με τον εντερικό αυλό. Αποτελείται από συνδετικό ιστό και, σε ορισμένες περιπτώσεις, μια λεπτή στοιβάδα λείων μυϊκών ινών.

Submucosa (Υποβλεννογόνος): Η στοιβάδα που βρίσκεται κάτω από τον βλεννογόνο και περιλαμβάνει συνδετικό ιστό, αγγεία και νευρικές δομές.

Muscularis (Μυϊκός χιτώνας): Μυϊκή στοιβάδα του τοιχώματος του εντέρου που αποτελείται κυρίως από μυϊκές ίνες και συμβάλλει στις κινήσεις του εντέρου.

ΕΙΣΑΓΩΓΗ

Η τεράστια εξέλιξη της Τεχνητής Νοημοσύνης και ιδιαίτερα των Συνελκτικών Νευρωνικών Δικτύων (CNNs) έχουν συμβάλει σημαντικά στην πρόοδο, στον τομέα της επεξεργασίας εικόνας, δίνοντας έτσι την δυνατότητα αποτελεσματικότερης αναγνώρισης προτύπων, εξαγωγής σύνθετων χαρακτηριστικών και εντοπισμού περιοχών ενδιαφέροντος με μεγάλη ακρίβεια. Τα συγκεκριμένα μοντέλα CNN προφέρουν πληθώρα πλεονεκτημάτων στο πεδίο της ιατρικής απεικόνισης και στην ανάλυση ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου. Υποστηρίζουν τμηματοποίηση δομών ιστού, εντοπίζουν ευρήματα και αξιοποιούν την πληροφορία για ποσοτική αξιολόγηση χαρακτηριστικών του ιστού. Τα CNNs συμβάλλουν έτσι, στην ενίσχυση της αντικειμενικότητας της οπτικής εκτίμησης και δίνουν τη δυνατότητα για χρήση αποτελεσματικότερων αυτοματοποιημένων μεθόδων στην υποστήριξη διάγνωσης σε ιστοπαθολογική εξέταση. Το παραπάνω συνδέεται άμεσα με την παρούσα διπλωματική εργασία, στην οποία αξιοποιούνται σύγχρονες τεχνικές Συνελκτικών Νευρωνικών Δικτύων για την ανάλυση εικόνων βιοψιών του παχέος εντέρου.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η διερεύνηση του τρόπου με τον οποίο τα Συνελκτικά Νευρωνικά Δίκτυα συμβάλλουν στην ανάλυση ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου, ειδικότερα στην τμηματοποίηση περιοχών ιστού και στην ποσοτική εκτίμηση χαρακτηριστικών του. Συγκεκριμένα, η παρούσα εργασία εστιάζει στην εφαρμογή του μοντέλου Mask R-CNN για την τμηματοποίηση των περιοχών mucosa, muscularis και submucosa, που βρίσκονται στο τοίχωμα του παχέος εντέρου, και στην μετέπειτα χρήση των масκών τμηματοποίησης που παράγονται για τον υπολογισμό του ποσοστού κολλαγόνου εντός αυτών, μέσω του δείκτη CPA (Collagen Proportionate Area).

Οι **στόχοι** της εργασίας είναι:

1. Η παρουσίαση του θεωρητικού πλαισίου που αφορά στις βιοψίες του παχέος εντέρου και τον καρκίνο του παχέος εντέρου, την Τεχνητή Νοημοσύνη, τα Τεχνητά Νευρωνικά Δίκτυα και τα Συνελκτικά Νευρωνικά Δίκτυα.
2. Η παρουσίαση των βασικών εφαρμογών των Συνελκτικών Νευρωνικών Δικτύων στην βιοϊατρική εικόνα και ειδικότερα στις βιοψίες του παχέος εντέρου.

3. Η ανάπτυξη και εφαρμογή μεθοδολογίας για την αυτόματη τμηματοποίηση περιοχών (κλάσεων) σε εικόνες από βιοψίες του παχέος εντέρου.
4. Η ποσοτική ανάλυση του ποσοστού κολλαγόνου σε ιστολογικές περιοχές, μέσω χρωματικής ομαδοποίησης με τη χρήση του αλγορίθμου k-means, και ο υπολογισμός του δείκτη CPA.
5. Η αξιολόγηση της προτεινόμενης μεθοδολογίας σχετικά με την ικανότητά της να χρησιμοποιηθεί, με τέτοιο τρόπο ώστε να υποστηρίζει με αντικειμενικό τρόπο και αξιόπιστα, στην ανάλυση ιστοπαθολογικών εικόνων.

Η εργασία δομείται σε έξι κεφάλαια. Στο **πρώτο κεφάλαιο** παρουσιάζεται το θεωρητικό υπόβαθρο σχετικά με τις βιοψίες του παχέος εντέρου και τον καρκίνο του παχέος εντέρου. Στο **δεύτερο κεφάλαιο** παρουσιάζονται βασικές έννοιες σχετικά με την Τεχνητή Νοημοσύνη, τους τομείς εφαρμογής της, τον τρόπο με τον οποίο έχει συμβάλει στον τομέα της ιατρικής, καθώς επίσης παρουσιάζονται και οι θεμελιώδεις αρχές της Μηχανικής Μάθησης. Στο **τρίτο κεφάλαιο** παρουσιάζεται η βασική δομή της λειτουργίας των Τεχνητών νευρωνικών Δικτύων, όπως και η διαδικασία μάθησης και εκπαίδευσής τους. Στο **τέταρτο κεφάλαιο** παρουσιάζονται τα Συνελκτικά Νευρωνικά Δίκτυα, ο ρόλος τους στο πεδίο της Τεχνητής Νοημοσύνης, τα δομικά τους στοιχεία, η εξέλιξή τους, η υπολογιστική τους λογική, οι τεχνικές βελτίωσης και συγκεκριμένες αρχιτεκτονικές CNNs, που σχετίζονται με την παρούσα εργασία, όπως είναι η Resnet. Στο **πέμπτο κεφάλαιο** εξετάζονται οι εφαρμογές των CNNs στην βιοϊατρική εικόνα και ιδιαίτερα η χρήση τους στην ιατρική διάγνωση στην ιστοπαθολογία. Παρουσιάζονται επίσης χαρακτηριστικά παραδείγματα εφαρμογών τους, όπως είναι η ταξινόμηση περιοχών ιστού με χρήση Res Net και ο εντοπισμός ευρημάτων σε βιοψίες του παχέος εντέρου με χρήση αρχιτεκτονικών R-CNN. Στο **έκτο κεφάλαιο** παρουσιάζεται το πειραματικό τμήμα της εργασίας, στο οποίο γίνεται ανάλυση ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου με χρήση του μοντέλου Mask R-CNN, όπου γίνεται αρχικά τμηματοποίηση ιστολογικών περιοχών και στη συνέχεια υπολογισμός του ποσοστού κολλαγόνου εντός των περιοχών, με χρήση χρωματικής ομαδοποίησης που έγινε μέσω του αλγορίθμου k-means. Τέλος, παρουσιάζονται η αξιολόγηση του μοντέλου, τα αποτελέσματα, τα συμπεράσματα, καθώς και προτάσεις για περαιτέρω έρευνα.

Κεφάλαιο 1ο

1. Θεωρητικό Υπόβαθρο

Η ιστοπαθολογική εξέταση βιοψιών του παχέος εντέρου παίζει σημαντικό ρόλο στην αναγνώριση και αξιολόγηση ευρημάτων που σχετίζονται κυρίως με φλεγμονές, προκαρκινικές βλάβες ή νεοπλασματικές ασθένειες όπως και με εκφυλιστικές αλλοιώσεις. Η μελέτη των ιστών του τοιχώματος του παχέος εντέρου επιτρέπει την εκτίμηση των μορφολογικών χαρακτηριστικών και την κατηγοριοποίηση των αλλοιώσεων, συμβάλλοντας έτσι στην έγκαιρη διάγνωση και την επιλογή της κατάλληλης αντιμετώπισης[1]. Ως εκ τούτου, η συστηματική μελέτη των ευρημάτων σε βιοψίες του παχέος εντέρου αποτελεί σημαντικό πεδίο ενδιαφέροντος για τους παθολογοανατόμους αλλά και για την ανάπτυξη και εφαρμογή σύγχρονων τεχνικών ανάλυσης ιατρικής εικόνας[2].

Στα παθολογικά ευρήματα που εντοπίζονται κυρίως στις βιοψίες του παχέος εντέρου, μεγαλύτερο ενδιαφέρον παρουσιάζουν οι προκαρκινικές και νεοπλασματικές αλλοιώσεις, αφού κάποιες από αυτές μπορεί να είναι πρώιμες βλάβες οι οποίες δυνητικά θα εξελιχθούν σε καρκίνωμα. Η έγκαιρη ιστοπαθολογική εξέταση και η αξιολόγηση των αλλοιώσεων αυτών συμβάλλει στον διαχωρισμό καλοηθών, από τις δυσπλασίες και κακοήθεις μορφές και στον έγκαιρο εντοπισμό βλαβών που αφορούν είτε σε αδενώματα – καρκινώματα, είτε σε καρκινογένεση. Για το λόγο αυτό, η μελέτη του καρκίνου του παχέος εντέρου έχει ιδιαίτερο ενδιαφέρον, μιας και διεθνώς αποτελεί μια από τις συχνότερες κακοήθειες όπου η έγκαιρη ανίχνευση βλαβών και η αντιμετώπισή τους βοηθάει σημαντικά στην μείωση των επιπτώσεων αλλά και της θνησιμότητας[3].

1.1. Η ασθένεια του καρκίνου του παχέος εντέρου

Ο καρκίνος του παχέος εντέρου αποτελεί ένα από τα πιο κοινά νεοπλάσματα, είναι ο δεύτερος σε συχνότητα εμφάνισης καρκίνος στον ανδρικό πληθυσμό και ο τρίτος στον γυναικείο πληθυσμό παγκοσμίως. Σύμφωνα με στοιχεία του Παγκόσμιου Οργανισμού Υγείας για το 2020, είναι ο τρίτος σε διάγνωση και ο δεύτερος σε θνησιμότητα. Ειδικότερα, υπήρξαν πάνω από 1,9 εκατομμύρια νέες περιπτώσεις καρκίνου του παχέος εντέρου, ενώ έχουν καταγραφεί πάνω από 930 χιλιάδες θάνατοι εξαιτίας αυτού[4]. Τα ποσοστά αυτά δεν διαφοροποιήθηκαν ιδιαίτερα το 2022 αφού, σύμφωνα με τον Παγκόσμιο Οργανισμό υγείας, σημειώθηκαν 1.926.425 νέα περιστατικά και 904.019 θάνατοι.[5] Τα υψηλότερα ποσοστά καρκίνου του παχέος εντέρου σημειώθηκαν στην Ευρώπη, την Αυστραλία και την Νέα Ζηλανδία, ενώ τα ποσοστά θνησιμότητα ήταν μεγαλύτερα στην Ανατολική Ευρώπη. Τα δεδομένα αυτά διαφοροποιούνται ανάλογα με την κοινωνικοοικονομική ανάπτυξη των χωρών[6]. Τα ποσοστά εμφανίζονται μειωμένα σε χώρες με υψηλό βιοτικό επίπεδο, λόγω της ύπαρξης προγραμμάτων πρωτοβάθμιας περίθαλψης στα πλαίσια προσυμπτωματικού ελέγχου. Τέλος, οι εκτιμήσεις δείχνουν ότι μέχρι το 2040 οι νέες περιπτώσεις καρκίνου του παχέος εντέρου θα αυξηθούν σε 63%, με 3,2 εκατομμύρια νέους ασθενείς ετησίως, ενώ θα παρατηρηθεί και αύξηση του ποσοστού θνησιμότητας σε 73% με 1,6 εκατομμύρια θανάτους ετησίως[4].

Οι παράγοντες που συμβάλλουν στην αύξηση της πιθανότητας εκδήλωσης της νόσου, καθώς και στην θνησιμότητα από αυτή, συνήθως αποτελούν έναν συνδυασμό από γενετικούς και περιβαλλοντικούς παράγοντες. Το οικογενειακό ιστορικό, κληρονομικοί και γενετικοί παράγοντες, η ηλικία, η υιοθέτηση συνηθειών του σύγχρονου τρόπου ζωής όπως οι διατροφικές συνήθειες, το αυξημένο σωματικό βάρος, η έλλειψη σωματικής δραστηριότητας, το κάπνισμα, αφορούν κάποιους από τους παράγοντες αυτούς[7].

Στις περισσότερες περιπτώσεις ο καρκίνος του παχέος εντέρου ξεκινάει από την εμφάνιση μικρών καλοθών μαζών κυττάρων, πολύποδες, όπου με το πέρασμα του χρόνου εξελίσσονται σε κακοήθεις όγκους με μεγάλη πιθανότητα να δημιουργηθούν μεταστάσεις και σε άλλα όργανα, κυρίως στο ήπαρ [8] Ως εκ τούτου, ο προσυμπτωματικός έλεγχος είναι ζωτικής σημασίας, καθώς συμβάλλει σημαντικά στην μη εξέλιξη προ καρκινικών πολυπόδων σε καρκίνο. Επιπλέον, η ανίχνευση του καρκίνου του παχέος εντέρου, και η αφαίρεσή του σε πρώιμο στάδιο, συμβάλλει σημαντικά στην αντιμετώπισή του και στην θεραπεία[9].

1.2. Αίτια ανάπτυξης της νόσου

Η πιθανότητα εμφάνισης της νόσου οφείλεται σε μία ποικιλία παραγόντων, τόσο γενετικών, όσο και περιβαλλοντικών[10]. Κληρονομικά σύνδρομα, γενετικές διαταραχές, οικογενειακό ιστορικό, καθώς και οι παθήσεις του εντέρου, όπως πολύποδες, επηρεάζουν σημαντικά στην εμφάνιση της νόσου. Από την άλλη, η διατροφή που βασίζεται σε αυξημένη κατανάλωση κόκκινου κρέατος και επεξεργασμένων τροφών, καθώς και η έλλειψη σωματικής δραστηριότητας αποτελούν επίσης παράγοντα κινδύνου για την εμφάνιση της νόσου.

1.2.1. Οικογενειακό Ιατρικό Ιστορικό και Γενετική

Έχει αποδειχθεί ότι η κληρονομική και γενετική προδιάθεση αυξάνουν τον κίνδυνο εμφάνισης καρκίνου του παχέος εντέρου. Η πιθανότητα εμφάνισης της νόσου, σχετίζεται με την ανάπτυξη καρκίνου του παχέος εντέρου σε συγγενείς πρώτου βαθμού, καθώς και με την ηλικία κατά την οποία εμφάνισαν τη νόσο. Ρόλο επίσης παίζει και η συνεμφάνιση άλλων τύπων καρκίνου στο οικογενειακό περιβάλλον[11]. Μελέτες έχουν δείξει ότι η ύπαρξη συγγενή πρώτου βαθμού με την νόσο, αυξάνει έως και δύο φορές κατά μέσο όρο την πιθανότητα εμφάνισης της νόσου σε κάποιο άτομο, σε σχέση με όσους δεν έχουν οικογενειακό ιστορικό [12] [13].

Κληρονομικά σύνδρομα που προέρχονται από την παρουσία γενετικών μεταλλάξεων, έχουν επίσης συνδεθεί με την εμφάνιση καρκίνου του παχέος εντέρου. Τα πιο γνωστά είναι το σύνδρομο Lynch και η οικογενής αδενωματούδης πολυποδίαση (FAP). Το σύνδρομο Lynch, γνωστό και ως μη-πολυποδιακός ορθοκολικός καρκίνος – Hereditary nonpolyposis colorectal cancer (HNPCC), το οποίο οφείλεται σε μεταλλάξεις γονιδίων που είναι υπεύθυνα για την επιδιόρθωση λαθών στο DNA. Η πιθανότητα να κληρονομήσουν τα παιδιά από τους γονείς αυτές τις μεταλλάξεις, είναι 50% [11].

Η Οικογενής Αδενωματούδης Πολυποδίαση – Familial Adenomatous Polyposis (FAP) είναι ένα κληρονομικό σύνδρομο που ευθύνεται για το 1% των περιπτώσεων καρκίνου του παχέος εντέρου. Το συγκεκριμένο σύνδρομο προκαλεί μία μετάλλαξη στο ογκοκατασταλτικό γονίδιο APC. Τα παιδιά ασθενών με το σύνδρομο αυτό, έχουν 50% πιθανότητες να το κληρονομήσουν[14][15].

1.2.2. Προσωπικό Ιατρικό Ιστορικό

Άτομα με ιστορικό κάποιου τύπου καρκίνου, πολύποδες του παχέος εντέρου, φλεγμονώδεις νόσους του εντέρου καθώς και σακχαρώδη διαβήτη διατρέχουν αυξημένο κίνδυνο εμφάνισης καρκίνου του παχέος εντέρου.

Η φλεγμονώδης νόσος του εντέρου αφορά σε διαταραχές που δημιουργούν χρόνια και ανεξέλεγκτη φλεγμονή στο γαστρεντερικό σύστημα. Οι πιο κοινές μορφές φλεγμονώδους νόσου του εντέρου είναι η νόσος του Crohn και η Ελκώδης Κολίτιδα. Λόγω του γεγονότος ότι η χρόνια φλεγμονή εξελίσσεται στην ανάπτυξη όγκου, το άτομο που νοσεί έχει δύο έως έξι φορές περισσότερες πιθανότητες να αναπτύξει καρκίνο του παχέος εντέρου, σε σχέση με υγιή άτομα [11] [16] [17] [18].

Οι πολύποδες του παχέος εντέρου ταξινομούνται σε δύο βασικές κατηγορίες: μη νεοπλασματικοί και νεοπλασματικοί – αδενωματώδεις πολύποδες. Οι αδενωματώδεις πολύποδες δυνητικά αποτελούν προκαρκινικές βλάβες, χωρίς αυτό να σημαίνει ότι εξελίσσονται όλοι σε κακοήθειες. Η εξέλιξη του αδενωματώδη πολύποδα σε καρκίνο του παχέος εντέρου εξαρτάται από την ηλικία, τα χρόνια παραμονής στον οργανισμό (10 – 15 χρόνια), το μέγεθός του καθώς και τον βαθμό δυσπλασίας. Έρευνες έχουν δείξει ότι το 40% των ατόμων ηλικίας 50 ετών και άνω εμφανίζουν αδενωματώδεις πολύποδες και άρα είναι υψίστης σημασίας ο έγκαιρος εντοπισμός και η αφαίρεσή τους, ώστε να περιοριστεί η πιθανότητα να εξελιχθούν σε καρκίνο του παχέος εντέρου [11] [16] [17].

Ιατρικά δεδομένα δείχνουν ότι ο σακχαρώδης διαβήτης αποτελεί έναν ανεξάρτητο παράγοντα κινδύνου για καρκίνους του γαστρεντερικού και άρα και για καρκίνο του παχέος εντέρου[19]. Μελέτες έχουν δείξει ότι τα διαβητικά άτομα έχουν δύο έως και τρεις φορές περισσότερες πιθανότητες να αναπτύξουν καρκίνο του παχέος εντέρου, σε σχέση με τα μη διαβητικά άτομα[20]. Ο αριθμός των ατόμων με διαβήτη το 2019 ανέρχονταν στα 463 εκατομμύρια παγκοσμίως, ενώ όπως αναμένεται, το 2030 θα αυξηθεί σε 578 εκατομμύρια και το 2045 σε 700 εκατομμύρια[21]. Η υπερινσουλιαιμία, μέσω μιας φλεγμονώδης κατάστασης που δημιουργεί και η οποία πυροδοτεί τον πολλαπλασιασμό των κυττάρων του παχέος εντέρου, συμβάλλει άμεσα στην καρκινογένεση του παχέος εντέρου [19].

1.2.3. Τρόπος ζωής – Διατροφικοί παράγοντες

Η διατροφή αναμφίβολα αποτελεί τον βασικότερο εξωγενή παράγοντα, που έχει εντοπιστεί ως σήμερα, ως αιτία ανάπτυξης καρκίνου του παχέος εντέρου. Μελέτες δείχνουν ότι, σημαντικός παράγοντας κινδύνου για την ανάπτυξη καρκίνου του παχέος εντέρου, είναι η συχνή κατανάλωση κόκκινου και επεξεργασμένου κρέατος [11] [22]. Επιπλέον, έχει αποδειχθεί ότι η κατανάλωση τροφών πλούσιων σε φυτικές ίνες, ενδεχομένως να μειώνει τον κίνδυνο εμφάνισης καρκίνου του παχέος εντέρου έως και 50%. Παρόλο που δεν έχει τεκμηριωθεί πλήρως ο ακριβής τρόπος αντικαρκινικής δράσης των φυτικών ινών που βρίσκονται στα φρούτα και τα λαχανικά, αποτελέσματα μελετών δείχνουν ότι η συχνή κατανάλωση λαχανικών και φρούτων, τα οποία μέσω των βιοδραστικών ενώσεων, των βιταμινών, των μετάλλων και ιχνοστοιχείων που περιέχουν, προσφέρουν στον οργανισμό ισχυρές αντιφλεγμονώδεις και αντιοξειδωτικές ιδιότητες, που ενδεχομένως αναστέλλουν την κυτταρική βλάβη [23] [11] [24] [18].

Σύμφωνα με το Αμερικανικό Ινστιτούτο για τον καρκίνο, η κατανάλωση γαλακτοκομικών προϊόντων ίσως συμβάλλει θετικά στην μείωση του κινδύνου ανάπτυξης καρκίνου του παχέος εντέρου[23]. Η πιθανή θετική συμβολή των γαλακτοκομικών προϊόντων, οφείλεται κυρίως στο ασβέστιο που περιέχουν, έχει αποδειχθεί ότι το ασβέστιο δεσμεύει χολικά και λιπαρά οξέα με τέτοιο τρόπο, ώστε να ελαττώνει την πιθανότητα τροποποίησης της βλεννογόνου του εντέρου, περιορίζοντας έτσι την πιθανότητα καρκινογένεσης. Σε αυτό βοηθάει επίσης και το γεγονός ότι έχει βρεθεί ότι το ασβέστιο συμβάλλει στην αναστολή πολλαπλασιασμού των καρκινικών κυττάρων[11] [25] [26]. Επιπλέον, και η βιταμίνη D, που περιέχεται στα γαλακτοκομικά, φαίνεται να συμβάλλει θετικά στην προστασία του οργανισμού από την ανάπτυξη καρκίνου του παχέος εντέρου. Το ασβέστιο συνδέεται άμεσα με την βιταμίνη D, καθώς η κύρια λειτουργία της είναι η διατήρηση της ομοιόστασης του ασβεστίου, ενισχύοντας έτσι την απορρόφησή της από το έντερο. Μελέτες έχουν δείξει ότι η βιταμίνη D επηρεάζει πολλές φυσιολογικές λειτουργίες του οργανισμού οι οποίες είναι μεγάλης σημασίας για τον έλεγχο του καρκίνου [27]

1.2.4. Άλλοι Παράγοντες

Η ηλικία αποτελεί έναν από τους σημαντικότερους παράγοντες ανάπτυξης καρκίνου του παχέος εντέρου, αφού κατά 90% περίπου των περιπτώσεων εμφάνιση της νόσου αφορά κυρίως άτομα άνω των 50 ετών[28], [29]. Η πιθανότητα ανάπτυξης καρκίνου του παχέος εντέρου σε άτομα άνω των 65 ετών είναι κατά τρεις φορές μεγαλύτερη σε σχέση με άτομα ηλικίας 50 έως 64 ετών., με την μέση ηλικία εμφάνισης της νόσου να είναι τα 68 έτη για τους άνδρες και τα 72 έτη για τις γυναίκες[30].

Το κάπνισμα επίσης, αποτελεί σημαντικό παράγοντα κινδύνου για την εμφάνιση διαφόρων τύπων καρκίνων όπως και για την ανάπτυξη καρκίνου του παχέος εντέρου. Έρευνες έδειξαν ότι οι καπνιστές διατρέχουν 2 έως και 3 φορές μεγαλύτερο κίνδυνο εμφάνισης καρκίνου του παχέος εντέρου, σε σχέση με τους μη καπνιστές, ενώ σε αυτό οφείλεται το 12% των θανάτων από καρκίνο αυτού του τύπου[31]. Το κάπνισμα σχετίζεται θετικά με την ανάπτυξη μεγάλων ορθοκολικών αδενωμάτων, τα οποία αποτελούν βλάβες που εξελίσσονται σε καρκίνο στις περισσότερες περιπτώσεις [8]. Πάνω από 60 χημικές ουσίες, από τις χιλιάδες που περιέχει ο καπνός του τσιγάρου, χαρακτηρίζονται ως καρκινογόνες και βλάπτουν το DNA [32].

Μελέτες δείχνουν ότι η κατανάλωση αλκοόλ, αποτελεί σταθερά παράγοντα υψηλού κινδύνου για την εμφάνιση καρκίνου του παχέος εντέρου[33]. Ο κίνδυνος είναι αυξημένος, κατά 20% περίπου, στα άτομα που καταναλώνουν δύο έως τρία ποτά σε καθημερινή βάση, ενώ όταν η ποσότητα είναι πάνω από τρία ποτά ημερησίως, το ποσοστό αυτό αυξάνεται σε 40% και σε 52% αντίστοιχα όταν η κατανάλωση φτάνει στα τέσσερα ή και πέντε ποτά καθημερινά[24], [30], [34] [35].

Επιδημιολογικά στοιχεία δείχνουν ότι η έλλειψη σωματικής δραστηριότητας συνδέεται με την αυξανόμενη εμφάνιση καρκίνου του παχέος εντέρου, ειδικά στους πληθυσμούς των ανεπτυγμένων χωρών, όπου ο καθιστικός τρόπος ζωής αποτελεί κύριο χαρακτηριστικό τους. Έχει αποδειχθεί ότι η συχνή σωματική άσκηση λειτουργεί ως προστασία απέναντι στην ανάπτυξη καρκίνου του παχέος εντέρου, μιας και τα οφέλη της σχετίζονται με την ενίσχυση του ανοσοποιητικού συστήματος, την διατήρηση φυσιολογικού σωματικού βάρους, την μείωση της φλεγμονής και του στρες και της ρύθμισης των ορμονών του οργανισμού[36]. Οι άνθρωποι που δεν έχουν εντάξει την σωματική άσκηση στην καθημερινότητά τους έχουν κατά 50% μεγαλύτερη πιθανότητα

να αναπτύξουν καρκίνο του παχέος εντέρου, σε σχέση με όσους ασκούνται τακτικά. [11], [34], [37].

1.3. Συμπτώματα του καρκίνου του παχέος εντέρου

Λόγω του γεγονότος ότι ο καρκίνος του παχέος εντέρου σε πρώιμο στάδιο δεν εμφανίζει συμπτώματα, ή αυτά δεν είναι άμεσα αντιληπτά, ο προληπτικός έλεγχος είναι υψίστης σημασίας. Τα συμπτώματα ποικίλλουν και είναι ανάλογα με το μέγεθος του όγκου, την θέση του στο έντερο και το μέγεθος της αλλοίωσης που έχει προκαλέσει. Ανάμεσα στα συμπτώματα που εμφανίζονται σε πρώιμο στάδιο της νόσου, είναι το κοιλιακό άλγος, η αλλαγή στις συνήθειες του εντέρου, η απώλεια βάρους χωρίς αιτία και η κόπωση. Σε μεταγενέστερα στάδια της νόσου, μπορεί να είναι η απόφραξη του εντέρου, επίμονη δυσκοιλιότητα, αίμα στα κόπρανα και έντονη απώλεια βάρους.

1.4. Στάδια του καρκίνου του παχέος εντέρου

Ο καρκίνος του παχέος εντέρου είναι ένας από τους πιο ιάσιμους καρκίνους με την προϋπόθεση να έχει διαγνωσθεί εγκαίρως. Η έκβαση της νόσου εξαρτάται από το στάδιο στο οποίο γίνεται η διάγνωση με δεδομένο ότι συμπτώματα εμφανίζονται συνήθως σε μεταγενέστερο στάδιο. Η ακριβής σταδιοποίηση είναι μεγίστης σημασίας για την επιλογή της κατάλληλης θεραπείας και την παρακολούθηση της ενδεχόμενης εξέλιξης της νόσου[38].

Για την εκτίμηση και αξιολόγηση της νόσου χρησιμοποιούνται διάφορα κριτήρια, ενώ η μέθοδος σταδιοποίησης που χρησιμοποιείται κατά κύριο λόγο βασίζεται στο σύστημα TNM (Tumor / Node / Metastasis), σύμφωνα με την Διεθνή Ένωση Καρκίνου (UICC) και την Αμερικανική Κοινή Ένωση για τον Καρκίνο (AJCC).[38] Το TNM είναι ένα σύστημα που χρησιμοποιείται για την αξιολόγηση της εμφάνισης και του μεγέθους του όγκου (Tumor), την εξάπλωση σε λεμφαδένες (Node), και της μετάστασης σε άλλα όργανα όπως το ήπαρ ή οι πνεύμονες (Metastasis).

TNM Σύστημα Σταδιοποίησης:

T (Tuma) - Όγκος

T0: δείχνει ότι δεν υπάρχουν ενδείξεις όγκου

Tis: καρκίνωμα in situ (αποκλειστικά στον βλεννογόνο)

T1: επέκταση της διήθησης ως τον υποβλεννογόνο χιτώνα

T2: επέκταση της διήθησης ως τον μυϊκό χιτώνα

T3: επέκταση της διήθησης ως τον ορογόνο (εξωτερικό προστατευτικό περίβλημα του γαστρεντερικού σωλήνα) ή όπου δεν υπάρχει ορογόνος προς τους περικολικούς ή περιορθικούς ιστούς

T4: ο όγκος διηθεί κατ' επέκταση άλλα όργανα ή διηθεί το σπλαχνικό περιτόναιο (λεπτός υμένας ο οποίος αποτελείται από δύο πέταλα: το τοιχωματικό που επενδύει το κοιλιακό τοίχωμα και το σπλαχνικό που επενδύει τα σπλάγχνα)

N (Nodes) – Λεμφαδένες

N0: δεν υπάρχει μετάσταση σε επιχώριους (τοπικούς) λεμφαδένες

N1: μετάσταση σε 1 έως 3 επιχώριους λεμφαδένες

N2: μετάσταση σε 4 ή περισσότερους επιχώριους λεμφαδένες.

M (Metastasis) – Μεταστάσεις

M0: δεν υπάρχουν απομακρυσμένες μεταστάσεις

M1: υπάρχουν απομακρυσμένες μεταστάσεις [39].

Στάδιο 0:

Πρόκειται για καρκίνο σε πρώιμο στάδιο, βρίσκεται στον βλεννογόνο (εσωτερικό στρώμα του παχέος εντέρου), ενδεχομένως να επεκταθεί και σε γειτονικούς υγιείς ιστούς. Ονομάζεται και ενδοβλεννογόνο καρκίνωμα ή καρκίνος in situ. [8], [40].

Στάδιο I T1 N0M0 – T2 N0M0:

Ο καρκίνος έχει επεκταθεί στον υποβλεννογόνο, μέσω του βλεννογόνου, ή και στον μυϊκό χιτώνα του παχέος εντέρου, χωρίς όμως να έχει προχωρήσει σε τοπικούς λεμφαδένες.

Στάδιο ΙΑ - T3 N0 M0:

Ο καρκίνος επεκτείνεται στον ορογόνο χιτώνα, μέσω του μυϊκού στρώματος του τοιχώματος του παχέος εντέρου, χωρίς να έχει επεκταθεί σε επιχώριους λεμφαδένες ή απομακρυσμένες περιοχές.

Στάδιο ΙΙΒ – T4a N0 M0:

Σε αυτό το στάδιο ο καρκίνος έχει διηθήσει στο σπλαχνικό περιτόναιο και έχει προχωρήσει και στο εξωτερικό του παχέος εντέρου σε γειτονικούς ιστούς ή και όργανα. Δεν έχει επεκταθεί ακόμη σε επιχώριους λεμφαδένες ή σε απομακρυσμένα όργανα.

Στάδιο ΙΙΙΑ – T1 – 2 N1 M0:

Ο καρκίνος έχει επεκταθεί προς τον υποβλεννογόνιο, μέσω του βλεννογόνου τοιχώματος του παχέος εντέρου και έχει εξαπλωθεί σε 1 έως 3 γειτονικούς λεμφαδένες ή και σε 4 έως 6 γειτονικούς λεμφαδένες.

Στάδιο ΙΙΙΒ:

Ο καρκίνος έχει επεκταθεί μέσω του μυϊκού στρώματος του τοιχώματος του παχέος εντέρου στο σπλαχνικό περιτόναιο ή και στον ιστό που περιλαμβάνει τα όργανα της κοιλιάς

Στάδιο ΙΙΙC:

Ο καρκίνος έχει επεκταθεί στο εξωτερικό στρώμα του παχέος εντέρου (σπλαχνικό περιτόναιο) και σε τουλάχιστον 4 έως 6 λεμφαδένες

ή

μέσω του ορογόνου έχει εξαπλωθεί σε γειτονικά όργανα ή σε ιστό κοντά στους λεμφαδένες.

Στάδιο ΙV:

Ο καρκίνος έχει κάνει μετάσταση σε άλλα όργανα του σώματος όπως το ήπαρ, οι πνεύμονες, τα οστά και οι ωοθήκες [41].

Η σταδιοποίηση του καρκίνου συνδέεται με τη σοβαρότητα της νόσου καθώς παρατηρείται ότι τα ποσοστά επιβίωσης μειωμένα ιδιαίτερα στα μεταγενέστερα στάδια.

Συγκεκριμένα, έρευνες έχουν δείξει ότι ο καρκίνος του παχέος εντέρου έχει ποσοστό επιβίωσης σε βάθος πενταετίας 74% για το στάδιο I αλλά μόνο 5% για το στάδιο IV. [42]

1.5. Διάγνωση

Στις περισσότερες περιπτώσεις ο καρκίνος του παχέος εντέρου χρειάζεται χρόνια για να αναπτυχθεί και συνήθως δεν παρουσιάζονται εμφανή συμπτώματα, άρα ο προγνωστικός έλεγχος είναι υψίστης σημασίας μιας και συμβάλλει στην μείωση της θνησιμότητας[8]. Μείωση της θνησιμότητας επιτυγχάνεται συνήθως με την έγκαιρη ανίχνευση, μέσω του προσυμπτωματικού ελέγχου και της αφαίρεσης του καρκίνου του παχέος εντέρου ή μέσω της ανίχνευσης και αφαίρεσης προκαρκινικών βλαβών όπως είναι τα αδενώματα και οι πολύποδες του παχέος εντέρου[43] [44] [45].

Ο προγνωστικός έλεγχος βασίζεται κυρίως σε εξετάσεις με βάση τα κόπρανα, ενδοσκοπικές και απεικονιστικές εξετάσεις[46]. Συγκεκριμένα, στις εξετάσεις προγνωστικού ελέγχου περιλαμβάνονται οι παρακάτω εξετάσεις:

Ανοσοχημική εξέταση κοπράνων για αιμοσφαιρίνη (FIT): χρησιμοποιείται για την διερεύνηση αιμορραγίας σε τμήματα του γαστρεντερικού συστήματος.

Εξέταση κρυφού αίματος κοπράνων (FOBT): αποτελεί εργαστηριακή εξέταση για τον έλεγχο δειγμάτων κοπράνων για κρυφό αίμα, η ύπαρξη του οποίου ενδεχομένως υποδηλώνει την ύπαρξη καρκίνου του παχέος εντέρου ή πολυπόδων[47].

Οπτική κολonosκόπηση (OC): βασική εξέταση που πραγματοποιείται για την πρόληψη αλλά και την διάγνωση καρκίνου του παχέος εντέρου αλλά και την ανίχνευση πολυπόδων στο έντερο. Χρησιμοποιείται για έλεγχο του βλεννογόνου του εντέρου, προληπτικά στις ηλικίες άνω των 50 ετών, αλλά και σε περιπτώσεις εμφάνισης καρκίνου του παχέος εντέρου. Σύμφωνα με μελέτη που πραγματοποιήθηκε, τα αποτελέσματα έδειξαν ότι υπάρχει συσχετισμός της κολonosκόπησης προσυμπτωματικού ελέγχου, με αισθητά μειωμένο κίνδυνο θνησιμότητας. Συγκεκριμένα, συσχετίστηκε με μείωση κατά 67% της θνησιμότητας για καρκίνους του δεξιού παχέος εντέρου και 75% για καρκίνους του αριστερού παχέος εντέρου[48].

Ευέλικτη σιγμοειδοσκόπηση (FS): αποτελεί διαγνωστική εξέταση με την οποία διερευνάται το κατώτερο τμήμα του παχέος εντέρου. Συνιστάται σε περιπτώσεις όπου υπάρχουν συμπτώματα όπως διάρροια ή δυσκοιλιότητα, αιμορραγία, αλλά και στην περίπτωση ύπαρξης πολυπόδων ή αιμορροϊδων. Γίνεται με σκοπό να διαγνωστεί ή να

επιβεβαιωθεί καρκίνος του παχέος εντέρου, να εντοπιστούν πολύποδες ή αιμορροΐδες ώστε να ληφθεί βιοψία για την ιστολογική εξέταση για την αφαίρεση αυτών (πολύποδες και αιμορροΐδες)[47].

Ψηφιακή ορθική εξέταση (DRE): έλεγχος που χρησιμοποιείται για την εξέταση του κάτω όρθου και άλλων εσωτερικών οργάνων. Χρησιμοποιείται για τον έλεγχο της αιμορραγίας του γαστρεντερικού σωλήνα ή του καρκίνου του παχέος εντέρου [49].

1.6. Θεραπεία

Ο καρκίνος του παχέος εντέρου είναι ένας από τους πιο θεραπεύσιμους καρκίνους, ειδικά εάν διαγνωσθεί εγκαίρως. Παρόλο που το ποσοστό εμφάνισης της νόσου είναι αυξημένο, ιδιαίτερα στις αναπτυγμένες δυτικές κοινωνίες, παρατηρείται μείωση θανάτων καθώς και αύξηση του ποσοστού πενταετούς επιβίωσης (διπλασιάστηκε). Συγκεκριμένα, στις αναπτυγμένες χώρες η πενταετής επιβίωση φτάνει στο 50% ακόμη και στο 60% [50].

Βασικός παράγοντας για την επιλογή της κατάλληλης θεραπείας και την έκβαση της νόσου, είναι το στάδιο στο οποίο έχει διαγνωστεί[38]. Το ποσοστό πενταετούς επιβίωσης, σύμφωνα με το σύστημα σταδιοποίησης TNM, σε ασθενείς με καρκίνο σταδίου I είναι 93%, ενώ στο στάδιο IIA φτάνει στο 85%. Αυτό οφείλεται στο γεγονός ότι ο καρκίνος του παχέος εντέρου αποτελεί ένα μόρφωμα που αναπτύσσεται τοπικά και ως εκ τούτου είναι υψίστης σημασίας σε άτομα υψηλού κινδύνου, να επιτευχθεί η έγκαιρη διάγνωση της νόσου[50].

Η μόνη θεραπευτική μέθοδος που μπορεί να επιφέρει ίαση σε διαγνωσμένο καρκίνο του παχέος εντέρου είναι η χειρουργική εκτομή[51]. Όταν διαγνωστεί καρκίνος του παχέος εντέρου, πρέπει να γίνεται σταδιοποίησή του. Επιπλέον, πρέπει να γίνεται έλεγχος του ήπατος και του θώρακα με αξονική τομογραφία, ώστε να γίνει ανίχνευση τυχόν μεταστάσεων του καρκίνου[38]. Η επικουρική χημειοθεραπεία, έπειτα από την εκτομή, εφαρμόζεται σε καρκίνο του παχέος εντέρου σταδίου III με σκοπό την εξάλειψη μικρομεταστάσεων ενώ τα οφέλη της είναι εμφανή και στην εξάπλωση της νόσου στους επιχώριους λεμφαδένες[50].

Κατά το στάδιο I η νόσος αντιμετωπίζεται με χειρουργική επέμβαση (εκτομή του όγκου και των τοπικών λεμφαδένων).[38] Ο καθαρισμός των τοπικών λεμφαδένων παίζει σημαντικό ρόλο, τόσο για θεραπευτικούς, όσο και για προγνωστικούς λόγους, λόγω του

γεγονότος ότι ο αριθμός των λεμφαδένων στους οποίους έχει γίνει διήθηση συνδέεται με την επιβίωση του ασθενούς αλλά και με την αναγκαιότητα χορήγησης επικουρικής χημειοθεραπείας [52]. Σύμφωνα με Αμερικανική Κοινή Ένωση για τον Καρκίνο (AJCC) και το Διεθνές Φόρουμ Ποιότητας (National Quality Forum), η αφαίρεση 12 λεμφαδένων θεωρείται απαραίτητη για να αξιολογηθεί η κατάστασή τους[53].

Οι βλάβες T1 είναι κακοήθεις πολύποδες όπου τα καρκινικά κύτταρα έχουν εισβάλει στο υποβλενογόνιο, και αποτελούν το 12% των πολυπόδων[38]. Ακόμη και στην περίπτωση που στην ενδοσκόπηση φαίνονται καλοήθεις, αν στην ιστοπαθολογική εξέταση ανιχνευθεί κακοήθης πολύποδας, πρέπει να εκτιμηθεί και να αποφασιστεί αν η ενδοσκοπική εκτομή αρκεί ή απαιτείται ενδοσκοπική βλενογονική εκτομή EMR (Endoscopic Submucosal Resection), ή Ενδοσκοπική υποβλενογονική ανατομή ESD (Endoscopic Submucosal Dissection) ή τμηματική εκτομή του παχέος εντέρου[38].

Ο κακοήθης πολύποδας θεραπεύεται με πολυεκτομή snare (snare polypectomy) ενδοσκοπικά. Όταν το όριο εκτομής είναι 2mm ή περισσότερο ή η ιστολογική δεν είναι καθόλου διαφοροποιημένη και δεν έχει διηθήσει στους λεμφαδένες[54].

Όταν το όριο εκτομής είναι μικρότερο από 2mm, η ιστολογική εξέταση είναι διαφοροποιημένη πολύ λίγο και έχει γίνει λεμφαγγειακή διήθηση, ενδείκνυται η τμηματική εκτομή του εντέρου. Αν δεν επιτρέπεται να υποβληθεί ο ασθενής σε χειρουργική επέμβαση, τότε η Ενδοσκοπική Εκτομή του Βλεννογόνου (EMR) είναι κατάλληλη για αλλοίωση $\leq 2\text{cm}$ και επιφανειακή υποβλεννογόνια διήθηση, ενώ για βλάβη μεγαλύτερη από 2cm και επιφανειακή υποβλεννογόνια διήθηση βλάβη μικρότερη από 2 cm και σημαντική υποβλεννογόνια διήθηση είναι κατάλληλη η Ενδοσκοπική Υποβλεννογόνια Εκτομή (ESD)[38].

Η μετάσταση του καρκίνου στους λεμφαδένες εξαρτάται από το βάθος στο οποίο έχουν διηθήσει τα καρκινικά κύτταρα στον πολύποδα του παχέος εντέρου. Όταν πρόκειται για έμμοσχο κακοήθη πολύποδα (Penduculated Polyp), εισβολή επιπέδου 4, οι μεταστάσεις στους λεμφαδένες είναι έως και 27% ενώ όταν πρόκειται για άμισχο πολύποδα, εισβολή των καρκινικών κυττάρων στο κατώτερο 3ο επίπεδο του υποβλεννογόνου, οι μεταστάσεις φτάνουν στο 23%[55].

Ο καρκίνος του παχέος εντέρου σταδίου II αντιμετωπίζεται με χειρουργική επέμβαση και εκτομή του όγκου, ενώ δεν συνιστάται η χορήγηση επικουρικής χημειοθεραπείας[38].

Ο καρκίνος του παχέος εντέρου σταδίου III αντιμετωπίζεται με χειρουργική εκτομή του όγκου, ενώ χορηγείται επικουρική ακτινοχημειοθεραπεία η οποία λειτουργεί θετικά σε ασθενείς στους οποίους υπάρχει διασπορά της νόσου στους επιχώριους λεμφαδένες. Σε αυτό το στάδιο του καρκίνου η επικουρική χημειοθεραπεία συμβάλλει στην μείωση κατά 30% του κινδύνου υποτροπής ενώ συμβάλλει θετικά επίσης και στην μείωση της θνησιμότητας κατά 22% έως 32% [56].

Η χορήγηση επικουρικής χημειοθεραπείας στους ασθενείς ξεκινάει συνήθως τέσσερις έως έξι εβδομάδες έπειτα από την χειρουργική επέμβαση και περιλαμβάνει συνδυασμό λευκοβορίνης 5 – φθοριοουρακίλης και οξαλιπλατίνης (FOLFOX) ενώ χορηγούνται επίσης και φθοριοπυριμιδίνες[57].

Ο καρκίνος στο στάδιο IV έχει εξαπλωθεί, εκτός της αρχικής εστίας στο έντερο, και σε άλλα όργανα όπως το ήπαρ και οι πνεύμονες. Ως εκ τούτου, είναι δύσκολο συνήθως να υπάρχει ίαση του καρκίνου σε αυτό το στάδιο, μπορεί όμως να χορηγηθεί θεραπεία που να τον ελέγχει αλλά και να παρατείνει την εξέλιξή του για αρκετό χρονικό διάστημα. Στις περισσότερες περιπτώσεις ασθενών με καρκίνο του παχέος εντέρου σταδίου IV η χορήγηση χημειοθεραπείας έχει ως στόχο την εξασφάλιση καλύτερης ποιότητας ζωής στον ασθενή μειώνοντας την ένταση των συμπτωμάτων. [57] Πριν την χορήγηση της θεραπείας θα πρέπει να γίνει εκτίμηση της θέσης του καρκίνου, της μετάλλαξης καθώς και οι συννοσηρότητες του ατόμου [58], Η θεραπεία που χρησιμοποιείται συνήθως στον καρκίνο του παχέος εντέρου σταδίου IV είναι συνδυασμός χημειοθεραπείας, στοχευμένης μοριακής θεραπείας, ανοσοθεραπείας, ακτινοθεραπείας και παρηγορητικής χειρουργικής[38]. Παρηγορητική ακτινοθεραπεία χορηγείται πιο σπάνια στον καρκίνο του παχέος εντέρου σταδίου IV και έχει ως στόχο την συρρίκνωση του όγκου, όταν αυτός προκαλεί συμπτώματα όπως είναι η αιμορραγία[38]. Από την άλλη, όταν χρησιμοποιείται παρηγορητική χειρουργική στοχεύει κυρίως στην βελτίωση της ποιότητας ζωής του ασθενούς και όχι στην ίαση. Ειδικότερα, χρησιμοποιείται για την απόφραξη του παχέος εντέρου και την αντιμετώπιση της αιμορραγίας[59]. Το Stent του παχέος εντέρου αποτελεί επίσης μία καλή πρακτική που χρησιμοποιείται για την ρύθμιση της απόφραξης του παχέος εντέρου και έχει πάνω από 90% επιτυχία, αποφεύγοντας έτσι σοβαρές χειρουργικές επεμβάσεις[60].

Οι παράγοντες από τους οποίους εξαρτάται η επιλογή του κατάλληλου σχήματος χημειοθεραπείας αφορούν τα γενικότερα χαρακτηριστικά της νόσου, την κατάσταση στην οποία βρίσκεται ο όγκος, καθώς και την κατάσταση του ασθενούς. Κάποια από τα

σχήματα χημειοθεραπείας που χρησιμοποιούνται πιο συχνά στο στάδιο IV του καρκίνου του παχέος εντέρου είναι τα παρακάτω:

FOLFIRI (5-FU, LV, και Irinotecan): 5-Fluorouracil (5-FU) – 5-φθοροουρακίλη (5-FU), Leucovorin (folinic acid) – Λευκοβορίνη (φυλλινικό οξύ), Irinotecan – ιρινοτεκίνη [61].

FOLFOX (5 FU, LV και Oxaliplatin): 5-Fluorouracil (5-FU) – 5-φθοροουρακίλη (5-FU), Leucovorin (folinic acid) – Λευκοβορίνη (φυλλινικό οξύ), Oxaliplatin – Οξαλιπλατίνη [61].

CAPIRI (Capecitabine και Irinotecan): Καπεσιταβίνη και Ιρινοτεκίνη [61].

CAPOX (Capecitabine και Oxaliplatin): Καπεσιταβίνη και Οξαλιπλατίνη [61].

Η χημειοθεραπεία χορηγείται σε συνδυασμό με στοχευμένη θεραπεία. Η επιλογή του κατάλληλου τύπου στοχευμένης θεραπείας εξαρτάται κυρίως από το εάν ο καρκίνος είναι θετικός ή όχι στην μετάλλαξη KRAS (Kristen Rat Sarcoma Viral Oncogene Homolog), δηλαδή αν ο καρκίνος είναι επιθετικός ή όχι [38]. Η στοχευμένη θεραπεία (μονοκλωνικά αντισώματα) που χρησιμοποιείται στο στάδιο IV, για τον έλεγχο δηλαδή του καρκίνου σε προχωρημένο στάδιο περιλαμβάνει τα εξής:

Bevacizumab – Μπεβασιζουμάμπη: Πρόκειται για ένα μονοκλωνικό αντίσωμα, τύπος πρωτεΐνης, που παράγεται από το ανοσοποιητικό σύστημα του ανθρώπου για προστασία απέναντι σε λοιμώξεις και τον καρκίνο. Είναι ένας αγγειακός παράγοντας κατά του αγγειακού ενδοθηλιακού αυξητικού παράγοντα (VEGF). Δρα αποτρέποντας την ανάπτυξη καινούριων αιμοφόρων αγγείων τα οποία τροφοδοτούν τον όγκο με θρεπτικά συστατικά και οξυγόνο, εμποδίζοντας έτσι την ανάπτυξή του [38], [61].

Cetuximab – Σετουξιμάμπη: πρόκειται για μονοκλωνικό αντίσωμα το οποίο δεσμεύεται στον υποδοχέα του επιδερμικού αυξητικού παράγοντα EGFR (Epidermal Growth Factor Receptor), ένα αντιγόνο το οποίο ενεργοποιεί πρωτεΐνες που ευθύνονται για την ανάπτυξη και εξέλιξη του καρκίνου [38], [61].

Ramucirumab – Ραμουσιρουμάμπη: Είναι ένα μονοκλωνικό αντίσωμα που στοχεύει στον υποδοχέα του αγγειακού ενδοθηλιακού αυξητικού παράγοντα 2, VEGFR – 2 (Vascular Endothelial Growth Factor). Αναστέλλει την αγγειογένεση, δηλαδή την ανάπτυξη νέων αιμοφόρων αγγείων που τροφοδοτούν τον όγκο με θρεπτικά συστατικά και οξυγόνο [38], [61].

Regorafenib – Ρεγκοραφενίμπη: χρησιμοποιείται σε μεταστατικούς καρκίνους του παχέος εντέρου κυρίως όταν οι άλλες θεραπείες δεν είναι πλέον αποτελεσματικές. Είναι κινασικός αναστολέας που δρα αναστέλλοντας την αγγειογένεση, την ογκογένεση και το μικροπεριβάλλον του όγκου, αναστέλλοντας έτσι την δημιουργία και εξάπλωση καρκινικών κυττάρων [38], [61]

Aflibercept – Αφλιβερσέπτη: χορηγείται σε συνδυασμό με χημειοθεραπεία σε μεταστατικούς καρκίνους του παχέος εντέρου. Πρόκειται για έναν αναστολέα αγγειογένεσης ο οποίος εξουδετερώνει τον αγγειακό ενδοθηλιακό αυξητικό παράγοντα VEGF (Vascular Endothelial Growth Factor). Είναι μια πρωτεΐνη που δεσμεύει αυξητικούς παράγοντες VEGF, με αποτέλεσμα να μειώνει την ροή του αίματος προς τον όγκο, αναστέλλοντας έτσι την αγγειογένεση και άρα την ανάπτυξη του όγκου[61] [38]

Οι παραπάνω δραστικές ουσίες αναστέλλουν την αγγειογένεση του όγκου, αυξάνοντας έτσι την αποτελεσματικότητα της χημειοθεραπείας [61][62]. Στην δοκιμή AVF2107g παρατηρήθηκε ότι το Bevacizumab (Μπεβασιζουμάμπη) σε συνδυασμό με το χημειοθεραπευτικό σχήμα FOLFIRI σημείωσε βελτίωση στην παράταση της επιβίωσης σε 20,3 μήνες αντί για 15,6 όταν αυτά χορηγήθηκαν σε ασθενείς με μεταστατικό καρκίνο του παχέος εντέρου [60].

Κεφάλαιο 2ο

2. Τεχνητή Νοημοσύνη

2.1. Εισαγωγή

Η Τεχνητή Νοημοσύνη αποτελεί έναν από τους σημαντικότερους τομείς της τεχνολογίας που αφορά στην ανάπτυξη υπολογιστικών συστημάτων τέτοια ώστε να είναι ικανά να εκτελούν εργασίες οι οποίες συνήθως απαιτούν ανθρώπινη νοημοσύνη. Ήδη από το 1950 ο Alan Turing εξερεύνησε την μαθηματική δυνατότητα της Τεχνητής Νοημοσύνης σε μελέτη σχετικά με την Τεχνητή Νοημοσύνη διατυπώνοντας το ερώτημα αν μπορούν οι μηχανές να σκεφτούν[63]. Στην εργασία του “Computer Machinery and Intelligence” πρότεινε τον τρόπο κατασκευής έξυπνων μηχανών και τον τρόπο με τον οποίο μπορεί να δοκιμαστεί η ευφυία τους. Η ιδέα του Turing δεν προχώρησε, λόγω της έλλειψης δυνατότητας αποθήκευσης εντολών στους υπολογιστές της εποχής εκείνης, μπορούσαν μόνο να εκτελούν εντολές, καθώς και του υψηλού κόστους χρήσης τους. Για να εξασφαλιστούν πηγές χρηματοδότησης έπρεπε να υπάρχει απόδειξη της ιδέας του και υποστήριξη και από άλλα άτομα υψηλού προφίλ [64].

Ο όρος τεχνητή Νοημοσύνη διατυπώθηκε για πρώτη φορά το 1956 στο συνέδριο ερευνητών και επιστημόνων στο ερευνητικό πανεπιστήμιο Dartmouth College στο Hanover των ΗΠΑ, από τους επιστήμονες John McCarthy, Marvin Minsky, Claude Shannon και πολλούς άλλους, μελετώντας την δυνατότητα προσομοίωσης της ανθρώπινης νοημοσύνης σε μηχανή. Το γεγονός αυτό, επηρέασε σημαντικά την έρευνα για την Τεχνητή Νοημοσύνη τα επόμενα χρόνια [63][65].

Στην εξέλιξη της Τεχνητής Νοημοσύνης συνέβαλαν, σε ερευνητικό επίπεδο, επιστήμες όπως η φιλοσοφία, μέσω του συμπεράσματος ότι ενεργούμε ορθολογικά με βάση την γνώση και την συλλογιστική, η ψυχολογία, η οποία αποκωδικοποιώντας την ανθρώπινη συμπεριφορά και τον τρόπο σκέψης, κατέληξε στο συμπέρασμα ότι ο ίδιος ο άνθρωπος μπορεί να θεωρηθεί ως μια μηχανή επεξεργασίας πληροφοριών. Επιπλέον, τα μαθηματικά προσφέρουν εργαλεία διεξαγωγής συμπερασμάτων, μέσω των πιθανοτήτων,

της λογικής και της θεωρίας υπολογισμού. Επίσης, με βάση τις νευροεπιστήμες αναπτύχθηκαν τεχνικές μίμησης βιολογικών διεργασιών των νευρώνων του εγκεφάλου, όπως για παράδειγμα τα Τεχνητά Νευρωνικά Δίκτυα. Τέλος, η επιστήμη των υπολογιστών αποτελεί την βάση για την ανάπτυξη της Τεχνητής Νοημοσύνης, μέσω της ανάπτυξης αλγορίθμων και μαθηματικών μοντέλων επιτυγχάνεται η αυτοματοποιημένη μάθηση για την λήψη αποφάσεων από μηχανές, η δυνατότητα επεξεργασίας μεγάλου όγκου δεδομένων (Big Data), τεχνικές Μηχανικής Μάθησης και η χρήση Τεχνητών Νευρωνικών Δικτύων κάνει εφικτή την δημιουργία ευφυών συστημάτων που μιμούνται την ανθρώπινη νοημοσύνη [66].

Η Τεχνητή νοημοσύνη είναι πλέον παντού, βρίσκεται εφαρμογή σε αρκετούς κλάδους όπως η τεχνολογία, η υγεία, το μάρκετινγκ, η οικονομία, η βιομηχανία, η ψυχαγωγία και η εκπαίδευση με σκοπό την βελτίωση της αποτελεσματικότητας και την ανάπτυξη καινοτόμων συστημάτων. [64]. Ερευνητές από διαφορετικές επιστημονικές περιοχές χρησιμοποιούν συχνά την Τεχνητή Νοημοσύνη με σκοπό να αναζητήσουν εργαλεία αυτοματοποίησης των λογικών βημάτων που χρησιμοποιούν στις εργασίες τους [65].

2.2. Τι είναι η Τεχνητή Νοημοσύνη

Μέχρι σήμερα έχουν υπάρξει διάφοροι ορισμοί της Τεχνητής Νοημοσύνης, κάποιιοι από αυτούς να αφορούν στην διαδικασία σκέψης και συλλογισμού και κάποιιοι άλλοι στην συμπεριφορά. Οι Rusel και Norving ταξινόμησαν τους ορισμούς που αφορούν την Τεχνητή Νοημοσύνη σε τέσσερις μεγάλες κατηγορίες, οι οποίες εξετάζουν την περιοχή από διαφορετική οπτική γωνία σε σχέση με τον σκοπό της [65][66].

Σύμφωνα με την πρώτη κατηγορία σκοπός της Τεχνητής Νοημοσύνης είναι η ανάπτυξη συστημάτων που **σκέφτονται όπως οι άνθρωποι** με χαρακτηριστικό παράδειγμα τον Haugeland που δίνει τον εξής ορισμό για την Τεχνητή Νοημοσύνη: *“Η προσπάθεια να κατασκευάσουμε υπολογιστές με διανοητική ικανότητα με την πλήρη και κυριολεκτική έννοια του όρου”*.

Στην δεύτερη κατηγορία κατατάσσονται οι ορισμοί σύμφωνα με τους οποίους στόχος της Τεχνητής Νοημοσύνης είναι η ανάπτυξη συστημάτων που **σκέφτονται λογικά**, με ενδεικτικό παράδειγμα για το τι είναι η Τεχνητή Νοημοσύνη, να αποτελεί ο ορισμός που

δόθηκε από τον Winston: “*Η μελέτη των υπολογισμών που καθιστούν εφικτή την αντίληψη, την λογική σκέψη και την αντίδραση*”.

Η Τρίτη κατηγορία εστιάζει σε συστήματα που **συμπεριφέρονται όπως οι άνθρωποι**, σύμφωνα με τους Rich και Knight που ορίζουν την Τεχνητή Νοημοσύνη ως εξής: “*Η μελέτη του πως να κάνουμε τους υπολογιστές να κάνουν πράγματα στα οποία αυτή τη στιγμή οι άνθρωποι είναι καλύτεροι*”.

Η τέταρτη περιλαμβάνει τους ορισμούς που δηλώνουν ότι η Τεχνητή Νοημοσύνη αποσκοπεί στην ανάπτυξη συστημάτων που **αντιδρούν λογικά** με χαρακτηριστικό παράδειγμα τον ορισμό του Luger που αναφέρει για την Τεχνητή Νοημοσύνη το εξής: “*Ο τομέας της επιστήμης των υπολογιστών που ασχολείται με την αυτοματοποίηση της ευφυούς συμπεριφοράς*” [65][66].

Συνεπώς ένας γενικός ορισμός που περιλαμβάνει στοιχεία από τους ορισμούς των παραπάνω κατηγοριών είναι ο ακόλουθος:

“*Τεχνητή Νοημοσύνη είναι ο τομέας της Επιστήμης των Υπολογιστών που ασχολείται με την σχεδίαση και υλοποίηση υπολογιστικών συστημάτων τα οποία είναι ικανά να μιμηθούν τις ανθρώπινες γνωστικές ικανότητες, εμφανίζοντας έτσι χαρακτηριστικά που αποδίδουμε συνήθως σε ανθρώπινη συμπεριφορά, όπως για παράδειγμα η επίλυση προβλημάτων, η αντίληψη και κατανόηση εικόνων, η μάθηση, η εξαγωγή συμπερασμάτων, η κατανόηση φυσικής γλώσσας κτλ*” [65].

Ως εκ τούτου, η Τεχνητή Νοημοσύνη αποτελεί ένα συνεχώς εξελισσόμενο τομέα της επιστήμης των υπολογιστών που έχει ως στόχο να μας προσφέρει μηχανές οι οποίες δεν απαιτούν εξειδικευμένες γνώσεις για τον χειρισμό τους, θα προσαρμόζονται στις ανάγκες των χρηστών και θα επιλύουν δύσκολα καθημερινά προβλήματα. Η έρευνα στην Τεχνητή Νοημοσύνη προσφέρει σημαντικά αποτελέσματα σε διάφορους τομείς όπως η επεξεργασία φυσικής γλώσσας, η μηχανική μάθηση, η ρομποτική, η μηχανική όραση, βέλτιστη επίλυση προβλημάτων και άλλα [65].

Οι δύο βασικές προσεγγίσεις με διαφορετική φιλοσοφία και εφαρμογές συνέβαλλαν εξίσου στην ανάπτυξη της σύγχρονης τεχνητής Νοημοσύνης:

Η συμβολική Τεχνητή Νοημοσύνη (Symbolic AI), η οποία με βάση την κατανόηση των νοητικών διεργασιών εστιάζει στην προσομοίωση της ανθρώπινης νοημοσύνης χρησιμοποιώντας αλγορίθμους, που βασίζονται στον χειρισμό συμβόλων, για την επίλυση προβλημάτων και την λήψη αποφάσεων. Παραδείγματα αυτής της προσέγγισης

αποτελούν οι εφαρμογές Τεχνητής Νοημοσύνης για αναπαράσταση γνώσης όπως κανόνες, δένδρα απόφασης, πλαίσια κ.α.

Η υπολογιστική Νοημοσύνη (Computational Intelligence), η οποία βασίζεται στην μίμηση της λειτουργίας του ανθρώπινου εγκεφάλου εστιάζοντας στην δυνατότητα των συστημάτων να μαθαίνουν από δεδομένα. Πεδία αυτής της προσέγγισης αποτελούν η Μηχανική Μάθηση, που χρησιμοποιεί αλγόριθμους για ανάλυση δεδομένων, εξαγωγή συμπερασμάτων και λήψη αποφάσεων, τα Τεχνητά Νευρωνικά Δίκτυα και τα μοντέλα Βαθιάς Μάθησης[65].

2.3. Πεδία εφαρμογής της Τεχνητής Νοημοσύνης

Η ανάπτυξη της τεχνολογίας, σε συνδυασμό με την διάδοση του διαδικτύου, καθώς και η ανάπτυξη προγραμματιστικών περιβαλλόντων, συνέβαλλαν στην δημιουργία ευφών συστημάτων τα χαρακτηριστικά των οποίων αποτελούν τα βασικά πεδία εφαρμογής της Τεχνητής Νοημοσύνης [65], [66].

Μέσω Τεχνητής Νοημοσύνης αναπτύχθηκαν συστήματα τα οποία διαθέτουν την ικανότητα μάθησης με την εξαγωγή μοντέλων ή προτύπων. Τα συστήματα αυτά έχουν την δυνατότητα να μαθαίνουν μέσω της εκπαίδευσή τους με πρωτογενή δεδομένα και συσχέτισης αυτών. Τα συστήματα επεξεργασίας φυσικής γλώσσας αποτελούν επίσης ένα πεδίο εφαρμογής της Τεχνητής Νοημοσύνης, μέσω των οποίων παρέχεται η δυνατότητα επικοινωνίας της μηχανής με τον άνθρωπο. Επιπλέον, η μηχανική όραση όπου, μέσω αλγορίθμων, οι υπολογιστές έχουν την δυνατότητα να επεξεργάζονται και να κατανοούν οπτικά δεδομένα με σκοπό να λαμβάνουν αποφάσεις. Η ρομποτική αποτελεί επίσης ένα από τα πεδία εφαρμογής της Τεχνητής Νοημοσύνης για την ανάπτυξη συστημάτων τα οποία διαθέτουν την δυνατότητα κίνησης. Ο τομέας της αναπαράστασης γνώσης που στοχεύει στην αποθήκευση πληροφοριών και η αυτοματοποιημένη συλλογιστική που βοηθάει στην αξιοποίησή τους για την εξαγωγή συμπερασμάτων και την απάντηση ερωτημάτων [65], [66].

Τα σημαντικότερα επιτεύγματα της Τεχνητής Νοημοσύνης των τελευταίων ετών, οφείλονται κυρίως στην διαθεσιμότητα μεγάλου όγκου δεδομένων, στην μεγάλη υπολογιστική ισχύ, πλέον χρησιμοποιούνται ειδικοί επεξεργαστές (GPUs), και στην παραγωγή νέων αποδοτικών αλγορίθμων μηχανικής μάθησης. Πλέον έχουμε στην

διάθεσή μας εφαρμογές που χρησιμοποιούν ευφυείς μηχανισμούς και υποσυστήματα, με σκοπό την βελτίωση της αποτελεσματικότητας και της ακρίβειας καθώς και την δυνατότητα λήψης αποφάσεων σε πολλούς τομείς όπως:

Στην ιατρική: η Τεχνητή Νοημοσύνη προσφέρει εργαλεία που βοηθούν στην ανάλυση ιατρικών εικόνων, στην αποτελεσματική διάγνωση, την ανάπτυξη νέων φαρμάκων κ.α.

Στην **βιομηχανία**, βοηθώντας στην αυτοματοποίηση στις διαδικασίες παραγωγής.

Στις **μεταφορές** μέσω ευφύων συστημάτων πλοήγησης ευφύων συστημάτων πλοήγησης αυτόνομων συστημάτων και συστημάτων διαχείρισης της κυκλοφορίας.

Στην **οικονομία**, όπου συστήματα Τεχνητής Νοημοσύνης χρησιμοποιούνται στις τράπεζες για την αυτοματοποίηση των συναλλαγών.

Στις **επιχειρήσεις** όπου χρησιμοποιούνται εφαρμογές Τεχνητής Νοημοσύνης στρατηγικού σχεδιασμού και λήψης αποφάσεων για την δημιουργία στοχευμένων διαφημίσεων, πρόβλεψη ζήτησης κ.α.

Στην **ψυχαγωγία** με την χρήση αλγορίθμων τεχνητής Νοημοσύνης για την δημιουργία περιεχομένου και συστημάτων προτάσεων σε πλατφόρμες περιεχομένου με βάση του προφίλ του χρήστη[65].

2.4. Η Τεχνητή Νοημοσύνη στην Ιατρική

Καθώς διανύουμε την εποχή της εξατομικευμένης ιατρικής, δημιουργείται η ανάγκη για εύρεση αποτελεσματικότερων διαγνωστικών εργαλείων, μοντέλων πρόβλεψης, ακριβείς προγνωστικούς δείκτες και μοντέλων σχεδιασμού εξατομικευμένης θεραπείας. Στην αντιμετώπιση αυτών των προκλήσεων στην ιατρική συμβάλλει η Τεχνητή Νοημοσύνη με σκοπό την παροχή καλύτερης υγειονομικής περίθαλψης. Η Μηχανική Μάθηση βοηθάει στην αξιολόγηση ακτινολογικών εικόνων καθώς και στην διαχείριση των ηλεκτρονικών ιατρικών φακέλων ασθενών, ενισχύοντας έτσι τις διαδικασίες διάγνωσης και την επιλογή της κατάλληλης θεραπείας [67].

Στην ιατρική, καθοριστικοί παράγοντες για την φροντίδα των ασθενών, αποτελούν η γνώση και η εμπειρία. Όσον αφορά τον ανθρώπινο παράγοντα, ο κύριος περιορισμός για την απόκτηση γνώσης και εμπειρίας είναι οι χρονικοί περιορισμοί. Πλέον, στην εποχή της Τεχνητής Νοημοσύνης, τεράστιες ποσότητες δεδομένων ασθενών μπορούν να

αποκτηθούν και να επεξεργαστούν ώστε να μετασχηματιστούν κατάλληλα προς όφελος των ασθενών.

Οι εφαρμογές υπολογιστών, μέσω αλγορίθμων, παρέχουν την δυνατότητα απόκτησης εμπειρίας σε αισθητά μικρότερο χρόνο, συγκριτικά με τον άνθρωπο. Επιπλέον, η ταχύτητα ανάγνωσης αξονικών τομογραφιών μέσω Τεχνητής Νοημοσύνης είναι σημαντικά μεγαλύτερη σε σχέση με τον μέσο άνθρωπο [67].

Η ανάλυση μεγάλου όγκου δεδομένων και η χρήση αλγορίθμων Μηχανικής Μάθησης συμβάλλει στην λήψη ταχύτερων και ακριβέστερων αποφάσεων, βελτιώνοντας έτσι την ποιότητα περίθαλψης των ασθενών. Επιπλέον, η δυνατότητα ανάλυσης απεικονιστικών δεδομένων, τα συστήματα πρόγνωσης ασθενειών και η χρήση ρομποτικής στον τομέα της χειρουργικής, κάνουν απαραίτητη την χρήση εργαλείων Τεχνητής νοημοσύνης σε πολλούς τομείς της ιατρικής [67].

Ο όρος τεχνητή νοημοσύνη στην ιατρική επικεντρώνεται κυρίως στους παρακάτω όρους:

Ανάλυση εικόνας: πρόκειται για μαθηματική διαδικασία σύμφωνα με την οποία εισάγεται μια εικόνα και εξάγεται με μεγαλύτερη ευκρίνεια. Επιτρέπει την ανάλυση ιατρικών εικόνων εντοπίζοντας με ακρίβεια πιθανά προβλήματα με την χρήση αλγορίθμων Μηχανικής Μάθησης ενισχύοντας έτσι την διάγνωση από τους ειδικούς με αποτέλεσμα να προσφέρει πιο έγκυρες και εξατομικευμένες λύσεις για τους ασθενείς [67].

Μηχανική Μάθηση: είναι ο κλάδος της Τεχνητής Νοημοσύνης που εστιάζει στην ανάπτυξη μοντέλων και αλγορίθμων που επιτρέπουν στα συστήματα να μάθουν από την εμπειρία. Οι υπολογιστές χρησιμοποιούν δεδομένα με τα οποία εκπαιδεύονται ώστε να αναγνωρίζουν μοτίβα, να λαμβάνουν αποφάσεις και να κάνουν προβλέψεις. Στην σύγχρονη ιατρική η Μηχανική μάθηση χρησιμοποιείται σε αλγορίθμους πρόβλεψης για την πιθανότητα εμφάνισης νόσων, στην ανάλυση εικόνας για τον εντοπισμό καρκινικών κυττάρων κ.α. [67].

Υπολογιστική Όραση: αφορά στην ανάπτυξη συστημάτων τα οποία επιτρέπουν στους υπολογιστές να δέχονται ως είσοδο ψηφιακές εικόνες, να τις επεξεργάζονται, να τις κατανοούν και να τις ερμηνεύουν. Στην ιατρική χρησιμοποιείται στην ανάλυση ιατρικών εικόνων και την διάγνωση ασθενειών, αλλά και στην ρομποτική χειρουργική [67].

Βαθιά Μάθηση: αποτελεί ένα πεδίο της Μηχανικής Μάθησης δομημένο με τέτοιο τρόπο ώστε να προσαρμόζει την λειτουργία του ανθρώπινου εγκεφάλου. Οι τεχνικές της Βαθιάς Μάθησης βασίζονται σε δίκτυα με πολλά επίπεδα, τα οποία δέχονται ως είσοδο σύνολα

δεδομένων. Κάθε επίπεδο αξιολογεί διαδοχικά τα δεδομένα προσθέτοντας πληροφορίες βελτιώνοντας έτσι την ακρίβεια του τελικού αποτελέσματος. Τα επίπεδα αυτά των δικτύων ονομάζονται “κρυφά” επίπεδα και λειτουργούν με τέτοιο τρόπο ώστε η έξοδος ενός επιπέδου να αποτελεί είσοδο στο επόμενο. Αν, για παράδειγμα, σε ένα τέτοιο δίκτυο εισαχθεί ως είσοδος μια εικόνα κολονοσκόπησης για την ανίχνευση πολύποδα, η εικόνα αρχικά διασπάται σε πολλαπλά υποσύνολα και στη συνέχεια επεξεργάζεται με διαφορετικά φίλτρα. Σε κάθε επίπεδο εξάγεται μία βαθμολογία η οποία μεταβιβάζεται στο επόμενο επίπεδο επεξεργασίας ως είσοδος και αυτή η ροή επεξεργασίας συνεχίζεται έως ότου παραχθεί το τελικό αποτέλεσμα. Στην προκειμένη περίπτωση, το τελικό αποτέλεσμα μπορεί να αφορά την διάγνωση ή τον εντοπισμό ενός πολύποδα [67].

Τεχνητά Νευρωνικά Δίκτυα: πρόκειται για μαθηματικό μοντέλο που βασίζεται σε μη γραμμικά εργαλεία μοντελοποίησης δεδομένων τα οποία αναλύουν πολύπλοκες σχέσεις μεταξύ εισόδων και εξόδων. Είναι εμπνευσμένο από τον τρόπο λειτουργίας των βιολογικών νευρώνων του ανθρώπινου εγκεφάλου όσον αφορά την επεξεργασία διαφόρων τύπων δεδομένων και την δημιουργία προτύπων, τα οποία χρησιμοποιούνται σε διαδικασίες λήψης αποφάσεων. Σε ένα Τεχνητό νευρωνικό Δίκτυο η είσοδος επεξεργάζεται από ένα σύνολο αλγορίθμων και η έξοδος που θα προκύψει θα εισαχθεί ως είσοδος σε ένα άλλο σύνολο αλγορίθμων προκειμένου να παραχθεί ένα τελικό αποτέλεσμα. Η δυνατότητα των Τεχνητών Νευρωνικών Δικτύων να αναλύουν μεγάλο όγκο δεδομένων, να αναγνωρίζουν πρότυπα και να προβλέπουν αποτελέσματα συμβάλλει σημαντικά στην ιατρική τόσο στην διάγνωση όσο και στην επιλογή θεραπείας [67].

Συνελκτικικά Νευρωνικά Δίκτυα: αποτελούν έναν τύπο Τεχνητού Νευρωνικού Δικτύου για ανάλυση δεδομένων που βασίζεται σε αλγορίθμους βαθιάς μάθησης χρησιμοποιώντας πολλά κρυφά επίπεδα. Χρησιμεύουν ιδιαίτερα στην αναγνώριση εικόνων και την ανάλυση βίντεο στην ιατρική απεικόνιση [67].

Οι αλγόριθμοι Βαθιάς Μάθησης (Deep Learning – DL) και Συνελκτικικών Νευρωνικών Δικτύων (Convolutional Neural Networks – CNN) χρησιμοποιούνται ευρέως στη σύγχρονη ιατρική. Συμβάλλουν σημαντικά στη διαχείριση μεγάλου όγκου δεδομένων, τον έλεγχο πληροφοριών, στη διάγνωση και τη χορήγηση φαρμάκων. Από την άλλη, κλάδοι της φυσικής ασχολούνται με την παροχή ιατρικού εξοπλισμού και την ανάπτυξη συστημάτων ρομπότ τα οποία χρησιμοποιούνται ευρέως στην χειρουργική.

2.5. Εφαρμογή Τεχνητής Νοημοσύνης στην ιατρική

Η Τεχνητή Νοημοσύνη βρίσκει εφαρμογή σε αρκετούς κλάδους της ιατρικής όπως η ακτινολογία, η ογκολογία, η καρδιολογία, η οφθαλμολογία, η γαστρεντερολογία, η δερματολογία, και η παθολογία. Η χρήση εργαλείων Τεχνητής Νοημοσύνης συμβάλλουν σημαντικά στη διάγνωση του καρκίνου καθώς και στον προσδιορισμό του σταδίου της νόσου [67].

Η Τεχνητή Νοημοσύνη βρίσκει ευρεία εφαρμογή στην ακτινολογία χάρη στην πρόοδο που έχει σημειωθεί στην ανάλυση εικόνας. Διάφορες ερευνητικές ομάδες έχουν δημιουργήσει αλγορίθμους επεξεργασίας εικόνας και υπολογιστικής όρασης με σκοπό την έγκαιρη διάγνωση και την ακριβέστερη οπτικοποίηση των πιθανών παθολογιών. Στόχος είναι να υπάρχουν εφαρμογές που θα εστιάζουν σε πληροφορίες που δεν μπορεί να εντοπίσει η ανθρώπινη όραση, αλλά και να παραχθεί γνώση σε σύντομο χρονικό διάστημα. Συγκεκριμένα, η ισραηλινή εταιρεία Max Q – AI Ltd, η οποία εξειδικεύεται στην ανάπτυξη εργαλείων υποβοήθησης λήψης αποφάσεων σε επείγουσες ιατρικές καταστάσεις, ανέπτυξε μία πλατφόρμα για την ανίχνευση ενδοκρανιακής αιμορραγίας, επιδεικνύοντας την περιοχή στους ακτινολόγους μέσω επεξεργασίας 3D δεδομένων αξονικής τομογραφίας.

Η συγκεκριμένη ιδέα αναπτύχθηκε περαιτέρω από την εταιρεία Viz.AI.inc από το Πανεπιστήμιο του Στάνφορντ. Η εταιρεία ανέπτυξε μια πλατφόρμα υποστήριξης κλινικών αποφάσεων, εγκεκριμένη από την Υπηρεσία Τροφίμων και Φαρμάκων της Αμερικής (FDA), που χρησιμοποιεί ανάλυση εικόνας αξονικής τομογραφίας εγκεφάλου για να στέλνει άμεσα ειδοποίηση στον ειδικό νευρολόγο, όταν ανιχνεύσει απόφραξη αγγείων του εγκεφάλου – κύρια αιτία εγκεφαλικών επεισοδίων. Το πλεονέκτημα της εφαρμογής έγκειται στο ότι επιτρέπει την έγκαιρη παρέμβαση του ειδικού, συχνά νωρίτερα από το καθορισμένο πρότυπο περίθαλψης. Σύμφωνα με τα αποτελέσματα κλινικής μελέτης που πραγματοποιήθηκε σε 300 εικόνες αξονικής τομογραφίας εγκεφάλου, έδειξαν ότι η εφαρμογή μπορεί να ειδοποιήσει τον ειδικό νευρολόγο νωρίτερα από δύο εξειδικευμένους νευρολόγους, στην περίπτωση υποψίας απόφραξης [67], [68].

Η χρήση τέτοιων αλγορίθμων σε συστήματα αρχειοθέτησης και επικοινωνίας εικόνας (PACS) εφαρμόζεται σε πάνω από πενήντα νοσοκομεία παγκοσμίως αναλύοντας πάνω από ένα εκατομμύριο απεικονίσεις. Πλέον υπάρχει η δυνατότητα αυτόματης ανίχνευσης

καρδιαγγειακών, πνευμονολογικών, ηπατικών καθώς και οστικών παθήσεων. Ενδεικτικά παραδείγματα αποτελούν ο αυτόματος εντοπισμός σπονδυλικών καταγμάτων με χρήση αλγορίθμων Τεχνητής Νοημοσύνης και η ανίχνευση οστεοπόρωσης με την χρήση αλγορίθμων μέτρησης οστικής πυκνότητας [67], [69].

Η διάγνωση και η σταδιοποίηση του καρκίνου του μαστού είναι τομέας της ιατρικής όπου η χρήση Τεχνητής Νοημοσύνης μπορεί να επιφέρει καλύτερα αποτελέσματα από τον άνθρωπο. Μελέτες έχουν αποδείξει την αξιοπιστία της Μηχανικής Μάθησης στην διάγνωση του καρκίνου. Συγκεκριμένα, σε μελέτη του Somashekhar έδειξε ότι το Υπολογιστικό Σύστημα Watson της IBM είχε κατά 93% συμφωνία με ένα σύνολο επαγγελματιών υγείας σχετικά με τα θεραπευτικά πρωτόκολλα που εφαρμόζονται στην αντιμετώπιση του καρκίνου του μαστού.

Επιπλέον, σε μελέτη του Benjordi και της ερευνητικής του ομάδας, όπου χρησιμοποιήθηκε σύνολο εκπαίδευσης αποτελούμενο από 129 φωτογραφικές διαφάνειες, εκ των οποίων οι 49 αντιστοιχούσαν σε μεταστάσεις στους λεμφαδένες, ο αλγόριθμος πέτυχε καλύτερη απόδοση στη διάγνωση σε σύγκριση με 11 παθολόγους. Στην συγκεκριμένη μελέτη, αξιολογήθηκε το γεγονός ότι οι γιατροί χρειάστηκαν 30 ώρες για να αξιολογήσουν και τις 129 φωτογραφικές διαφάνειες, σε αντίθεση με τον ελάχιστο χρόνο που χρειάστηκε ο αλγόριθμος για την επεξεργασία τους[67].

Οι αλγόριθμοι Τεχνητής Νοημοσύνης έχουν αποδειχθεί πιο αποτελεσματικοί, σε σχέση με τον άνθρωπο, και στην ανίχνευση του καρκίνου του πνεύμονα. Σε μελέτη του Kun-Hsing You και της ομάδας του, όπου χρησιμοποιήθηκαν 2.186 ιστοπαθολογικές εικόνων ολικής αλλοίωσης του αδενοκαρκινώματος και του ακανθοκυτταρικού καρκινώματος του πνεύμονα αποδείχθηκε η αποτελεσματικότητα της Τεχνητής Νοημοσύνης στη διάγνωση του καρκίνου του πνεύμονα. Αυτό συνέβαλλε στη βελτίωση της φροντίδας των ασθενών και στην επιλογή της κατάλληλης θεραπείας. [67].

Η Τεχνητή Νοημοσύνη έχει σημειώσει πρόοδο και στη δερματολογία, ιδίως στη διάγνωση και ταξινόμηση των δερματικών ασθενειών με βάση τις εικόνες. Σε μελέτη που δημοσιεύθηκε στο περιοδικό *Nature* (Andre Esteva), χρησιμοποιήθηκε ένα μόνο CNN, το οποίο εκπαιδεύτηκε σε ένα σύνολο δεδομένων 129.450 κλινικών εικόνων – που συλλέχθηκαν έπειτα από βιοψία – δύο ομάδων: η μία αντιπροσώπευε συνηθισμένους καρκίνους και η άλλη καρκίνους με υψηλότερη θνησιμότητα. Η σύγκριση των αποτελεσμάτων του CNN με εκείνα 21 πιστοποιημένων δερματολόγων έδειξε ότι τα

αποτελέσματα της Τεχνητής Νοημοσύνης συμφωνούσαν πλήρως με αυτά των δερματολόγων σε όλες τις περιπτώσεις και για τις δύο ομάδες εικόνων [67].

2.6. Η Τεχνητή Νοημοσύνη στον Καρκίνο του Παχέος Εντέρου

Η συμβολή της Τεχνητής Νοημοσύνης παίζει σημαντικό ρόλο κυρίως στη διάγνωση, αλλά και στη θεραπεία του καρκίνου του παχέος εντέρου. Η εφαρμογή της στον προ συμπτωματικό έλεγχο μπορεί να βελτιώσει σε σημαντικό βαθμό τον έγκαιρο εντοπισμό, μειώνοντας έτσι την συχνότητα εμφάνισης και την θνησιμότητα από τη νόσο. Οι αλγόριθμοι Μηχανικής Μάθησης και τα Συνελκτικά Νευρωνικά Δίκτυα, τα οποία είναι ιδιαίτερα αποτελεσματικά στην ταξινόμηση εικόνων, έχουν ενσωματωθεί σε σημαντικό βαθμό σήμερα στην ιατρική. Παράλληλα, καινοτόμες συσκευές όπως τα ρομποτικά χειρουργικά συστήματα και τα νανορομπότ συμβάλλουν ιδιαίτερα στη διάγνωση του καρκίνου του παχέος εντέρου [70]

Η ενσωμάτωση Τεχνητής Νοημοσύνης με γονιδιωματικά δεδομένα έχει δείξει σημαντικά αποτελέσματα στη διαχείριση του καρκίνου του παχέος εντέρου. Έχουν υπάρξει μελέτες που έδειξαν την αποτελεσματικότητα της Τεχνητής Νοημοσύνης στην έρευνα για τη νόσο αναφορικά με την πρόβλεψη κινδύνου και την ανίχνευση μεταλλάξεων.[70] Συγκεκριμένα, οι Hu et al. σε πείραμα προσομοίωσης χρησιμοποίησαν Νευρωνικά Δίκτυα ταξινομώντας με επιτυχία ασθενείς με καρκίνο του παχέος εντέρου, έπειτα από χειρουργική επέμβαση, σε δυο ομάδες: υποτροπιάζουσα και μη υποτροπιάζουσα, με βάση πρότυπα γονιδιακής έκφρασης. Πρότειναν το δίκτυο S – Kohonen ως το πιο αποτελεσματικό για την ταξινόμηση του καρκίνου του παχέος εντέρου υπογραμμίζοντας έτσι τις δυνατότητες της Τεχνητής Νοημοσύνης στην πρόβλεψη μετεγχειρητικών αποτελεσμάτων [71]. Επιπλέον, οι Xu et al. χρησιμοποίησαν Support Vector Machines (SVM) σε μελέτη για τον εντοπισμό γονιδίων διαφορετικής έκφρασης (DEGs) όπου εντόπισαν μία υπογραφή 15 γονιδίων η οποία μπορεί να λειτουργήσει ως προγνωστικός παράγοντας του κινδύνου υποτροπής από καρκίνο του παχέος εντέρου[70].

Μία νέα προσέγγιση η οποία βασίστηκε στη φασματοσκοπία NIR σε συνδυασμό με Τεχνητό Νευρωνικό Δίκτυο Αντίθετης Διάδοσης(CP – ANN – Counter propagation Artificial Neural Network) όπου επιτρέπει την ακριβή ανίχνευση μεταλλάξεων του γονιδίου BRAF V600E, με ποσοστό ευαισθησίας 100% και ακρίβεια διάγνωσης 93,8% [70]. Επίσης, ερευνητική ομάδα το 2015 σχεδίασε Τεχνητό Νευρωνικό Δίκτυο για τη συσχέτιση γενετικών και περιβαλλοντικών παραγόντων, προσφέροντας σημαντικά δεδομένα για επιγενετικές αλλαγές που σχετίζονται με τον καρκίνο του παχέος εντέρου. Σε διάφορες μελέτες χρησιμοποιήθηκαν δεδομένα προφίλ γονιδιακής έκφρασης από τη Βάση Δεδομένων The Cancer Genome Atlas, τα οποία αναλύθηκαν μέσω αλγορίθμων Τεχνητής Νοημοσύνης δημιουργώντας έτσι διαγνωστικά μοντέλα , με ποσοστά ακριβείας κοντά στο 100%. [70]

Τα νέα στοιχεία δείχνουν ότι τα μη κωδικοποιημένα RNA (nc-RNAs) επηρεάζουν κρίσιμες σηματοδοτικές οδούς σχετικά με την ανάπτυξη και εξέλιξη του όγκου σε κάθε στάδιο. Υπάρχουν μελέτες που αναδεικνύουν το ρόλο των ncRNAs στην έγκαιρη διάγνωση, την πρόγνωση και τη στοχευμένη θεραπεία. [70] Ερευνητές έχουν εντοπίσει υπογραφή τριών microRNA (Mir 139 – 5p, Mir – 3, Mir – 17 – 92) για την πρόβλεψη της κατάστασης του όγκου στο στάδιο II [72]. Επιπλέον, αναπτύχθηκε και ένας υπολογιστικός αλγόριθμος, ο οποίος Χρησιμοποίησε ένα ταξινομητή Naïve Bayes, με σκοπό τον εντοπισμό αλληλεπιδράσεων μεταξύ των microRNA και των mRNAs και ο οποίος θα μπορούσε να βοηθήσει σημαντικά στην ανάπτυξη νέων στοχευμένων φαρμάκων.

Μέχρι σήμερα, η χρήση εργαλείων Τεχνητής Νοημοσύνης έχει βοηθήσει τους κλινικούς γιατρούς στην πρόβλεψη και την πρόγνωση της ασθένειας του καρκίνου του παχέος εντέρου. Οι τρέχουσες εξελίξεις δείχνουν ότι η Τεχνητή Νοημοσύνη αποτελεί σημαντικό εργαλείο για την κατανόηση της βιολογίας του καρκίνου του παχέος εντέρου με σκοπό την βελτιωμένη φροντίδα των ασθενών. Αρκετές μελέτες έχουν ασχοληθεί με το μικροπεριβάλλον των όγκων καθώς και με το ρόλο των ανοσοκυττάρων σε αυτό. Για παράδειγμα, η χρήση του αλγορίθμου CIBERSORT στην ανάλυση ανοσοκυττάρων βοηθάει τους κλινικούς γιατρούς να επιλέξουν στοχευμένη ανοσοθεραπεία [70].

2.7. Μηχανική Μάθηση

Η προσπάθεια του ανθρώπου για την κατανόηση του περιβάλλοντός του βασίζεται στην παρατήρηση και τη δημιουργία απλοποιημένων εκδοχών του, και έτσι οδηγείται στη δημιουργία προτύπων. Επιπλέον, ο άνθρωπος έχει τη δυνατότητα να οργανώνει και να συσχετίζει τις εμπειρίες του σχηματίζοντας έτσι νέες δομές που ονομάζονται πρότυπα. Οι δύο αυτές έννοιες, τα μοντέλα και τα πρότυπα, σχετίζονται με την ανθρώπινη μάθηση. Συνεπώς, η δημιουργία μοντέλων ή προτύπων μέσω ενός υπολογιστικού συστήματος βασιζόμενο σε ένα σύνολο δεδομένων, ονομάζεται Μηχανική Μάθηση. [65] Βασικός στόχος της Μηχανικής Μάθησης είναι η δημιουργία μοντέλων που να μπορούν να γενικεύουν καλά σε νέα δεδομένα.

Οι τεχνικές Μηχανικής Μάθησης, οι οποίες χρησιμοποιούνται ανάλογα με τη φύση του προβλήματος, διακρίνονται σε Μάθηση με Επίβλεψη, Μάθηση Χωρίς Επίβλεψη και Ενισχυτική Μάθηση. Στη μάθηση με επίβλεψη, το υπολογιστικό σύστημα μαθαίνει (από δεδομένα) μια συνάρτηση στόχο, η οποία αντιπροσωπεύει το μοντέλο που περιγράφει τα δεδομένα. Η συνάρτηση στόχος χρησιμοποιείται για την εκτίμηση της τιμής μιας

μεταβλητής, γνωστή ως εξαρτημένη μεταβλητή ή μεταβλητή εξόδου, με βάση τις τιμές ενός συνόλου μεταβλητών που ονομάζονται ανεξάρτητες μεταβλητές ή μεταβλητές εισόδου. Το σύνολο των δυνατών τιμών εισόδου (πεδίο ορισμού), στη συνάρτηση στόχο, ονομάζεται στιγμιότυπο. [65] Στη μάθηση με επίβλεψη, το υπολογιστικό σύστημα εκπαιδεύεται να κατασκευάζει μοντέλα ή πρότυπα με βάση ένα σύνολο δεδομένων όπου ένα “δάσκαλος” του παρέχει καθοδήγηση αντιστοιχίζοντας κάποιες από τις εισόδους στις επιθυμητές εξόδους.

Στη μάθηση χωρίς επίβλεψη το υπολογιστικό σύστημα στοχεύει στην ανακάλυψη συσχετίσεων και ομάδων, σε δεδομένες τα οποία δεν διαθέτουν ετικέτες ή προϋπάρχουσα κατηγοριοποίηση. Με άλλα λόγια δημιουργεί πρότυπα ή μοντέλα χωρίς να υπάρχει προηγούμενη γνώση για τη δομή του συνόλου δεδομένων [65].

Η ενισχυτική μάθηση αφορά ένα σύνολο τεχνικών στις οποίες το σύστημα εκπαιδεύεται μέσω αλληλεπίδρασης με το περιβάλλον. Ο τρόπος που μαθαίνει το μοντέλο, είναι εμπνευσμένος από βιολογικά μοντέλα μάθησης, τα οποία βασίζονται στις αρχές της επιβράβευσης και της τιμωρίας. Στόχος είναι η μεγιστοποίηση ενός αριθμητικού σήματος ενίσχυσης – ανταμοιβή – μαθαίνοντας ποια ενέργεια είναι η καταλληλότερη σε κάθε κατάσταση. [65]

Εφαρμογές της ενισχυτικής μάθησης περιλαμβάνουν τον έλεγχο ρομπότ, τη βελτιστοποίηση εργασιών, τη μάθηση επιτραπέζιων παιχνιδιών, την αυτοματοποίηση σύνθετων συστημάτων κ.τ.λ. Είναι καταλληλότερη για διαδραστικά προβλήματα όπου το σύστημα έχει την ικανότητα να μαθαίνει μέσα από τη δική του εμπειρία, καθώς δεν υπάρχει καθοδήγηση [65].

Στη μάθηση με επίβλεψη συναντάμε δύο βασικά είδη προβλημάτων, την ταξινόμηση που εστιάζει στη δημιουργία μοντέλων για την πρόβλεψη διακριτών κατηγοριών και την παρεμβολή που εστιάζει στη δημιουργία μοντέλων για την πρόβλεψη συνεχών αριθμητικών τιμών [65]. Στην παρούσα εργασία εξετάζεται η χρήση προηγμένων μεθόδων Μηχανικής Μάθησης που χρησιμοποιούνται στην ιατρική. Τα Τεχνητά Νευρωνικά Δίκτυα, και ειδικότερα τα Συνελκτικά Τεχνητά Νευρωνικά Δίκτυα (CNNs) αποτελούν ισχυρά εργαλεία ανάλυσης ιατρικών εικόνων με σκοπό να εξάγουν σύνθετα χαρακτηριστικά από δεδομένα ιστοπαθολογικών εικόνων από βιοψίες, επιτυγχάνοντας έτσι ακρίβεια και αξιοπιστία στον εντοπισμό ευρημάτων.

Κεφάλαιο 3ο

3. Τεχνητά Νευρωνικά Δίκτυα

3.1. Εισαγωγή

Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούν παραμετρικά μοντέλα Μηχανικής Μάθησης, των οποίων η αρχιτεκτονική είναι εμπνευσμένη από τη λειτουργία των βιολογικών νευρωνικών δικτύων, και χρησιμοποιούνται εκτενώς σε προβλήματα αναγνώρισης προτύπων [73]. Προβλήματα από διάφορες επιστήμες όπως, η φυσική, η χημεία, η ιατρική, η οικονομία κ.α, μπορούν να μετασχηματιστούν σε προβλήματα αναγνώρισης προτύπων και να λυθούν μέσω Τεχνητών Νευρωνικών Δικτύων [73]. Έχουν την ικανότητα να μαθαίνουν από δεδομένα, να αναγνωρίζουν πρότυπα και να λαμβάνουν αποφάσεις, βασιζόμενα σε ένα σύνολο αλληλένδετων υπολογιστικών μονάδων, τους τεχνητούς νευρώνες. Έχουν χρησιμοποιηθεί σε μια ευρεία γκάμα εφαρμογών όπως διάγνωση ασθενειών, η μηχανική όραση, και η πρόβλεψη χρονοσειρών [65].

Παρόλο που οι βασικές ιδέες των Νευρωνικών Δικτύων είχαν ερευνηθεί από τη δεκαετία του 1950, η πρακτική εφαρμογή τους εφαρμόστηκε δεκαετίες μετά, εξαιτίας της έλλειψης αποδοτικών μεθόδων εκπαίδευσής τους. Αυτό άλλαξε τη δεκαετία του 1980 με την επαναφορά του αλγορίθμου οπίσθιας τροφοδότησης (backpropagation). Στη συνέχεια, η εμφάνιση άλλων μοντέλων Μηχανικής Μάθησης, όπως οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), η τεχνική Booting καθώς και τα Random Forests, περιόρισε τη χρήση των Τεχνητών Νευρωνικών Δικτύων. Τα νέα αυτά μοντέλα υπερτερούσαν ως προς τα αποτελέσματα και απαιτούσαν λιγότερη παραμετροποίηση, σε σχέση με τα Τεχνητά Νευρωνικά Δίκτυα [74].

Από τις αρχές της δεκαετίας του 2010 τα Τεχνητά Νευρωνικά Δίκτυα επανήλθαν ως “Βαθιά Μάθηση” (Deep Learning) σημειώνοντας μεγάλη πρόοδο σε προβλήματα όπως η αναγνώριση εικόνας και βίντεο και η επεξεργασία κειμένου. Η άνοδος που έχει σημειώσει η Βαθιά Μάθηση, ιδιαίτερα την τελευταία δεκαετία, οφείλεται σε διάφορους παράγοντες. Η τεράστια άνοδος της υπολογιστικής ισχύος κατέστησε δυνατή την

εκπαίδευση Νευρωνικών Δικτύων ακόμη και σε προσωπικούς υπολογιστές. Επιπλέον, σημαντικό ρόλο έπαιξε και η διαθεσιμότητα μεγάλου όγκου δεδομένων, καθώς η αύξηση της αποθηκευτικής ικανότητας σε συνδυασμό με το Διαδίκτυο συνέβαλλαν στη συλλογή και διανομή των δεδομένων για χρήση τους στη Μηχανική Μάθηση. Τέλος, η εμφάνιση αλγοριθμικών βελτιώσεων συνέβαλλε στην εκπαίδευση των Νευρωνικών Δικτύων.

3.2. Νευρωνικά Δίκτυα – Βασική Δομή και Λειτουργία

Ένα Τεχνητό Νευρωνικό Δίκτυο, ή Νευρωνικό Δίκτυο, είναι ένας μεγάλης κλίμακας παράλληλος επεξεργαστής με κατανομημένη αρχιτεκτονική. Αποτελείται από βασικές μονάδες επεξεργασίας, τους Τεχνητούς Νευρώνες, ή Νευρώνες, και έχει την ικανότητα να αποθηκεύει εμπειρική γνώση με τέτοιο τρόπο ώστε να είναι διαθέσιμη για μελλοντική χρήση [75]. Τα Νευρωνικά Δίκτυα παρουσιάζουν ομοιότητες με τους νευρώνες του ανθρώπινου εγκεφάλου σε 2 σημεία: αποκτούν γνώση από το περιβάλλον τους, μέσω μιας διαδικασίας μάθησης, και για να την αποθηκεύσουν χρησιμοποιείται η ισχύς των συνδέσεων μεταξύ των νευρώνων, που ονομάζεται συναπτικό βάρος. Η μάθηση επιτυγχάνεται μέσω μιας διαδικασίας που ονομάζεται που λέγεται Αλγόριθμος Μάθησης, ο ρόλος του οποίου είναι να προσαρμόζει τα συναπτικά βάρη των νευρώνων με τέτοιο τρόπο, ώστε να οδηγήσει στην επίτευξη του επιθυμητού αποτελέσματος (στόχου) [75].

Ουσιαστικά, κάθε Νευρώνας δέχεται ένα πλήθος n εισόδων, αντιστοιχίζοντας ένα συγκεκριμένο βάρος σε κάθε μία από τις εισόδους. Στη συνέχεια, υπολογίζει το άθροισμά τους, το οποίο διοχετεύει σε μια μη γραμμική συνάρτηση μέσω της οποίας παράγεται η τελική έξοδος.

3.2.1. Το μοντέλο του Τεχνητού Νευρώνα

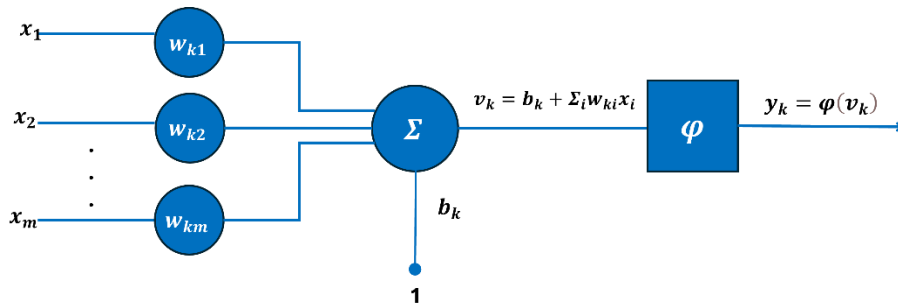
Το μοντέλο του Τεχνητού Νευρώνα αποτελείται από τρία βασικά μέρη:

1. **Ένα σύνολο από Συναψεις**, κάθε μία από τις οποίες χαρακτηρίζεται από ένα “βάρος”. Σε ένα σήμα x_i στην είσοδο της σύναψης i , που συνδέεται με τον Νευρώνα k πολλαπλασιάζεται με το συναπτικό βάρος W_{ki} . Το συναπτικό βάρος W_{ki} μπορεί να λαμβάνει θετικές τιμές, αν η σύναψη ωθεί τον νευρώνα να αποκριθεί στη διέγερση, ή αρνητικές τιμές αν η σύναψη αποτρέπει τον νευρώνα να παράγει μια απόκριση.

2. **Τον Αθροιστή**, ο οποίος αθροίζει τα σήματα εισόδου με βάση τα συναπτικά βάρη του νευρώνα.

3. **Τη Συνάρτηση Ενεργοποίησης**, η οποία χρησιμοποιείται για τη μείωση του εύρους της εξόδου του νευρώνα, περιορίζοντάς την σε ένα διάστημα της μορφής $[0,1]$ ή $[-1, 1]$ [75].

Η παραπάνω διαδικασία απεικονίζεται στο παρακάτω σχήμα:



Σχήμα 1 Μοντέλο Τεχνητού Νευρώνα

Το μοντέλο περιλαμβάνει επίσης και μια εξωτερικά εφαρμοζόμενη πόλωση (bias) b_k με στόχο να αυξήσει ή να μειώσει τη δικτυακή διέγερση της συνάρτησης ενεργοποίησης, αναλόγως με το αν είναι θετική ή αρνητική αντίστοιχα.

Ένας νευρώνας k περιγράφεται με μαθηματικούς όρους από τις παρακάτω εξισώσεις:

$$u_k = \sum_{j=1}^m w_{kj} \cdot x_j \quad (3.1)$$

$$y_k = \varphi(u_k + b_k) \quad (3.2)$$

$$v_k = u_k + b_k \quad (3.3)$$

Όπου:

x_1, x_2, \dots, x_m είναι τα σήματα εισόδου,

$w_{k1}, w_{k2}, \dots, w_{km}$ είναι τα συναπτικά βάρη του νευρώνα,

u_k είναι η γραμμική συνδυαστική έξοδος του νευρώνα που προκύπτει από τα σήματα εισόδου,

φ είναι η συνάρτηση ενεργοποίησης,

y_k είναι η έξοδος του νευρώνα και

v_k είναι το τοπικό πεδίο ή δυναμικό ενεργοποίησης.

3.2.2. Συνάρτηση Ενεργοποίησης

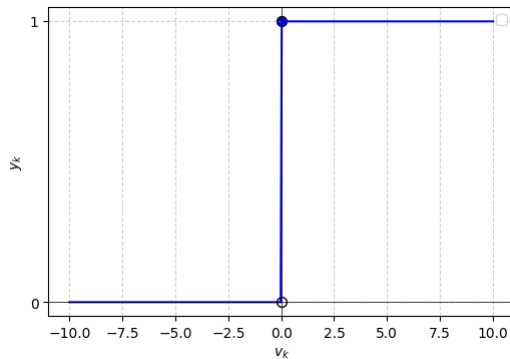
Η συνάρτηση ενεργοποίησης $\varphi(v)$ καθορίζει την έξοδο ενός νευρώνα με βάση τα εισερχόμενα σήματα εισόδου. Η συμπεριφορά του μοντέλου του νευρώνα διαφοροποιείται ανάλογα με τον τύπο της συνάρτησης ενεργοποίησης φ που χρησιμοποιείται. Υπάρχουν δύο βασικές συναρτήσεις ενεργοποίησης, η Συνάρτηση Κατωφλίου (Threshold Function) και η Σιγμοειδής Συνάρτηση (Sigmoid Function).

1. Συνάρτηση Κατωφλίου:

Η έξοδος ενός νευρώνα k , όταν χρησιμοποιεί τη συνάρτηση κατωφλίου είναι:

$$y_k = \begin{cases} 1 & \text{αν } v_k \geq 0 \\ 0 & \text{αν } v_k < 0 \end{cases} \quad (4) \quad (3.4)$$

Όπου v_k είναι το τοπικό πεδίο του Τεχνητού Νευρώνα.

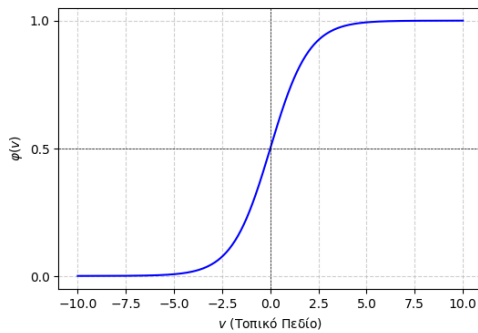


Σχήμα 2 Συνάρτηση Κατωφλίου

2. Σιγμοειδής Συνάρτηση

Αποτελεί την πιο συνηθισμένη μορφή συνάρτησης ενεργοποίησης που χρησιμοποιείται στα Τεχνητά Νευρωνικά Δίκτυα. Είναι μια αυστηρά αύξουσα συνάρτηση η οποία παρουσιάζει εξομάλυνση και ασυμπτωτικές ιδιότητες. Η γραφική παράσταση της συνάρτησης έχει χαρακτηριστικό σχήμα “S”. Ένα παράδειγμα σιγμοειδής συνάρτησης αποτελεί η λογιστική συνάρτηση που ορίζεται από την παρακάτω σχέση:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (3.5)$$



Σχήμα 3 Σιγμοειδής Συνάρτηση

3.2.3. Το Perceptron

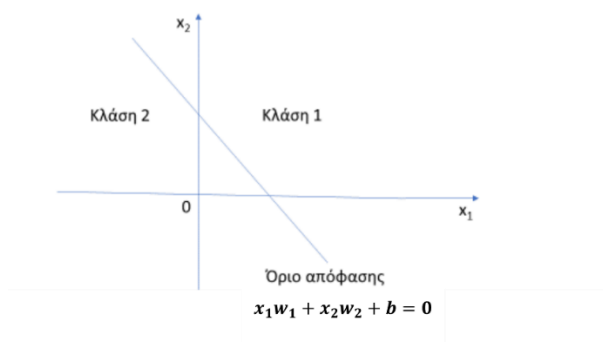
Το perceptron είναι η απλούστερη μορφή Τεχνητού Νευρωνικού Δικτύου το οποίο λειτουργεί ως βασικό θεμέλιο στην κατανόηση σύνθετων αρχιτεκτονικών Τεχνητών Νευρωνικών Δικτύων. Αποτελείται από έναν μόνο τεχνητό νευρώνα με προσαρμόσιμα συναπτικά βάρη και πόλωση που συμβολίζεται με b (bias). Η έξοδος του Perceptron προκύπτει από τον υπολογισμό του τοπικού πεδίου ως γραμμικός συνδυασμός των εισόδων με τα αντίστοιχα βάρη, πάνω στον οποίο εφαρμόζεται ένας απότομος περιοριστής (κατώφλι). Συγκεκριμένα, παράγει ως απόκριση +1, αν η είσοδος του απότομου περιοριστή είναι θετική και -1 αν είναι αρνητική. Η δυαδική αυτή έξοδος επιτρέπει στο Perceptron να ταξινομεί το σύνολο των εισόδων του σε δύο κλάσεις, την κλάση +1 και την κλάση -1.

Η έξοδος του μοντέλου που προκύπτει από τον υπολογισμό της εισόδου (τοπικό πεδίο) δίνεται από τη σχέση:

$$u = \sum_{i=1}^m w_i x_i + b \quad (3.6)$$

Το άθροισμα εισόδων περνάει από μια βηματική συνάρτηση ενεργοποίησης όπου δίνει την έξοδο y .

$$y = \begin{cases} +1, & \text{αν } u > 0 \\ -1, & \text{αν } u \leq 0 \end{cases} \quad (3.7)$$



Σχήμα 4 Το όριο και οι περιοχές απόφασης

Στόχος του Perceptron είναι να διαχωρίσει γραμμικά τα πρότυπα εισόδου στις δυο παραπάνω κλάσεις. Στην περίπτωση που οι είσοδοι είναι γραμμικά διαχωρίσιμες, το Perceptron μπορεί να επιτύχει τέλεια ταξινόμηση [65].

Η εξίσωση του υπερεπιπέδου που διαχωρίζει τις δύο περιοχές απόφασης είναι:

$$\sum_{i=1}^m w_i x_i + b = 0 \Leftrightarrow w_1 x_1 + w_2 x_2 + \dots + w_m x_m = -b \quad (3.8)$$

Όπου, w_i τα συναπτικά βάρη, x_i η είσοδος και b η πόλωση.

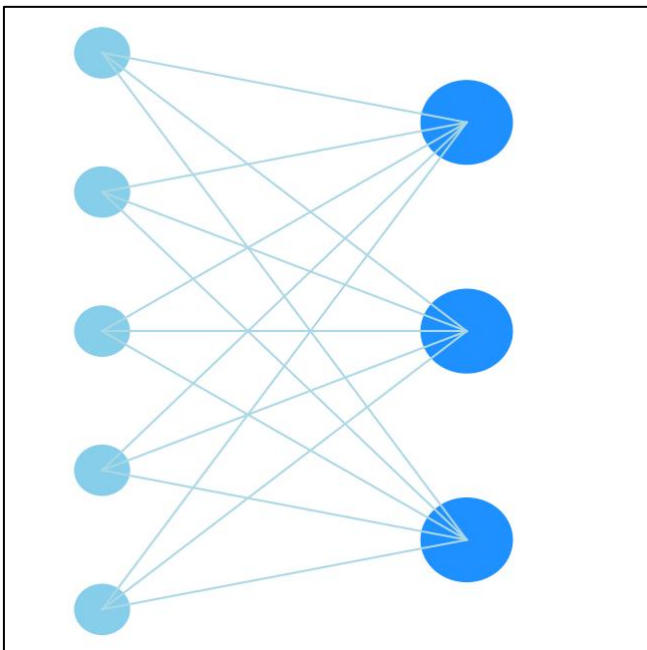
Τα συναπτικά βάρη στο Perceptron προσαρμόζεται μέσω μιας διαδικασίας εκπαίδευσης. Η σύνδεση πολλών Perceptron σε διαδοχικά επίπεδα σχηματίζει τα πολυεπίεδα Perceptron.

3.2.4. Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα διακρίνονται σε διάφορες κατηγορίες ανάλογα με τον τρόπο με τον οποίο συνδέονται οι νευρώνες, τη ροή της πληροφορίας και το είδος του προβλήματος που λύνουν [76]. Στην αρχιτεκτονική ενός δικτύου συμπεριλαμβάνονται η διαμόρφωση των επιπέδων, ο αριθμός των νευρώνων κάθε επιπέδου, ο μεταξύ τους τύπος σύνδεσης και οι συναρτήσεις ενεργοποίησης.

Οι τρεις κατηγορίες αρχιτεκτονικών δικτύων είναι η αρχιτεκτονική ενός επιπέδου δικτύων πρόσθιας τροφοδότησης, πολυεπίεδων δικτύων πρόσθιας τροφοδότησης και αναδρομικών δικτύων. Βασικότερη κατηγορία αποτελούν τα Δίκτυα Πρόσθιας Τροφοδότησης (Feedforward Neural Networks) στα οποία η πληροφορία ρέει προς τα εμπρός χωρίς ανατροφοδότηση.

Στην αρχιτεκτονική δικτύου πρόσθιας τροφοδότησης ενός επιπέδου, το επίπεδο εισόδου, το οποίο αποτελείται από πηγαίους κόμβους, συνδέεται άμεσα με το επίπεδο εξόδου το οποίο περιλαμβάνει τους νευρώνες εξόδου. Στο συγκεκριμένο δίκτυο η ροή της πληροφορίας γίνεται μόνο προς τη μία κατεύθυνση, από την είσοδο προς τους νευρώνες εξόδου, δεν υπάρχει ενδιάμεσο επίπεδο. Το δίκτυο ονομάζεται ενός επιπέδου πρόσθιας τροφοδότησης γιατί οι πηγαίοι κόμβοι δεν εκτελούν υπολογισμούς[76]. Στο σχήμα 5 παρουσιάζεται ένα δίκτυο πρόσθιας τροφοδότησης ενός επιπέδου με 5 πηγαίους κόμβους και 3 νευρώνες εξόδου.

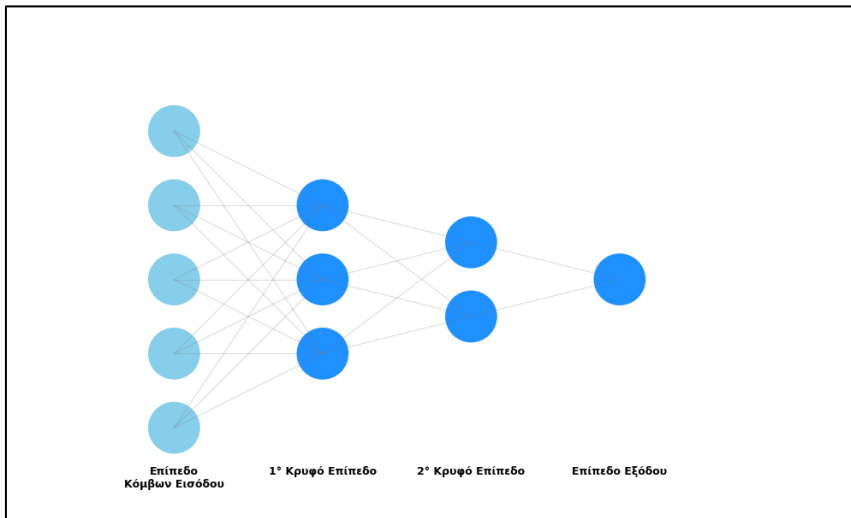


Σχήμα 5 Δίκτυο ενός επιπέδου πρόσθιας τροφοδότησης

Στην περίπτωση των πολυεπίπεδων δικτύων πρόσθιας τροφοδότησης (Multilayer Feedforward Networks) η αρχιτεκτονική τους περιλαμβάνει ένα επίπεδο εισόδου αποτελούμενο από πηγαίους κόμβους (input nodes), τροφοδοτώντας ένα ή περισσότερα κρυφά επίπεδα (Hidden Layers) νευρώνων. Τα ενδιάμεσα επίπεδα ονομάζονται κρυφά επειδή οι ενεργοποιήσεις τους δεν είναι προσβάσιμες ούτε από τα δεδομένα εισόδου, αλλά ούτε και από την τελική έξοδο.

Σε τέτοιου τύπου αρχιτεκτονικές δικτύων κάθε επίπεδο νευρώνων λαμβάνει ως είσοδο τα σήματα εξόδου που έχουν προκύψει από το προηγούμενο επίπεδο διαμορφώνοντας έτσι μια αυστηρά ιεραρχική ροή πληροφορίας. Τα σήματα εξόδου των νευρώνων στο επίπεδο εξόδου καθορίζονται από τη συνολική απόκριση του δικτύου σε ένα δοσμένο στιγμιότυπο εισόδου. Έπειτα από την επεξεργασία που γίνεται διαδοχικά μέσω των κρυφών επιπέδων.

Στο Σχήμα 6 παρουσιάζεται ένα παράδειγμα πολυεπίπεδου δικτύου το οποίο περιλαμβάνει πέντε πηγαίους κόμβους εισόδου, δύο κρυφά επίπεδα, το πρώτο με τρεις νευρώνες και το δεύτερο με δυο επίπεδα, και ένα νευρώνα εξόδου, το συγκεκριμένο δίκτυο ονομάζεται πλήρως συνδεδεμένο αφού κάθε επίπεδο του δικτύου συνδέεται με κάθε κόμβο του επόμενου επιπέδου.



Σχήμα 6 Πολυεπίπεδο Δίκτυο Πρόσθιας Τροφοδότησης

Σε ένα πολυεπίπεδο δίκτυο πρόσθιας τροφοδότησης η έξοδος κάθε νευρώνα υπολογίζεται ως εξής:

$$v_j = w_{j1} \cdot x_1 + w_{j2} \cdot x_2 + \dots + w_{jn} \cdot x_n \cdot b_j \quad (3.9)$$

Όπου:

w_{ji} τα συναπτικά βάρη που συνδέουν τις εισόδους x_i με τον νευρώνα j

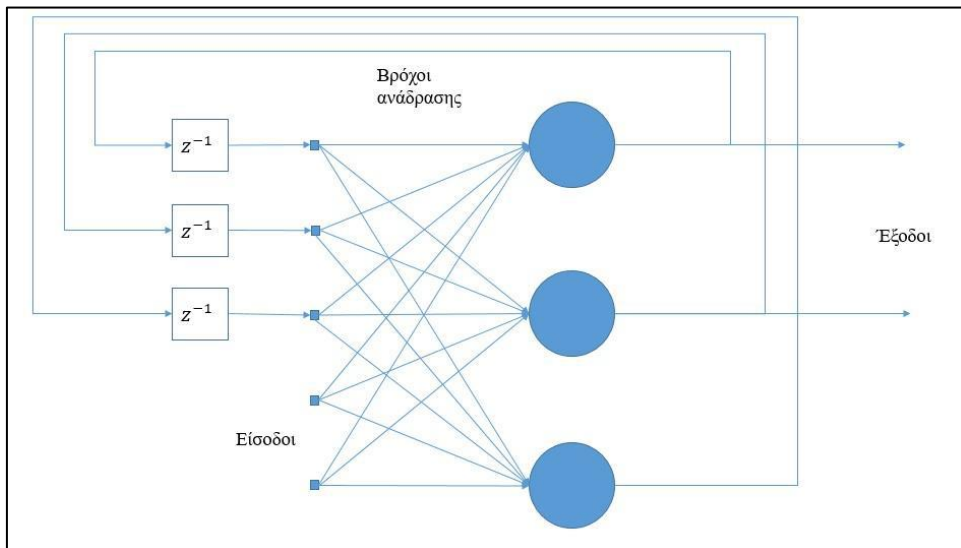
b_j η πόλωση (bias)

Στη συνέχεια υπολογίζεται η έξοδος του νευρώνα μέσω της συνάρτησης ενεργοποίησης φ

Η έξοδος του δικτύου προκύπτει όταν αυτή η διαδικασία επαναληφθεί διαδοχικά σε όλα τα επίπεδα, όπου κάθε επίπεδο λαμβάνει ως είσοδο την έξοδο του προηγούμενου.

Στα αναδρομικά Νευρωνικά Δίκτυα (RNNs) υπάρχουν κυκλικές συνδέσεις (βρόχοι ανάδρασης). Η συγκεκριμένη αρχιτεκτονική αποτυπώνεται στο Σχήμα 7 Οι αναδρομικές

αυτές συνδέσεις επηρεάζουν σημαντικά την διαδικασία προσαρμογής των βαρών και άρα και την συνολική απόδοση του μοντέλου κατά την φάση πρόβλεψης ή ταξινόμησης.



Σχήμα 7 Αναδρομικό Νευρωνικό Δίκτυο

3.2.5. Διαδικασία Μάθησης

Οι διαδικασίες μάθησης διακρίνονται σε δύο βασικές κατηγορίες: μάθηση με επίβλεψη και μάθηση χωρίς επίβλεψη. Στην μάθηση με επίβλεψη γνωρίζουμε εκ των προτέρων την επιθυμητή έξοδο του δικτύου, ενώ οι παράμετροι του μοντέλου προσαρμόζονται έτσι ώστε να μειωθεί το σφάλμα, ελαχιστοποιείται δηλαδή η διαφορά ανάμεσα στην αναμενόμενη και την πραγματική έξοδο. Ένας από τους πιο διαδεδομένους αλγόριθμους μάθησης με επίβλεψη για πολυεπίπεδα perceptron είναι ο αλγόριθμος οπισθοδιάδοσης (Backpropagation) ο οποίος βασίζεται στον κανόνα διόρθωσης του λάθους (Error Correction Learning Rule) [75].

Από την άλλη, η μάθηση χωρίς επίβλεψη διακρίνεται σε δυο κύριες μορφές: την ενισχυτική μάθηση και την αυτοοργανούμενη μάθηση. Στην ενισχυτική μάθηση, η πληροφορία για την επιθυμητή έξοδο παρέχεται ποιοτικά και η διαδικασία εκμάθησης της αντίστοιχης εισόδου – εξόδου γίνεται μέσω συνεχούς αλληλεπίδρασης με το περιβάλλον, με σκοπό την ελαχιστοποίηση ενός κριτηρίου απόδοσης. Στην αυτοοργανούμενη μάθηση οι παράμετροι του Νευρωνικού Δικτύου προσαρμόζονται με τέτοιο τρόπο, ώστε να ταξινομήσουν τα δεδομένα εισόδου σε διαφορετικές κλάσεις (κατηγορίες), σύμφωνα με κάποιο κριτήριο ομοιότητας, χωρίς όμως να υπάρχει προηγούμενη γνώση για την κατηγορία στην οποία ανήκουν τα δεδομένα εισόδου [75].

3.2.6. Ο Αλγόριθμος Οπισθοδιάδοσης του λάθους (Error Back Propagation Algorithm)

Ο πιο διαδεδομένος αλγόριθμος εκμάθησης πολυεπίπεδου perceptron, είναι ο Αλγόριθμος Οπισθοδιάδοσης (Backpropagation). Μέσω του συγκεκριμένου αλγορίθμου τα συναπτικά βάρη των συνδέσεων του Νευρωνικού Δικτύου τροποποιούνται με τέτοιο τρόπο ώστε να ελαχιστοποιείται μια συνάρτηση σφάλματος η οποία μετράει τη διαφορά μεταξύ της επιθυμητής και της πραγματικής εξόδου.

Σε ένα πολυεπίπεδο Perceptron, ενός επιπέδου εισόδου, δυο κρυφών επιπέδων και ενός επιπέδου εξόδου, όπως εμφανίζεται στο σχήμα 6, το σύνολο εκπαίδευσης που χρησιμοποιείται για εκπαίδευση με επίβλεψη του δικτύου ορίζεται το:

$$T = \{x(n), d(b)_n^N = 1\} \quad (3.10)$$

Επίσης, καθορίζεται ότι το σήμα της εξόδου του νευρώνα j συμβολίζεται με $y_j(n)$ για το στιγμιότυπο εισόδου $x(n)$.

Η παρακάτω σχέση δίνει το σφάλμα που παράγεται στην έξοδο του νευρώνα j .

$$e_j(n) = d_j(n) - y_j(n) \quad (3.11)$$

Όπου $d_j(n)$ η επιθυμητή απόκριση για το στιγμιότυπο εισόδου $x(n)$.

Η συνολική στιγμιαία ενέργεια σφάλματος του δικτύου προκύπτει από το άθροισμα των τετραγώνων αυτών των σφαλμάτων για όλα τα σήματα εξόδου και ορίζεται ως:

$$E(n) = \sum_{j \in C} e_j^2(n) \quad (3.12)$$

Όπου το C περιλαμβάνει τους νευρώνες στο επίπεδο εξόδου.

Εφόσον το σύνολο εκπαίδευσής περιλαμβάνει N στιγμιότυπα, η μέση ενέργεια σφάλματος για ολόκληρο το σύνολο εκπαίδευσης ορίζεται ως:

$$E_{av} = \frac{1}{N} \sum_{n=1}^n E(n) = \frac{1}{2N} \sum_{n=1}^N \sum_{j \in C} e_j^2(n) \quad (3.13)$$

Η μάθηση μπορεί να είναι online μάθηση ή μαζική μάθηση ανάλογα με το αν τα συναπτικά βάρη του Δικτύου μεταβάλλονται. Στην online μάθηση, η συνάρτηση σφάλματος προς ελαχιστοποίηση ορίζεται ως η συνολική ενέργεια σφάλματος $E(n)$. Σύμφωνα με αυτό, ο αλγόριθμος οπισθοδιάδοσης εφαρμόζει μια διόρθωση $\Delta w_{ji}(n)$ για κάθε βάρος του δικτύου $w_{ji}(n)$, κατά ποσό ανάλογο με την μερική παράγωγο της συνάρτησης σφάλματος ως προς το βάρος $\frac{\partial E(n)}{\partial w(n)}$. Την προσαρμογή αυτή περιγράφει ο κανόνας Δέλτα, όπου ο ρυθμός μάθησης καθορίζει το μέγεθος του βήματος προσαρμογής. Η μεταβολή του βάρους υπολογίζεται ως:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \quad (3.14)$$

όπου η είναι ο ρυθμός μάθησης του αλγορίθμου οπισθοδιάδοσης.

Το αρνητικό πρόσημο στη σχέση ενημέρωσης βαρών εξασφαλίζει ότι η ενημέρωση γίνεται προς την κατεύθυνση που μειώνει το σφάλμα $E(n)$.

Μέσω του κανόνα της αλυσίδας γίνεται ο υπολογισμός της παραγώγου $\frac{\partial E(n)}{\partial w_{ji}(n)}$ και έτσι ορίζεται το τοπικό σφάλμα $\delta_j(n)$.

Για τους νευρώνες εξόδου ισχύει:

$$\delta_j(n) = e_j(n) \cdot \varphi'(v_j(n)) \quad (3.15)$$

Για τους κρυφούς νευρώνες ισχύει:

$$\delta_j(n) = \varphi'(v_j(n)) \sum_{k \in S} \delta_k(n) w_{kj}(n) \quad (3.16)$$

Όπου S είναι το σύνολο των νευρώνων στο επόμενο επίπεδο.

Ενημέρωση των βαρών:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) \quad (3.17)$$

Για το bias ισχύει:

$$b_j(n+1) = b_j(n) + \eta \delta_j(n) \quad (3.18)$$

Η διαδικασία περιλαμβάνει:

Εμπρόσθια διέλευση: υπολογισμό απόκρισης κάθε νευρώνα

Οπισθοδιάδοση: ενημέρωση των βαρών με βάση το σφάλμα

Ο αλγόριθμος συνεχίζει μέχρι το σφάλμα να μειωθεί σε επιθυμητό επίπεδο ή να φτάσουμε σε μέγιστο αριθμό εποχών.

Ο αλγόριθμος οπισθοδιάδοσης δίνει τη δυνατότητα στα Νευρωνικά Δίκτυα να μαθαίνουν από δεδομένα, προσαρμόζοντας τα βάρη με τέτοιο τρόπο ώστε να ελαχιστοποιηθεί το συνολικό σφάλμα.

Η επαναλαμβανόμενη πρόσθια διάδοση και οπισθοδιάδοση σε ολόκληρο το σύνολο εκπαίδευσης οδηγεί στην συνεχή βελτίωση της απόδοσης του Δικτύου, καθιστώντας τη χρήση του αλγορίθμου θεμελιώδη σε σύγχρονες εφαρμογές Deep Learning.

3.2.7. Ενημέρωση συναπτικών βαρών με Γενικευμένο Κανόνα Δέλτα

Ο Γενικευμένος Κανόνας δέλτα αποτελεί βασικό μηχανισμό της διαδικασίας εκπαίδευσης ενός πολυεπιπέδου Νευρωνικού Δικτύου πρόσθιας τροφοδότησης, μέσω του αλγορίθμου οπισθοδιάδοσης του σφάλματος. Είναι ο αλγόριθμος που επιτρέπει την εφαρμογή της μεθόδου της φθίνουσας κλήσης (Gradient Descent) σε δίκτυα με κρυφά στρώματα. Ο ρυθμός μάθησης αποτελεί έναν κρίσιμο παράγοντα του Κανόνα Δέλτα. Ο μικρός ρυθμός μάθησης εξασφαλίζει μικρές μεταβολές στα συναπτικά βάρη ανά επανάληψη, προκαλώντας πιο ομαλή πορεία στον χώρο των συναπτικών βαρών, επιβραδύνοντας όμως παράλληλα τη διαδικασία μάθησης. Αντίθετα, μεγάλος ρυθμός μάθησης επιταχύνει τη μάθηση, όμως οι μεγάλες μεταβολές στα βάρη μπορεί να οδηγήσουν σε αστάθεια του Νευρωνικού Δικτύου.

Για να αυξηθεί ο ρυθμός μάθησης χωρίς τον κίνδυνο αστάθειας, τροποποιείται ο Κανόνας Δέλτα μεν τέτοιο τρόπο ώστε να περιλαμβάνει έναν όρο ορμής (momentum term), δημιουργώντας τον Γενικευμένο Κανόνα Δέλτα σύμφωνα με την εξίσωση:

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \quad (3.19)$$

Όπου, α είναι θετικός αριθμός που ονομάζεται σταθερά ορμής η οποία βοηθάει στη διατήρηση μιας ομαλής και σταθερής διαδικασίας της εκπαίδευσης. Η εξίσωση (.19) ονομάζεται Γενικευμένος Κανόνας Δέλτα [75].

3.2.8. Αρχικοποίηση Συναπτικών Βαρών

Έχουν προταθεί αρκετές διαφορετικές μέθοδοι αρχικοποίησης των βαρών που μπορούν να συμβάλλουν στη γρήγορη σύγκλιση του αλγορίθμου οπισθοδιάδοσης του σφάλματος. Συνήθως τα βάρη αρχικοποιούνται με μικρές τυχαίες τιμές για να αποφευχθεί η συμμετρία στις ενημερώσεις των βαρών.

Στην περίπτωση που τα βάρη αρχικοποιούνται με την ίδια τιμή, οι νευρώνες ενός επιπέδου θα έδιναν πανομοιότυπες εξόδους και θα μάθαιναν τα ίδια χαρακτηριστικά. Η τυχαία αρχικοποίηση, από την άλλη, επιτρέπει στους νευρώνες να εξελιχθούν διαφοροποιημένα κατά την διαδικασία της μάθησης, οδηγώντας έτσι σε καλύτερη προσαρμογή στο πρόβλημα [76].

3.2.9. Προεπεξεργασία του διανύσματος εισόδου

Η προεπεξεργασία των δεδομένων εισόδου είναι βασικό στάδιο για την βελτίωση της απόδοσης και της σταθερότητας του δικτύου και περιλαμβάνει τεχνικές όπως:

Κανονικοποίηση (Normalization), όπου γίνεται μετασχηματισμός των τιμών εισόδου (στο διάστημα π.χ. [0,1] ή [-1, 1]) για την αποφυγή της κυριαρχίας κάποιων μεταβλητών λόγω μεγάλης κλίμακας.

- Η κανονικοποίηση σε ένα διάστημα [a,b] γίνεται με την χρήση της παρακάτω εξίσωσης:

$$X_{norm} = a + \frac{(X - X_{min})}{X_{max} - X_{min}} \quad (4.20)$$

- **Αντιμετώπιση Ελλιπών Δεδομένων** με χρήση μεθόδων όπως είναι η συμπλήρωση με τον μέσο όρο ή χρήση του αλγορίθμου K – Κοντινότερων Γειτόνων για την συμπλήρωση κενών τιμών.
- **Διαχείριση Θορύβου**, όπου μέσω αλγορίθμων συσταδοποίησης γίνεται αφαίρεση στιγμιοτύπων ακραίων τιμών (outliers) ή θορυβωδών δεδομένων.

3.2.10. Εύρεση βέλτιστης Αρχιτεκτονικής Δικτύου

Η επιλογή της βέλτιστης αρχιτεκτονικής ενός πολυεπίδεδου νευρωνικού δικτύου είναι κρίσιμη κατά τη διαδικασία της σχεδίασης και της εκπαίδευσής ως προς την ικανότητα γενίκευσης και την αποτελεσματικότητά του. Ως αρχιτεκτονική ορίζεται η εύρεση του πλήθους των επιπέδων του δικτύου, το πλήθος των νευρώνων σε κάθε επίπεδο, και ειδικά στα κρυφά επίπεδα.

Αν ένα Νευρωνικό Δίκτυο περιλαμβάνει περιορισμένο αριθμό κρυφών νευρώνων ή επιπέδων, παρουσιάζει μειωμένη εκφραστική ικανότητα και ενδεχομένως να αδυνατεί να προσεγγίσει με επάρκεια τη συνάρτηση που περιγράφει το προς εξέταση πρόβλημα, οδηγώντας έτσι σε υψηλό σφάλμα στο σύνολο εκπαίδευσης αλλά και στο σύνολο ελέγχου. Από την άλλη, η ύπαρξη μεγάλου αριθμού κρυφών επιπέδων αυξάνει την πολυπλοκότητα του μοντέλου και ενδεχομένως να οδηγήσει σε υπερπροσαρμογή (overfitting) στα δεδομένα εκπαίδευσης, δυσκολίες στη γενίκευση σε νέα άγνωστα δεδομένα και αυξημένο υπολογιστικό κόστος.

Δεν υπάρχει κάποιος γενικός θεωρητικός κανόνας για την εύρεση του βέλτιστου αριθμού κρυφών νευρώνων. Χρησιμοποιούνται ωστόσο εμπειρικοί κανόνες ο οποίοι παρέχουν αρχικές εκτιμήσεις για τον σχεδιασμό της αρχιτεκτονικής.

- Σύμφωνα με εμπειρικές προσεγγίσεις που αναφέρονται στη βιβλιογραφία προτείνεται να βρίσκεται ο αριθμός των κρυφών νευρώνων μεταξύ του πλήθους των νευρώνων εισόδου και του πλήθους των νευρώνων εξόδου. Συγκεκριμένα, προτείνεται ο αριθμός των κρυφών νευρώνων να ισούται με τα $2/3$ του αθροίσματος του πλήθους των νευρώνων εισόδου και του πλήθους των νευρώνων εξόδου.
- Άλλος εμπειρικός κανόνας προτείνει ότι το πλήθος των κρυφών νευρώνων θα πρέπει να είναι μικρότερος από το διπλάσιο του πλήθους των νευρώνων εισόδου συν μια μονάδα και ταυτόχρονα μεγαλύτερος από 2 [76].
- Επιπλέον, έχει προταθεί ότι το πλήθος των κρυφών νευρώνων μπορεί να προσεγγιστεί από την τετραγωνική ρίζα του γινομένου του πλήθους των νευρώνων εισόδου και του πλήθους των νευρώνων εξόδου [76].

Τέλος, σε κάποιες περιπτώσεις προτείνεται το πλήθος των κρυφών νευρώνων να ισούται με τα $2/3$ του πλήθους των νευρώνων εισόδου.

Οι παραπάνω εμπειρικοί κανόνες δεν εγγυώνται την επίτευξη εύρεσης της βέλτιστης αρχιτεκτονικής, χρησιμοποιούνται ωστόσο ως αφετηρία κατά τη διαδικασία σχεδιασμού. Η τελική επιλογή της αρχιτεκτονικής γίνεται συνήθως μέσω επαναληπτικής εκπαίδευσης και αξιολόγησης του Νευρωνικού Δικτύου, με βάση την απόδοση του Δικτύου στο σύνολο της εκπαίδευσης αλλά και σε ανεξάρτητα σύνολα ελέγχου.

3.2.11. Κριτήρια τερματισμού

Κατά τη διαδικασία εκπαίδευσης ενός Νευρωνικού Δικτύου απαιτείται να οριστούν τα κριτήρια τερματισμού για την αποφυγή της υπερεκπαίδευσης και της εξοικονόμησης υπολογιστικού χρόνου.

Συνήθη κριτήρια τερματισμού αποτελούν:

Μέγιστος αριθμός επαναλήψεων – εποχών 9 epochs): γίνεται τερματισμός όταν έχει επιτευχθεί τιμή σφάλματος μικρότερη από ένα προκαθορισμένο κατώφλι.

Έλεγχος της διακύμανσης των βαρών: γίνεται τερματισμός όταν υπάρξει ελάχιστη ή αμελητέα μεταβολή του μέσου τετραγωνικού σφάλματος μεταξύ διαδοχικών επαναλήψεων.

Τα παραπάνω κριτήρια επιτρέπουν, σε πρακτικό επίπεδο, τον έλεγχο της διαδικασίας μάθησης με τέτοιο τρόπο ώστε να εξασφαλίζεται ότι το Δίκτυο έχει φτάσει σε ικανοποιητικό επίπεδο προσαρμογής, χωρίς ταυτόχρονα να απαιτείται περιττό υπολογιστικό κόστος [75].

Κεφάλαιο 4ο

4. Συνελικτικά Νευρωνικά Δίκτυα

4.1. Η Θέση των Συνελικτικών Νευρωνικών Δικτύων στο Πεδίο της Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI) αποτελεί έναν ευρύ διεπιστημονικό κλάδο της επιστήμης των υπολογιστών, ο οποίος στοχεύει στη δημιουργία συστημάτων ικανών να εκτελούν εργασίες που παραδοσιακά απαιτούν ανθρώπινη γνωστική λειτουργία, όπως η αντίληψη, η λήψη αποφάσεων και η μάθηση. Ιστορικά, οι πρώιμες προσεγγίσεις AI βασίζονταν κυρίως σε συμβολικά συστήματα και «έμπειρα συστήματα» (expert systems), τα οποία επιχειρούσαν να κωδικοποιήσουν ρητά την ανθρώπινη γνώση σε ένα σύνολο κανόνων.

Η Μηχανική Μάθηση (Machine Learning - ML) αναδύθηκε ως ένα ισχυρό υποπεδίο της AI, επιφέροντας μια θεμελιώδη αλλαγή παραδείγματος. Αντί για ρητό προγραμματισμό κανόνων, η ML εστιάζει στην ανάπτυξη αλγορίθμων που μπορούν να «μαθαίνουν» αυτόματα από τα δεδομένα[77][78]. Σε αυτό το πλαίσιο, τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANNs), υπολογιστικά μοντέλα εμπνευσμένα από τη δομή και τη λειτουργία του βιολογικού εγκεφάλου, προσέφεραν μια ισχυρή μεθοδολογία για την προσέγγιση πολύπλοκων, μη-γραμμικών προβλημάτων[77][79].

Η τρέχουσα «χρυσή εποχή» της AI τροφοδοτείται σε μεγάλο βαθμό από τη Βαθιά Μάθηση (Deep Learning - DL), έναν εξειδικευμένο τομέα της ML που χρησιμοποιεί ANNs με πολλαπλά επίπεδα επεξεργασίας (layers). Αυτή η «βαθιά» αρχιτεκτονική επιτρέπει στα μοντέλα να μαθαίνουν αυτόματα ιεραρχικές αναπαραστάσεις των δεδομένων. Σε αντίθεση με τις παραδοσιακές μεθόδους ML, που απαιτούσαν εξειδικευμένη γνώση και χειροκίνητη σχεδίαση χαρακτηριστικών (hand-crafted feature extraction), τα μοντέλα DL μαθαίνουν τα βέλτιστα χαρακτηριστικά απευθείας από τα ακατέργαστα δεδομένα[79].

Στην καρδιά της επανάστασης της Βαθιάς Μάθησης, ειδικά στο πεδίο της υπολογιστικής όρασης, βρίσκονται τα **Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs)**[80] [81]. Τα CNNs αποτελούν μια εξειδικευμένη κατηγορία «βαθιών» δικτύων, η αρχιτεκτονική των οποίων είναι άμεσα εμπνευσμένη από τον βιολογικό οπτικό φλοιό. Η καινοτομία τους έγκειται στη χρήση ειδικών επιπέδων, των **συνελικτικών επιπέδων (convolutional layers)**, τα οποία εφαρμόζουν μαθηματικές πράξεις (συνελίξεις) χρησιμοποιώντας μικρά φίλτρα ή πυρήνες (kernels). Αυτή η προσέγγιση, σε συνδυασμό με την **κοινή χρήση βαρών (weight sharing)** και τα **επίπεδα συγκέντρωσης (pooling layers)**, επιτρέπει στα CNNs να εντοπίζουν ιεραρχικά τοπικά μοτίβα (local patterns) μέσα στα δεδομένα, καθιστώντας τα αναλλοίωτα ως προς τη θέση (translation invariant) και εξαιρετικά αποδοτικά για την ανάλυση δεδομένων με χωρική δομή πλέγματος, όπως οι εικόνες [78][79] [81].

Η εκρηκτική άνοδος των CNNs ξεκίνησε ουσιαστικά το 2012, όταν η αρχιτεκτονική AlexNet κυριάρχησε στον διαγωνισμό ImageNet. Αυτή η επιτυχία, η οποία κατέστη δυνατή χάρη στη διαθεσιμότητα τεράστιων συνόλων δεδομένων (Big Data) και την ωρίμανση των Μονάδων Επεξεργασίας Γραφικών (GPUs), απέδειξε την υπεροχή των CNNs έναντι όλων των προηγούμενων μεθόδων υπολογιστικής όρασης και πυροδότησε έναν καταιγισμό ερευνών που οδήγησε σε ολοένα και βαθύτερες και πολυπλοκότερες αρχιτεκτονικές (π.χ. VGG, GoogLeNet, ResNet) [78] [79].

4.2. Η Σημασία των CNNs στην Ιατρική Απεικόνιση και την Ιστοπαθολογία

Η ικανότητα των CNNs να εξάγουν αυτόματα πολύπλοκα και αφηρημένα χαρακτηριστικά από τις εικόνες, τα κατέστησε ιδανικό εργαλείο για τον τομέα της ιατρικής απεικόνισης [82] [83]. Αλγόριθμοι βασισμένοι σε CNNs εφαρμόζονται πλέον με εξαιρετικά αποτελέσματα στην ανάλυση ακτινογραφιών, αξονικών (CT) και μαγνητικών (MRI) τομογραφιών, βοηθώντας στην ταχύτερη και ακριβέστερη διάγνωση παθήσεων [81] [84] [85].

Ένα πεδίο όπου η επίδραση των CNNs αναμένεται να είναι μεταμορφωτική είναι η **Ψηφιακή Ιστοπαθολογία** [86] [87]. Η ιστοπαθολογική εξέταση, δηλαδή η μικροσκοπική ανάλυση δειγμάτων ιστού (βιοψιών), παραμένει ο «χρυσός κανόνας» (gold standard) για

την οριστική διάγνωση, την ταξινόμηση και τη σταδιοποίηση του καρκίνου. Ωστόσο, η διαδικασία αυτή αντιμετωπίζει σημαντικές προκλήσεις. Πρώτον, η ψηφιοποίηση των πλακιδίων δημιουργεί Εικόνες Ολόκληρης Τομής (Whole Slide Images - WSIs), οι οποίες είναι τεράστιες σε μέγεθος (συντά σε κλίμακα gigapixel), καθιστώντας την υπολογιστική τους ανάλυση εξαιρετικά απαιτητική [86][87]. Δεύτερον, η διάγνωση εξαρτάται σε μεγάλο βαθμό από την υποκειμενική εκτίμηση και την εμπειρία του παθολογοανατόμου, με αποτέλεσμα να παρατηρείται σημαντική μεταβλητότητα μεταξύ διαφορετικών παρατηρητών (inter-observer variability). Τέλος, ο διαγνωστικός φόρτος εργασίας αυξάνεται συνεχώς, καθιστώντας την ταχεία και ακριβή ανάλυση δυσχερή.

Τα CNNs προσφέρουν μια ισχυρή λύση σε αυτές τις προκλήσεις. Στον τομέα της ιστοπαθολογίας του **καρκίνου του παχέος εντέρου (Colorectal Cancer - CRC)**, ο οποίος αποτελεί μία από τις κύριες αιτίες θανάτου από καρκίνο παγκοσμίως, [88] τα CNNs δεν χρησιμοποιούνται απλώς για να αυτοματοποιήσουν την ταξινόμηση μεταξύ καλοήθους και κακοήθους ιστού. Πλέον, αλγόριθμοι Βαθιάς Μάθησης μπορούν να αναλύσουν τις εικόνες H&E (Αιματοξυλίνης & Ηωσίνης) για να επιτύχουν πολύ πιο σύνθετους στόχους[88] [87]:

1. **Ανίχνευση και Τμηματοποίηση:** Εντοπισμός και ακριβής οριοθέτηση καρκινικών περιοχών, αδενικών δομών (glands) ή μεμονωμένων κυττάρων[88] [84].
2. **Πρόγνωση:** Αναγνώριση ιστολογικών χαρακτηριστικών που σχετίζονται με την πρόγνωση, όπως η εκβλάστηση του όγκου (tumor budding) ή η πρόβλεψη πενταετούς επιβίωσης απευθείας από την εικόνα [87].
3. **Πρόβλεψη Βιοδεικτών:** Πρόβλεψη της μοριακής κατάστασης του όγκου, όπως η Μικροδορυφορική Αστάθεια (Microsatellite Instability - MSI) ή μεταλλάξεις γονιδίων (π.χ. KRAS, BRAF), απευθείας από την εικόνα H&E, χωρίς την ανάγκη δαπανηρών μοριακών εξετάσεων [88] [87].

Η παρούσα διπλωματική εργασία εντάσσεται σε αυτό το ταχέως αναπτυσσόμενο πεδίο, με κεντρικό άξονα την «Αξιοποίηση αρχιτεκτονικών Συνελικτικών Νευρωνικών Δικτύων για τον εντοπισμό ευρημάτων σε βιοψίες παχέος εντέρου».

4.3. Στόχοι και Δομή Κεφαλαίου

Στόχος του παρόντος κεφαλαίου είναι να παράσχει το απαραίτητο θεωρητικό υπόβαθρο για την κατανόηση των Συνελικτικών Νευρωνικών Δικτύων, τα οποία αποτελούν τον πυρήνα της μεθοδολογίας που αναπτύχθηκε. Η δομή του κεφαλαίου ακολουθεί τρεις βασικούς άξονες:

1. **Ανάλυση των βασικών αρχών των CNNs:** Θα εξεταστούν τα θεμελιώδη δομικά στοιχεία που συνθέτουν μια τυπική αρχιτεκτονική CNN. Αυτό περιλαμβάνει μια λεπτομερή περιγραφή του **συνελικτικού επιπέδου (convolutional layer)**, αναλύοντας τις έννοιες των φίλτρων (filters/kernels), του βήματος (stride) και του γεμίματος (padding). Θα παρουσιαστούν οι **συναρτήσεις ενεργοποίησης (activation functions)**, με έμφαση στην Rectified Linear Unit (ReLU), τα **επίπεδα συγκέντρωσης (pooling layers)** που εκτελούν υποδειγματοληψία (π.χ. Max και Average Pooling), και τα **πλήρως συνδεδεμένα επίπεδα (fully connected layers)** που συνήθως ολοκληρώνουν την εξαγωγή χαρακτηριστικών και οδηγούν στην τελική ταξινόμηση.
2. **Παρουσίαση προχωρημένων αρχιτεκτονικών (ResNet, R-CNN):** Πέρα από τις βασικές δομές, η σύγχρονη Βαθιά Μάθηση βασίζεται σε πιο εξελιγμένες αρχιτεκτονικές που επιλύουν συγκεκριμένα προβλήματα. Θα δοθεί έμφαση σε δύο οικογένειες μοντέλων που είναι κρίσιμες για την ιατρική εικόνα:
 - **ResNet (Residual Networks):** Μια αρχιτεκτονική-ορόσημο που εισήγαγε τις «παρακαμπτήριες συνδέσεις» (skip/identity connections). Αυτή η καινοτομία επιτρέπει την εκπαίδευση εξαιρετικά βαθιών δικτύων (π.χ. 50, 101, ή 152 επιπέδων) αντιμετωπίζοντας αποτελεσματικά το **πρόβλημα της εξαφανιζόμενης κλίσης (vanishing gradient problem)**.
 - **R-CNN (Region-based CNNs):** Μια οικογένεια αρχιτεκτονικών (R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN) σχεδιασμένη ειδικά για εργασίες **ανίχνευσης αντικειμένων (object detection)** και **τμηματοποίησης (segmentation)**. Αυτό είναι θεμελιώδες για τον στόχο της παρούσας εργασίας, που είναι ο «εντοπισμός ευρημάτων» και όχι απλώς η ταξινόμηση ολόκληρων εικόνων.
3. **Εφαρμογές σε Ιατρικές Εικόνες:** Τέλος, το κεφάλαιο θα συνδέσει τις θεωρητικές αρχιτεκτονικές (ResNet, R-CNN) με την πρακτική τους εφαρμογή στην ανάλυση βιοϊατρικών εικόνων. Θα συζητηθεί πώς αυτά τα μοντέλα αξιοποιούνται συγκεκριμένα σε ιστοπαθολογικές εικόνες για την τμηματοποίηση ιστών, την ανίχνευση καρκινικών

περιοχών και την ταξινόμηση ευρημάτων σε σύνολα δεδομένων παχέος εντέρου, όπως τα Kather et al. CRC dataset και TCGA.

4.4. Ιστορική αναδρομή και εξέλιξη των CNN

4.4.1. Από το Perceptron στο Neocognitron

Η πορεία προς τα σύγχρονα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) είναι μακρά και ξεκινά δεκαετίες πριν την πρόσφατη επανάσταση της Βαθιάς Μάθησης. Η ίδια η έννοια του Τεχνητού Νευρωνικού δικτύου (ANN) βασίζεται στην προσπάθεια προσομοίωσης της υπολογιστικής δομής του ανθρώπινου εγκεφάλου[77] [79].

Η πρώτη θεμελιώδης εργασία που πρότεινε ένα μαθηματικό μοντέλο για έναν βιολογικό νευρώνα δημοσιεύτηκε το 1943 από τους McCulloch και Pitts. Το μοντέλο τους ήταν ένας απλός λογικός κατωφλιακός επεξεργαστής (threshold logic unit), ικανός να εκτελεί βασικές λογικές πράξεις, αλλά του έλειπε ένας κρίσιμος μηχανισμός: η ικανότητα της μάθησης [78].

Αυτό το κενό ήρθε να καλύψει ο Frank Rosenblatt το 1957 (και δημοσιεύτηκε το 1958) με την εφεύρεση του **Perceptron** (Αντιληπτικού). Το Perceptron ήταν το πρώτο μοντέλο που εισήγαγε έναν αλγόριθμο εκπαίδευσης (τον κανόνα μάθησης perceptron). Ήταν ένα σύστημα ενός επιπέδου ικανό να μαθαίνει πώς να ταξινομεί δείγματα σε δύο κατηγορίες. Η μεγάλη του αδυναμία, ωστόσο, ήταν ότι μπορούσε να επιλύσει μόνο προβλήματα που ήταν **γραμμικά διαχωρίσιμα** (linearly separable)[78] [79].

Η αδυναμία αυτή αναδείχθηκε με έμφαση το 1969 από τους Minsky και Papert στο βιβλίο τους "Perceptrons", όπου απέδειξαν μαθηματικά ότι το απλό Perceptron αδυνατούσε να λύσει ακόμα και τη βασική λογική πράξη XOR. Η δημοσίευση αυτή είχε καταστροφικές συνέπειες, οδηγώντας σε δραματική μείωση της χρηματοδότησης και του ερευνητικού ενδιαφέροντος για τα νευρωνικά δίκτυα, εγκαινιάζοντας τον πρώτο «Χειμώνα της AI» .

Παρά το ακαδημαϊκό κλίμα, η έρευνα για την προσομοίωση της όρασης συνεχίστηκε. Μια κρίσιμη εξέλιξη, που έθεσε τα θεμέλια για τη σύγχρονη αρχιτεκτονική των CNNs, ήρθε από την Ιαπωνία. Το 1980 (και επακόλουθες εργασίες όπως το 1988), ο Kunihiro Fukushima ανέπτυξε το **Neocognitron**, ένα ιεραρχικό, πολυεπίπεδο νευρωνικό δίκτυο σχεδιασμένο ειδικά για την αναγνώριση οπτικών προτύπων, όπως οι χειρόγραφοι

χαρακτήρες [80] [77]. Το Neocognitron ήταν άμεσα εμπνευσμένο από τις νευροεπιστημονικές μελέτες των Hubel και Wiesel πάνω στον οπτικό φλοιό. Εισηγάγα τις δύο θεμελιώδεις αρχές των σύγχρονων CNNs:

1. **Τοπικά Πεδία Αντίληψης (Local Receptive Fields):** Οι νευρώνες σε ένα επίπεδο συνδέονταν μόνο με μια μικρή, τοπική περιοχή του προηγούμενου επιπέδου.
2. **Ιεραρχική Επεξεργασία με Υποδειγματοληψία:** Το δίκτυο αποτελούνταν από εναλλασσόμενα επίπεδα που εξήγαγαν χαρακτηριστικά (S-cells, αντίστοιχα των συνελκτικών επιπέδων) και επίπεδα που μείωναν τη χωρική ανάλυση, παρέχοντας ανοχή σε μικρές μετατοπίσεις (C-cells, αντίστοιχα των pooling layers) [79] [81].

Το Neocognitron ήταν πρωτοποριακό, αλλά του έλειπε ένας αποδοτικός αλγόριθμος εκπαίδευσης για ολόκληρη τη βαθιά δομή.

4.4.2. LeNet-5: Η Γέννηση του Σύγχρονου CNN

Η λύση στο πρόβλημα εκπαίδευσης των πολυεπίπεδων δικτύων ήρθε στα μέσα της δεκαετίας του 1980 με την εκ νέου ανακάλυψη και διάδοση του αλγορίθμου **Οπίσθιας Διάδοσης (Backpropagation)** από τους Rumelhart, Hinton και Williams (1986) [78] [79]. Το Backpropagation παρείχε έναν υπολογιστικά αποδοτικό τρόπο για τη χρήση της κατάβασης κλίσης (gradient descent) για την εκπαίδευση πολυεπίπεδων δικτύων (MLPs), επιλύοντας το πρόβλημα που είχε θέσει το Perceptron.

Ο Yann LeCun και η ομάδα του στα Bell Labs ήταν οι πρώτοι που συνδύασαν τις ιεραρχικές δομές του Neocognitron με τον αλγόριθμο εκπαίδευσης Backpropagation. Το αποτέλεσμα, που παρουσιάστηκε οριστικά το 1998, ήταν το **LeNet-5**. Αυτή θεωρείται η πρώτη σύγχρονη αρχιτεκτονική CNN [89][81].

Το LeNet-5 σχεδιάστηκε ειδικά για την αναγνώριση χειρόγραφων ψηφίων από το σύνολο δεδομένων MNIST. Η αρχιτεκτονική του καθιέρωσε τη δομή που ακολουθείται μέχρι σήμερα: μια ακολουθία από **συνελκτικά επίπεδα** (Convolution) και **επίπεδα υποδειγματοληψίας** (Subsampling, ή Pooling), ακολουθούμενη από ένα ή περισσότερα **πλήρως συνδεδεμένα επίπεδα** (Fully Connected - FC) στο τέλος για την τελική ταξινόμηση.

Παρά την τεράστια επιτυχία του στην αναγνώριση ψηφίων (που οδήγησε στην εμπορική του χρήση για την ανάγνωση επιταγών), το LeNet-5 και τα CNNs γενικότερα, δεν

κατάφεραν να κυριαρχήσουν άμεσα. Ο λόγος ήταν διπλός: πρώτον, η εκπαίδευσή τους ήταν υπολογιστικά πολύ απαιτητική για την τεχνολογία της εποχής και, δεύτερον, για πιο σύνθετα προβλήματα όρασης, ανταγωνίζονταν ισχυρές μεθόδους μηχανικής μάθησης όπως οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVMs) [78] [79].

4.4.3. Η Επανάσταση της Βαθιάς Μάθησης: AlexNet, Big Data και GPUs

Η πραγματική επανάσταση για τα CNNs ήρθε το 2012. Μια ομάδα από το Πανεπιστήμιο του Τορόντο, με επικεφαλής τους Alex Krizhevsky, Ilya Sutskever και Geoffrey Hinton, παρουσίασε την αρχιτεκτονική **AlexNet** [78] [89]. Συμμετέχοντας στον διαγωνισμό ImageNet Large Scale Visual Recognition Challenge (ILSVRC), το AlexNet πέτυχε ποσοστό σφάλματος 16.4%, συντρίβοντας τον ανταγωνισμό, του οποίου η επόμενη καλύτερη επίδοση (βασισμένη σε παραδοσιακές μεθόδους υπολογιστικής όρασης) ήταν 26.2%. Αυτή η στιγμή θεωρείται ευρέως ως η «Μεγάλη Έκρηξη» της Βαθιάς Μάθησης.

Η επιτυχία του AlexNet δεν οφειλόταν σε μία μόνο καινοτομία, αλλά στη σύγκλιση τριών κρίσιμων παραγόντων, της «Αγίας Τριάδας» της σύγχρονης AI :

1. **Διαθεσιμότητα Υπολογιστικής Ισχύος (GPU Acceleration):** Η αρχιτεκτονική του AlexNet ήταν πολύ μεγαλύτερη από το LeNet (61 εκατομμύρια παράμετροι έναντι 431 χιλιάδων του LeNet-5). Η εκπαίδευσή του θα ήταν αδύνατη χωρίς την αξιοποίηση των **Μονάδων Επεξεργασίας Γραφικών (GPUs)**. Τα CNNs, από τη φύση τους (μαζικοί παραλληλισμοί πινάκων και πράξεις συνέλιξης), είναι ιδανικά για την αρχιτεκτονική των GPUs. Το AlexNet σχεδιάστηκε ρητά για να εκπαιδευτεί παράλληλα σε δύο GPUs, μειώνοντας δραματικά τον χρόνο εκπαίδευσης[79].
2. **Διαθεσιμότητα Μεγάλων Δεδομένων (Big Data):** Τα βαθιά δίκτυα με εκατομμύρια παραμέτρους είναι επιρρεπή στην υπερπροσαρμογή (overfitting). Για να γενικεύσουν σωστά, απαιτούν τεράστιους όγκους επισημειωμένων δεδομένων. Το σύνολο δεδομένων **ImageNet**, με πάνω από 1 εκατομμύριο εικόνες υψηλής ανάλυσης ταξινομημένες σε 1000 κατηγορίες, παρείχε τα απαραίτητα δεδομένα εκπαίδευσης.
3. **Αλγοριθμικές Βελτιώσεις:** Το AlexNet εισήγαγε (ή μάλλον, διέδωσε) δύο τεχνικές που είναι πλέον θεμελιώδεις:

- Τη συνάρτηση ενεργοποίησης **ReLU (Rectified Linear Unit)** , η οποία επέλυσε το πρόβλημα κορεσμού και εξαφανιζόμενης κλίσης (vanishing gradient) που παρουσίαζαν οι παλαιότερες συναρτήσεις (όπως η sigmoid ή η tanh) [90] [81].
- Την τεχνική κανονικοποίησης **Dropout** , κατά την οποία ένα ποσοστό νευρώνων απενεργοποιείται τυχαία σε κάθε βήμα εκπαίδευσης, αποτρέποντας την υπερβολική αλληλεξάρτηση μεταξύ τους και βελτιώνοντας τη γενίκευση[78] [79].

4.4.4. Η πορεία προς το Βάθος και την Αποδοτικότητα

Μετά την επιτυχία του AlexNet, η ακαδημαϊκή και εμπορική έρευνα επικεντρώθηκε στη δημιουργία συνεχώς και πιο βαθιών και αποδοτικών αρχιτεκτονικών.

- **VGGNet (2014):** Αναπτύχθηκε από το Visual Geometry Group (VGG) της Οξφόρδης και ήταν ο επιλαχών στον διαγωνισμό ILSVRC 2014. Η φιλοσοφία του VGG ήταν η **ομοιομορφία και το βάθος**. Αντί για τα μεγάλα φίλτρα (11x11, 7x7) των προηγούμενων μοντέλων, το VGG χρησιμοποίησε αποκλειστικά πολύ μικρά φίλτρα **3x3** στοιβαγμένα το ένα πάνω στο άλλο. Απέδειξαν ότι μια στοιβία από 3x3 φίλτρα μπορεί να έχει το ίδιο αποτελεσματικό πεδίο αντίληψης με ένα μεγαλύτερο φίλτρο, αλλά με περισσότερη μη-γραμμικότητα και λιγότερες παραμέτρους . Οι εκδόσεις VGG-16 και VGG-19 (με 16 και 19 επίπεδα αντίστοιχα) έγιναν εξαιρετικά δημοφιλείς, αν και ήταν υπολογιστικά βαριά (το VGG-16 έχει ~138M παραμέτρους)[78][79].
- **GoogLeNet/Inception(2014):** Ο νικητής του ILSVRC 2014, που αναπτύχθηκε από την Google. Η βασική του καινοτομία ήταν η **υπολογιστική αποδοτικότητα**. Αντί να αυξάνει απλώς το βάθος, εισήγαγε τη «μονάδα Έναρξης» (**Inception module**) . Αυτή η μονάδα εκτελεί συνελίξεις με διαφορετικά μεγέθη φίλτρων (1x1, 3x3, 5x5) και pooling παράλληλα και έπειτα συνενώνει τα αποτελέσματά τους . Για να μειώσει το τεράστιο υπολογιστικό κόστος αυτής της παραλληλίας, χρησιμοποίησε εκτενώς **συνελίξεις 1x1** ως «λαιμό μπουκαλιού» (bottleneck) για τη μείωση των διαστάσεων (dimensionality reduction) πριν από τις πιο ακριβές 3x3 και 5x5 συνελίξεις . Το αποτέλεσμα ήταν ένα δίκτυο 22 επιπέδων με μόλις 7 εκατομμύρια παραμέτρους, πολύ λιγότερες από το AlexNet[78] [79].

- **ResNet (2015):** Ο νικητής του ILSVRC 2015, το ResNet (Residual Network) από τη Microsoft Research, έφερε την επόμενη μεγάλη τομή. Καθώς τα δίκτυα γίνονταν βαθύτερα (π.χ. 30, 40 επίπεδα), οι ερευνητές αντιμετώπισαν το πρόβλημα της **εξαφανιζόμενης/εκρηγνυόμενης κλίσης (vanishing/exploding gradients)** και το πρόβλημα της «υποβάθμισης» (degradation), όπου το βαθύτερο μοντέλο είχε χειρότερη απόδοση από ένα πιο ρηχό. Το ResNet εισήγαγε τις **παρακαμπτήριες συνδέσεις (skip connections)** ή **υπολειμματικές μονάδες (residual blocks)**. Σε αυτή τη δομή, η είσοδος ενός μπλοκ (x) προστίθεται απευθείας στην έξοδο του ($F(x)$), με αποτέλεσμα η τελική έξοδος να είναι $H(x) = F(x) + x$. Αυτό επιτρέπει στη κλίση να ρέει απρόσκοπτα προς τα πίσω (backpropagate) μέσω της σύνδεσης ταυτότητας (x), επιτρέποντας την επιτυχή εκπαίδευση εξαιρετικά βαθιών δικτύων (152 επιπέδων ή και βαθύτερων) . Το ResNet θεωρείται πλέον η ραχοκοκαλιά πολλών σύγχρονων αρχιτεκτονικών[79][91].
- **EfficientNet (2019):** Μετά το ResNet, η έρευνα μετατοπίστηκε από την απλή αύξηση του βάθους στη βελτιστοποίηση της συνολικής αποδοτικότητας (ακρίβεια έναντι υπολογιστικού κόστους). Το EfficientNet (Tan & Le, 2019) πρότεινε μια νέα μέθοδο, το "**compound scaling**". Αντί να αυξάνουν αυθαίρετα μόνο μία διάσταση του δικτύου (βάθος, πλάτος/κανάλια, ή ανάλυση εικόνας), οι ερευνητές της Google πρότειναν μια αρχή για την ισορροπημένη κλιμάκωση και των τριών διαστάσεων ταυτόχρονα. Χρησιμοποιώντας μια εξαιρετικά αποδοτική βασική αρχιτεκτονική (εμπνευσμένη από το MobileNet), το EfficientNet πέτυχε νέα state-of-the-art αποτελέσματα στο ImageNet με δραματικά λιγότερες παραμέτρους και υπολογισμούς (FLOPS) σε σχέση με τους προκατόχους του[78] [79].

4.5. Θεμελιώδη στοιχεία των Συνελικτικών Νευρωνικών Δικτύων

4.5.1. Δομή ενός CNN

Τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) αποτελούν μια εξειδικευμένη κατηγορία νευρωνικών δικτύων που είναι σχεδιασμένα για την επεξεργασία δεδομένων με χωρική δομή πλέγματος, όπως οι εικόνες. Η αρχιτεκτονική τους είναι εμπνευσμένη από τον βιολογικό οπτικό φλοιό και διαφέρει ριζικά από τα παραδοσιακά Πολυεπίπεδα Αντιληπτικά (MLPs), καθώς δεν αντιμετωπίζει την είσοδο ως ένα μονοδιάστατο διάνυσμα, αλλά διατηρεί τη χωρική της δομή.

Μια τυπική αρχιτεκτονική CNN αποτελείται από μια ακολουθία επιπέδων που σταδιακά μετασχηματίζουν τον αρχικό πίνακα εικονοστοιχείων (pixels) σε ένα σύνολο υψηλού επιπέδου χαρακτηριστικών, τα οποία τελικά τροφοδοτούν έναν ταξινομητή. Αυτή η ιεραρχική εξαγωγή χαρακτηριστικών είναι που καθιστά τα CNNs τόσο ισχυρά. Τα θεμελιώδη δομικά στοιχεία που απαρτίζουν σχεδόν κάθε σύγχρονο CNN είναι:

1. Το Επίπεδο Εισόδου (Input Layer)
2. Το Συνελικτικό Επίπεδο (Convolutional Layer)
3. Η Συνάρτηση Ενεργοποίησης (Activation Function)
4. Το Επίπεδο Συγκέντρωσης (Pooling Layer)
5. Το Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)
6. Το Επίπεδο Εξόδου (Output Layer)

Μια συνηθισμένη διάταξη περιλαμβάνει την επανάληψη των τριών πρώτων επιπέδων (Convolutional → Activation → Pooling) σε βάθος, πριν την τελική ταξινόμηση από τα πλήρως συνδεδεμένα επίπεδα.

4.5.2. Επίπεδο Εισόδου (Input Layer)

Το επίπεδο εισόδου είναι το σημείο εκκίνησης του δικτύου. Ο ρόλος του είναι να υποδεχθεί τα ακατέργαστα δεδομένα της εικόνας και να τα προετοιμάσει για επεξεργασία. Σε αντίθεση με τα MLPs που απαιτούν "ισοπέδωση" (flattening) της εικόνας σε ένα μονοδιάστατο διάνυσμα, τα CNNs δέχονται την είσοδο στη φυσική της μορφή, διατηρώντας τη χωρική σχέση μεταξύ των εικονοστοιχείων [79] [89].

Τα δεδομένα εισόδου αναπαρίστανται σχεδόν πάντα ως ένας πολυδιάστατος πίνακας (τανυστής / tensor) με διαστάσεις $H \times W \times D$ [78][79]:

- **H (Height):** Το ύψος της εικόνας σε pixels.
- **W (Width):** Το πλάτος της εικόνας σε pixels.
- **D (Depth):** Το βάθος, που αντιστοιχεί στον αριθμό των καναλιών.
 - **Grayscale (Ασπρόμαυρες):** Για εικόνες σε κλίμακα του γκρι, $D = 1$.
 - **RGB (Έγχρωμες):** Για τυπικές έγχρωμες εικόνες, $D = 3$, όπου κάθε κανάλι αντιστοιχεί στο Κόκκινο, το Πράσινο και το Μπλε.
 - **3D / Πολυφασματικές:** Στην ιατρική απεικόνιση, το βάθος μπορεί να αντιπροσωπεύει κάτι διαφορετικό, όπως τις διαδοχικές τομές (slices) σε μια 3D μαγνητική τομογραφία (MRI), ή διαφορετικά φασματικά κανάλια [83] [84].

Κατά την εισαγωγή, τα δεδομένα συχνά υφίστανται **κανονικοποίηση (normalization)**, όπως η αφαίρεση της μέσης τιμής (mean subtraction) ή η διαίρεση με την τυπική απόκλιση, μια διαδικασία που επιταχύνει τη σύγκλιση κατά την εκπαίδευση [79][81].

4.5.3. Συνελικτικό Επίπεδο (Convolutional Layer)

Το συνελικτικό επίπεδο είναι ο πυρήνας (core building block) του CNN και εκτελεί το μεγαλύτερο μέρος των υπολογισμών. Η λειτουργία του είναι να ανιχνεύει τοπικά χαρακτηριστικά (local features) στην εικόνα εισόδου. Αυτό επιτυγχάνεται μέσω της μαθηματικής πράξης της **συνέλιξης (convolution)** [80][81].

Τα βασικά στοιχεία του επιπέδου αυτού είναι:

- **Φίλτρα (Filters) ή Πυρήνες (Kernels):** Αυτά είναι τα «μάτια» του δικτύου. Ένα φίλτρο είναι ένας μικρός, δισδιάστατος πίνακας βαρών (π.χ., 3×3 ή 5×5) τα

οποία είναι *μαθησίσιμα* (learnable). Κάθε φίλτρο εκπαιδεύεται για να «ενεργοποιείται» (να παράγει υψηλή τιμή) όταν ανιχνεύει ένα συγκεκριμένο χαρακτηριστικό, όπως μια κάθετη ακμή, μια οριζόντια γραμμή, μια καμπύλη ή ένα συγκεκριμένο χρώμα[78][79].

- **Πράξη Συνέλιξης:** Το φίλτρο «γλιστράει» (slides) πάνω σε ολόκληρη την εικόνα εισόδου, βήμα προς βήμα. Σε κάθε θέση, υπολογίζεται το **εσωτερικό γινόμενο (dot product)** μεταξύ των τιμών του φίλτρου και των τιμών των εικονοστοιχείων της εικόνας που καλύπτει εκείνη τη στιγμή[79][81].
- **Χάρτης Ενεργοποίησης (Activation Map / Feature Map):** Ο 2D πίνακας που προκύπτει από την εφαρμογή ενός φίλτρου σε ολόκληρη την εικόνα ονομάζεται χάρτης χαρακτηριστικών ή χάρτης ενεργοποίησης¹⁸¹⁸¹⁸¹⁸. Ουσιαστικά, είναι μια αναπαράσταση του *πού* στην αρχική εικόνα εντοπίστηκε το χαρακτηριστικό που ανιχνεύει το φίλτρο [79][89].
- **Βάθος Εξόδου (Output Depth):** Ένα μεμονωμένο συνελκτικό επίπεδο δεν μαθαίνει μόνο ένα φίλτρο, αλλά έναν ολόκληρο *στόλο* από φίλτρα (π.χ., 32, 64 ή 128)²⁰²⁰²⁰²⁰. Κάθε φίλτρο παράγει τον δικό του χάρτη χαρακτηριστικών. Αυτοί οι χάρτες στοιβάζονται όλοι μαζί για να δημιουργήσουν τον τελικό όγκο εξόδου. Επομένως, το βάθος του όγκου εξόδου ενός συνελκτικού επιπέδου είναι ίσο με τον αριθμό των φίλτρων που χρησιμοποιήθηκαν[80][81].

Για τον έλεγχο της λειτουργίας και του μεγέθους εξόδου του συνελκτικού επιπέδου, ορίζονται τρεις βασικές **υπερ-παράμετροι** [78][79]:

1. **Μέγεθος Πυρήνα (Kernel Size):** Οι διαστάσεις (Height x Width) του φίλτρου. Σύγχρονες αρχιτεκτονικές, όπως το VGG, έχουν αποδείξει ότι η χρήση μικρών φίλτρων (π.χ., 3 X 3) σε διαδοχικά επίπεδα είναι πιο αποδοτική από τη χρήση ενός μεγάλου φίλτρου (π.χ., 7 X 7)[79].
2. **Βήμα (Stride):** Καθορίζει πόσα εικονοστοιχεία μετακινείται το φίλτρο σε κάθε βήμα πάνω στην εικόνα. Ένα βήμα 1 (Stride=1) είναι το πιο συνηθισμένο. Ένα βήμα 2 (Stride=2) θα έχει ως αποτέλεσμα τη μείωση του μεγέθους (subsampling) του χάρτη χαρακτηριστικών.[78]
3. **Γέμισμα (Padding):** Αναφέρεται στην προσθήκη εικονοστοιχείων (συνήθως με τιμή μηδέν) γύρω από το περίγραμμα της εικόνας εισόδου²⁵²⁵²⁵²⁵²⁵²⁵²⁵²⁵. Αυτό γίνεται για δύο λόγους[78][79]:

- **Controlling Output Size:** Χωρίς padding, η εικόνα συρρικνώνεται μετά από κάθε συνέλιξη. Χρησιμοποιώντας "same" padding (όπου $P = (F-1)/2$), διασφαλίζουμε ότι ο χάρτης χαρακτηριστικών εξόδου έχει το ίδιο ύψος και πλάτος με τον χάρτη εισόδου .
- **Border Pixels:** Χωρίς padding, τα εικονοστοιχεία στις γωνίες και τις άκρες της εικόνας "βλέπονται" από το φίλτρο πολύ λιγότερες φορές από τα κεντρικά pixels, με αποτέλεσμα την απώλεια πληροφορίας.

Η τελική διάσταση εξόδου (O) ενός συνελκτικού επιπέδου υπολογίζεται από τον τύπο:

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

όπου N είναι η διάσταση εισόδου (π.χ. ύψος), F είναι το μέγεθος του φίλτρου, P είναι το padding και S είναι το βήμα[78][79].

4.5.4. Συναρτήσεις Ενεργοποίησης (Activation Functions)

Αμέσως μετά από κάθε συνελκτικό επίπεδο (και μετά την πρόσθεση της πόλωσης/bias), εφαρμόζεται μια συνάρτηση ενεργοποίησης. Ο κρίσιμος ρόλος αυτών των συναρτήσεων είναι η εισαγωγή **μη-γραμμικότητας (non-linearity)** στο δίκτυο[78][79]. Χωρίς μη-γραμμικές συναρτήσεις, μια βαθιά στοίβα από συνελκτικά επίπεδα θα ήταν μαθηματικά ισοδύναμη με ένα μόνο, ρηχό συνελκτικό επίπεδο, καθιστώντας το δίκτυο ανίκανο να μοντελοποιήσει πολύπλοκες σχέσεις[89].

Ενώ παλαιότερα χρησιμοποιούνταν συναρτήσεις όπως η **sigmoid** ή η **υπερβολική εφαπτομένη (tanh)**, τα σύγχρονα CNNs κυριαρχούνται από την οικογένεια των **Rectified Linear Units (ReLU)**[90][81].

- **ReLU (Rectified Linear Unit):** Είναι η πιο δημοφιλής συνάρτηση ενεργοποίησης σήμερα. Η εξίσωσή της είναι εξαιρετικά απλή: $f(x) = \max(0, x)$.

Πλεονεκτήματα:

1. Επιλύει το **πρόβλημα της εξαφανιζόμενης κλίσης (vanishing gradient problem)**. Σε αντίθεση με τις σιγμοειδείς συναρτήσεις που έχουν κλίση (παράγωγο) σχεδόν μηδέν για μεγάλες θετικές ή αρνητικές τιμές, η ReLU έχει σταθερή παράγωγο ίση με 1 για όλες τις θετικές εισόδους, επιτρέποντας στη κλίση να ρέει απρόσκοπτα προς τα πίσω[79][81].
2. Είναι **υπολογιστικά ταχύτερη**, καθώς απαιτεί μόνο μια απλή σύγκριση με το μηδέν.

3. Προκαλεί **αραιότητα (sparsity)** στις ενεργοποιήσεις, καθώς όλοι οι αρνητικοί νευρώνες μηδενίζονται, καθιστώντας το δίκτυο πιο αποδοτικό[90].

Μειονέκτημα: Μπορεί να οδηγήσει στο φαινόμενο "Dying ReLU", όπου ένας νευρώνας που παίρνει μόνιμα αρνητική είσοδο "πεθαίνει" και σταματά να μαθαίνει (καθώς η παράγωγός του γίνεται 0)[90].

- **Leaky ReLU (LReLU):** Μια βελτίωση της ReLU που λύνει το παραπάνω πρόβλημα.
 - **Λειτουργία:** $f(x) = x$ για $x > 0$ και $f(x) = ax$ για $x \leq 0$, όπου a είναι μια μικρή σταθερά (π.χ., 0.01).
 - **Πλεονέκτημα:** Επιτρέπει μια μικρή, μη-μηδενική κλίση για αρνητικές τιμές, εμποδίζοντας τους νευρώνες να "πεθάνουν".
- **ELU (Exponential Linear Unit):** Μια άλλη εναλλακτική της ReLU.
 - **Λειτουργία:** $f(x) = x$ για $x > 0$ και $f(x) = \alpha(e^x - 1)$ για $x \leq 0$.
 - **Πλεονέκτημα:** Όπως η LReLU, επιτρέπει αρνητικές τιμές, αλλά τείνει να ωθεί τη μέση ενεργοποίηση προς το μηδέν, κάτι που, σύμφωνα με τους συγγραφείς της, επιταχύνει τη σύγκλιση[90].

4.5.5. Επίπεδο Συγκέντρωσης (Pooling Layer)

Τα επίπεδα συγκέντρωσης (γνωστά και ως επίπεδα **υποδειγματοληψίας/ subsampling**) είναι ένα ακόμη κρίσιμο στοιχείο της αρχιτεκτονικής των CNN. Συνήθως τοποθετούνται αμέσως μετά από ένα συνελκτικό επίπεδο (ή ένα ζεύγος συνελκτικών επιπέδων)[80][81].

Ο κύριος σκοπός τους είναι η **προοδευτική μείωση του χωρικού μεγέθους (ύψος και πλάτος)** των χαρτών χαρακτηριστικών[79].

Αυτό επιτυγχάνει τρεις στόχους:

1. **Μείωση Παραμέτρων:** Μειώνοντας τις διαστάσεις, ελαττώνεται ο αριθμός των παραμέτρων και ο υπολογιστικός φόρτος στα επόμενα επίπεδα[80][78].
2. **Έλεγχος Υπερ-προσαρμογής (Overfitting):** Η μείωση των παραμέτρων λειτουργεί ως μορφή κανονικοποίησης, βοηθώντας το μοντέλο να γενικεύει καλύτερα[79].

3. **Τοπική Αναλλοίωτη Μετατόπιση (Local Translation Invariance):** Το pooling καθιστά την αναπαράσταση των χαρακτηριστικών ελαφρώς αναλλοίωτη σε μικρές μετατοπίσεις. Για παράδειγμα, αν μια ακμή μετακινηθεί κατά ένα pixel, η έξοδος ενός Max Pooling πιθανότατα θα παραμείνει η ίδια. Αυτό είναι επιθυμητό, καθώς η ακριβής θέση ενός χαρακτηριστικού είναι συχνά λιγότερο σημαντική από τη σχετική του θέση ως προς άλλα[79][81].

Τα επίπεδα pooling λειτουργούν "γλιστρώντας" ένα παράθυρο (π.χ., 2×2) πάνω στον χάρτη χαρακτηριστικών, συνήθως με βήμα $S=2$. Σε αντίθεση με τη συνέλιξη, το pooling **δεν έχει μαθησίσιμες παραμέτρους (βάρη)**. Εκτελεί απλώς μια σταθερή αθροιστική συνάρτηση. Οι δύο πιο συνηθισμένοι τύποι είναι:

- **Max Pooling (Συγκέντρωση Μεγίστου):** Είναι ο πιο δημοφιλής τύπος. Επιστρέφει τη **μέγιστη τιμή** από την περιοχή που καλύπτει το παράθυρο. Είναι πολύ αποτελεσματικό στο να "συνοψίζει" την παρουσία ενός χαρακτηριστικού μέσα σε μια περιοχή[80][79].
- **Average Pooling (Συγκέντρωση Μέσου Όρου):** Επιστρέφει τη **μέση τιμή** όλων των εικονοστοιχείων μέσα στο παράθυρο[81].

4.5.6. Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer - Dense)

Έπειτα από αρκετούς γύρους συνέλιξης, ενεργοποίησης και pooling, τα χαρακτηριστικά υψηλού επιπέδου που έχουν εξαχθεί είναι έτοιμα για ταξινόμηση. Αυτό συμβαίνει στα πλήρως συνδεδεμένα επίπεδα, τα οποία βρίσκονται στο τέλος της αρχιτεκτονικής[78][79].

Αυτά τα επίπεδα λειτουργούν ακριβώς όπως τα επίπεδα σε ένα τυπικό Multi-Layer Perceptron (MLP):

1. **Επιπέδωση (Flattening):** Ο τρισδιάστατος όγκος εξόδου (π.χ., $7 \times 7 \times 512$) από το τελευταίο συνελκτικό/pooling επίπεδο "ισιώνεται" (flattened) σε ένα μακρύ, μονοδιάστατο διάνυσμα (π.χ., $7 \times 7 \times 512 = 25088$ στοιχεία)[78].
2. **Πλήρης Σύνδεση:** Αυτό το διάνυσμα τροφοδοτείται σε ένα ή περισσότερα "πυκνά" (dense) επίπεδα, όπου **κάθε νευρώνας** σε ένα επίπεδο συνδέεται με **κάθε νευρώνα** του προηγούμενου επιπέδου[79].

3. **Ταξινόμηση:** Το τελευταίο πλήρως συνδεδεμένο επίπεδο έχει συνήθως τόσους νευρώνες όσες είναι και οι κλάσεις-στόχοι (π.χ., 10 νευρώνες για τα ψηφία 0-9, 1000 νευρώνες για τις κλάσεις του ImageNet, ή 2 νευρώνες για το "καλοήθες"/"κακοήθες")[88][81].

Ο ρόλος αυτών των επιπέδων είναι να λάβουν τα ιεραρχικά χωρικά χαρακτηριστικά που έμαθε το συνελικτικό τμήμα και να τα συνδυάσουν για να καθορίσουν τη συνολική πιθανότητα για κάθε κλάση[89][81].

4.5.7. Επίπεδο Εξόδου (Output Layer)

Το επίπεδο εξόδου είναι το τελευταίο πλήρως συνδεδεμένο επίπεδο, το οποίο, σε συνδυασμό με μια εξειδικευμένη συνάρτηση ενεργοποίησης, παράγει την τελική πρόβλεψη του δικτύου[78][79]. Η επιλογή της συνάρτησης εξαρτάται από τον τύπο του προβλήματος:

- **Softmax:** Αυτή είναι η συνάρτηση εκλογής για προβλήματα **πολυ-κλασσοῦ ταξινόμησης (multi-class classification)**, όπου οι κλάσεις είναι αμοιβαία αποκλειόμενες (π.χ. μια εικόνα δείχνει ή γάτα, ή σκύλο, ή πουλί, αλλά όχι συνδυασμό)[79][89]. Η συνάρτηση Softmax παίρνει τις ακατέργαστες τιμές (logits) από τους K νευρώνες εξόδου και τις μετατρέπει σε μια κατανομή πιθανοτήτων, όπου όλες οι έξοδοι κυμαίνονται από 0 έως 1 και το άθροισμά τους είναι 1. Η κλάση με την υψηλότερη πιθανότητα είναι η τελική πρόβλεψη του μοντέλου.
- **Sigmoid (Λογιστική):** Αυτή η συνάρτηση χρησιμοποιείται τυπικά για προβλήματα **δυσδικής ταξινόμησης (binary classification)** (π.χ. "καρκίνος" ή "όχι καρκίνος")[88][83]. Η συνάρτηση "συνθλίβει" μία μόνο έξοδο (logit) σε μια τιμή μεταξύ 0 και 1, η οποία μπορεί να ερμηνευθεί ως η πιθανότητα η είσοδος να ανήκει στη θετική κλάση.
- **Γραμμική (Linear):** Για προβλήματα **παλινδρόμησης (regression)**, όπου η έξοδος δεν είναι μια κλάση αλλά μια συνεχής τιμή (π.χ. πρόβλεψη της ηλικίας ενός ασθενούς από την εικόνα του), το επίπεδο εξόδου δεν χρησιμοποιεί συνάρτηση ενεργοποίησης (ή ισοδύναμα, χρησιμοποιεί μια γραμμική συνάρτηση $f(x)=x$) [79].

Η έξοδος αυτή συγκρίνεται στη συνέχεια με την πραγματική ετικέτα (ground truth) μέσω μιας **συνάρτησης απώλειας (loss function)**, όπως η **Cross-Entropy** για ταξινόμηση ή

το **Μέσο Τετραγωνικό Σφάλμα (MSE)** για παλινδρόμηση, για τον υπολογισμό του σφάλματος που θα καθοδηγήσει την οπίσθια διάδοση (backpropagation) [79][89][81].

4.6. Υπολογιστική Λογική

Η θεμελιώδης διαφορά των CNNs από τα παραδοσιακά πλήρως συνδεδεμένα δίκτυα (MLPs) έγκειται στον τρόπο με τον οποίο επεξεργάζονται τα δεδομένα εισόδου [89][81]. Αντί να «ισοπεδώνουν» την εικόνα σε ένα μονοδιάστατο διάνυσμα, τα CNNs αξιοποιούν τη χωρική δομή της [78][79]. Αυτό επιτυγχάνεται μέσω τριών κεντρικών ιδεών: της τοπικής συνδεσιμότητας (sparse connectivity), της κοινής χρήσης βαρών (weight sharing) και της ιεραρχικής αναπαράστασης μέσω υποδειγματοληψίας. Η βασική μαθηματική πράξη που υλοποιεί τις δύο πρώτες ιδέες είναι η **συνέλιξη** [79][81].

4.6.1. Εξίσωση Συνελίξεων (Mathematical Definition)

Στο πλαίσιο των CNNs, η συνέλιξη είναι μια γραμμική πράξη μεταξύ ενός πίνακα βαρών, που ονομάζεται **πυρήνας (kernel)** ή **φίλτρο (filter)** (K), και ενός τμήματος της εικόνας εισόδου (I). Η πράξη αυτή εκτελείται επαναληπτικά σε ολόκληρη την εικόνα, παράγοντας έναν **χάρτη χαρακτηριστικών (feature map)** [78][79].

Για μια δισδιάστατη (2D) εικόνα εισόδου I και έναν δισδιάστατο πυρήνα K , η διακριτή συνέλιξη (discrete convolution) στο σημείο (m, n) ορίζεται μαθηματικά ως:

$$(I * K)(m, n) = \sum_i \sum_j I(m - i, n - j) K(i, j)$$

Στην πράξη της βαθιάς μάθησης, η λειτουργία που χρησιμοποιείται είναι τεχνικά η **διασταυρούμενη συσχέτιση (cross-correlation)**, η οποία είναι υπολογιστικά πανομοιότυπη με τη συνέλιξη αλλά χωρίς την αναστροφή του πυρήνα ¹. Η λειτουργία αυτή, που συχνά αναφέρεται απλώς ως "συνέλιξη" στη βιβλιογραφία των CNNs, υπολογίζεται ως το άθροισμα των κατά στοιχείο γινομένων (element-wise product) μεταξύ του πυρήνα και του τμήματος της εικόνας που καλύπτει [78]:

$$(I * K)(m, n) = \sum_i \sum_j I(m + i, n + j) K(i, j)$$

Σε ένα τυπικό συνελικτικό επίπεδο, η είσοδος x^{l-1} (από το προηγούμενο επίπεδο) συνελίσσεται με έναν πυρήνα k^l και προστίθεται μια πόλωση (bias) b^l . Το αποτέλεσμα

περνά μέσα από μια μη-γραμμική συνάρτηση ενεργοποίησης f (όπως η ReLU) για να παραχθεί η έξοδος x^l (ο χάρτης χαρακτηριστικών) [79] [81]:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right)$$

Όπου M_j αναφέρεται στην επιλογή των χαρτών εισόδου. Αυτή η εξίσωση δείχνει πώς τα χαρακτηριστικά από το προηγούμενο επίπεδο μετασχηματίζονται γραμμικά (συνέλιξη και πόλωση) και κατόπιν μη-γραμμικά (συνάρτηση ενεργοποίησης) για να δημιουργήσουν τα χαρακτηριστικά του επόμενου επιπέδου.

4.6.2. Παράδειγμα 2D Συνέλιξης

Για να γίνει κατανοητή η λειτουργία, ας θεωρήσουμε μια απλή ασπρόμαυρη εικόνα εισόδου 5×5 (όπου οι τιμές αντιπροσωπεύουν την ένταση των pixels) και έναν πυρήνα (φίλτρο) 3×3 , ο οποίος είναι σχεδιασμένος να ανιχνεύει κάθετες ακμές.

Είσοδος (Input Image) I (5×5):

$$I = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Πυρήνας (Kernel / Filter) K (3×3) (Ανιχνευτής Κάθετης Ακμής):

$$K = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

Χρησιμοποιούμε βήμα $S = 1$ και μηδενικό γέμισμα $P=0$. Το φίλτρο 3×3 θα "γλιστρήσει" πάνω στην εικόνα 5×5 . Το μέγεθος εξόδου θα είναι $(N - F + 2P)/S + 1 = (5 - 3 + 0)/1 + 1 = 3$. Άρα, ο χάρτης χαρακτηριστικών εξόδου θα είναι 3×3 .

Βήμα 1: Υπολογισμός Εξόδου $[0, 0]$

Το φίλτρο τοποθετείται στην πάνω αριστερή γωνία της εικόνας:

$$\text{Patch} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Υπολογίζεται το άθροισμα των κατά στοιχείο γινομένων:

$$\begin{aligned} & (1 \cdot 1) + (1 \cdot 0) + (1 \cdot -1) + (0 \cdot 1) + (1 \cdot 0) + (1 \cdot -1) + (0 \cdot 1) + (0 \cdot 0) \\ & \quad + (1 \cdot -1) \\ & = 1 + 0 - 1 + 0 + 0 - 1 + 0 + 0 - 1 = -2 \end{aligned}$$

Βήμα 2: Υπολογισμός Εξόδου [0, 1]

Το φίλτρο μετακινείται ένα βήμα δεξιά:

$$\text{Patch} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Υπολογισμός:

$$\begin{aligned} & (1 \cdot 1) + (1 \cdot 0) + (0 \cdot -1) + (1 \cdot 1) + (1 \cdot 0) + (1 \cdot -1) + (0 \cdot 1) + (1 \cdot 0) \\ & \quad + (1 \cdot -1) \\ & = 1 + 0 + 0 + 1 + 0 - 1 + 0 + 0 - 1 = 0 \end{aligned}$$

Βήμα 3: Υπολογισμός Εξόδου [0, 2]

Το φίλτρο μετακινείται ξανά δεξιά:

$$\text{Patch} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Υπολογισμός:

$$\begin{aligned} & (1 \cdot 1) + (0 \cdot 0) + (0 \cdot -1) + (1 \cdot 1) + (1 \cdot 0) + (0 \cdot -1) + (1 \cdot 1) + (1 \cdot 0) \\ & \quad + (1 \cdot -1) \\ & = 1 + 0 + 0 + 1 + 0 + 0 + 1 + 0 - 1 = 2 \end{aligned}$$

Συνεχίζοντας αυτή τη διαδικασία για όλες τις θέσεις, προκύπτει ο τελικός **Χάρτης Χαρακτηριστικών (Feature Map)**:

$$\text{Output} = \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix}$$

Αν εφαρμόσουμε μια συνάρτηση ενεργοποίησης **ReLU** σε αυτόν τον χάρτη (μηδενίζοντας τις αρνητικές τιμές), η τελική έξοδος του επιπέδου θα είναι:

$$\text{Output}_{\text{ReLU}} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Παρατηρούμε ότι το φίλτρο "ενεργοποιήθηκε" (παρήγαγε υψηλές τιμές) ακριβώς στη στήλη όπου η εικόνα άλλαξε από '1' σε '0', ανιχνεύοντας επιτυχώς την κάθετη ακμή[79][81].

4.6.3. Υπολογιστική Πολυπλοκότητα και Μείωση Παραμέτρων

Για να κατανοήσουμε την υπολογιστική αποδοτικότητα των CNNs, είναι απαραίτητο να τα συγκρίνουμε με τα παραδοσιακά Πλήρως Συνδεδεμένα Δίκτυα (MLPs).

Το Πρόβλημα του MLP (Curse of Dimensionality):

Σε ένα MLP, κάθε νευρώνας ενός επιπέδου συνδέεται με όλους τους νευρώνες (ή pixels) του προηγούμενου επιπέδου[77][78]. Αν εφαρμόζαμε αυτή τη λογική σε μια τυπική εικόνα, ο αριθμός των παραμέτρων (βαρών) θα ήταν αστρονομικός.

- Ας υποθέσουμε μια μικρή έγχρωμη εικόνα 32 X 32 X 3 (3 κανάλια RGB). Σύνολο εικονοστοιχείων = 3,072.
- Αν το πρώτο κρυφό επίπεδο έχει 1,000 νευρώνες.
- Ο αριθμός των βαρών μόνο για αυτό το επίπεδο θα ήταν: $3,072 \times 1,000 + 1,000$ (biases) ≈ 3.1 εκατομμύρια παράμετροι.
- Για μια μεγαλύτερη εικόνα, π.χ. 224 X 224 X 3 ($\approx 150,000$ pixels) και 4,096 νευρώνες (όπως στο AlexNet), οι παράμετροι θα ήταν $\$150,000 \times 4,096 \approx 600 \text{ \textit{εκατομμύρια}} \text{\$}$.

Αυτός ο τεράστιος αριθμός παραμέτρων καθιστά τα MLPs υπολογιστικά πανάκριβα και εξαιρετικά επιρρεπή στην **υπερπροσαρμογή (overfitting)**. Επιπλέον, η "ισοπέδωση" της εικόνας σε διάνυσμα καταστρέφει την πολύτιμη χωρική πληροφορία[79][89].

Η Λύση του CNN:

Τα CNNs επιλύουν αυτό το πρόβλημα δραστικά μέσω δύο βασικών αρχών:[78][80][89]

1. Αραιή Συνδεσιμότητα (Sparse Connectivity / Local Receptive Fields):

Ένας νευρώνας στον χάρτη χαρακτηριστικών δεν συνδέεται με όλη την προηγούμενη εικόνα. Συνδέεται μόνο με μια μικρή, τοπική περιοχή της εισόδου, αυτήν που καλύπτει ο πυρήνας (π.χ., 3×3). Αυτό από μόνο του μειώνει δραστικά τον αριθμό των συνδέσεων.

2. Κοινή Χρήση Βαρών (Weight Sharing):

Αυτή είναι η πιο σημαντική καινοτομία. Σε ένα MLP, κάθε σύνδεση (π.χ., από το pixel $[0,0]$ στον νευρώνα 1 και από το pixel $[0,1]$ στον νευρώνα 1) έχει το δικό της, μοναδικό βάρος [79].

Σε ένα CNN, τα βάρη που ορίζουν έναν πυρήνα (π.χ., τα 9 βάρη του 3×3 φίλτρου μας) είναι τα ίδια βάρη που χρησιμοποιούνται καθώς ο πυρήνας αυτός "γλιστράει" σε ολόκληρη την εικόνα 888. Το δίκτυο δεν μαθαίνει έναν ανιχνευτή ακμής για την πάνω αριστερή γωνία και έναν άλλο ανιχνευτή ακμής για το κέντρο. Μαθαίνει έναν ανιχνευτή ακμής και τον εφαρμόζει παντού [88][85].

Υπολογισμός Παραμέτρων σε CNN:

Παράμετροι για ένα τυπικό συνελκτικό επίπεδο, χρησιμοποιώντας τον τύπο που παρέχεται από τον Alom et al.

- Έστω είσοδος $224 \times 224 \times 3$.
- Έστω ένα συνελκτικό επίπεδο με $FM_l = 64$ φίλτρα (πυρήνες).
- Κάθε φίλτρο έχει μέγεθος $F=3$ (δηλαδή 3×3).
- Το βάθος κάθε φίλτρου πρέπει να ταιριάζει με το βάθος της εισόδου, άρα $FM_{l-1} = 3$.
- Μέγεθος κάθε πυρήνα: $3 \times 3 \times 3 = 27$ βάρη.
- Κάθε πυρήνας έχει επίσης 1 πόλωση (bias).
- Σύνολο παραμέτρων ανά φίλτρο = $27 + 1 = 28$.
- Συνολικές Παράμετροι Επιπέδου = (Παράμετροι ανά φίλτρο) \times (Αριθμός φίλτρων)
- Total = $((F \times F \times FM_{l-1}) + 1) \times FM_l = ((3 \times 3 \times 3) + 1) \times 64 = (27 + 1) \times 64 = 28 \times 64 = 1,792$.

Σύγκριση:

- **MLP:** ~600,000,000 παράμετροι.
- **CNN (1ο επίπεδο):** 1,792 παράμετροι.

Αυτή η δραματική μείωση του αριθμού των παραμέτρων (parameter reduction) είναι αυτό που καθιστά τα CNNs υπολογιστικά εφικτά, μειώνει δραστικά τις απαιτήσεις μνήμης και τα καθιστά πολύ λιγότερο επιρρεπή στην υπερ-προσαρμογή σε σύγκριση με τα MLPs.

4.7. Τεχνικές Βελτίωσης

Η εκπαίδευση των βαθιών νευρωνικών δικτύων είναι μια υπολογιστικά απαιτητική διαδικασία που αντιμετωπίζει δύο θεμελιώδεις προκλήσεις: [78][80][89]

1. **Σταθερότητα και Ταχύτητα Εκπαίδευσης:** Καθώς το δίκτυο γίνεται βαθύτερο, η διάδοση της κλίσης (gradient) προς τα πίσω (backpropagation) γίνεται ασταθής. Αυτό μπορεί να οδηγήσει στο φαινόμενο της «εξαφανιζόμενης κλίσης» (vanishing gradient) ή της «εκρηγνυόμενης κλίσης» (exploding gradient), καθιστώντας τη σύγκλιση του μοντέλου εξαιρετικά αργή ή αδύνατη [79][89].
2. **Υπερ-προσαρμογή (Overfitting):** Τα βαθιά δίκτυα διαθέτουν εκατομμύρια παραμέτρους. Με τόσο υψηλή χωρητικότητα, τείνουν να «απομνημονεύουν» τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένου του θορύβου, αντί να μαθαίνουν τα υποκείμενα μοτίβα. Αυτό έχει ως αποτέλεσμα το μοντέλο να αποδίδει άριστα στα δεδομένα που έχει δει, αλλά να αποτυγχάνει να γενικεύσει σε νέα, αθέατα δεδομένα [78][79].

Για την αντιμετώπιση αυτών των προκλήσεων, έχει αναπτυχθεί μια σειρά από ισχυρές τεχνικές βελτίωσης και κανονικοποίησης, οι οποίες αποτελούν πλέον αναπόσπαστο μέρος σχεδόν κάθε σύγχρονης αρχιτεκτονικής CNN.

4.7.1. Batch Normalization (Ομαδική Κανονικοποίηση)

Η Ομαδική Κανονικοποίηση (Batch Normalization - BN) είναι μια τεχνική που προτάθηκε από τους Ioffe και Szegedy (2015) και άλλαξε δραστικά τον τρόπο εκπαίδευσης των βαθιών δικτύων [92].

Το Πρόβλημα: "Internal Covariate Shift"

Κατά τη διάρκεια της εκπαίδευσης, τα βάρη σε κάθε επίπεδο ενημερώνονται συνεχώς. Αυτό έχει ως συνέπεια, η κατανομή των εισόδων (δηλαδή των ενεργοποιήσεων από το

προηγούμενο επίπεδο) σε κάθε κρυφό επίπεδο να αλλάζει συνεχώς. Αυτό το φαινόμενο ονομάστηκε "Internal Covariate Shift". Το δίκτυο αναγκάζεται να προσαρμόζεται συνεχώς σε έναν «κινούμενο στόχο», γεγονός που επιβραδύνει δραματικά τη σύγκλιση και καθιστά την εκπαίδευση ευαίσθητη στην αρχικοποίηση των βαρών [79].

Η Λύση του BN

Το Batch Normalization επιλύει αυτό το πρόβλημα κανονικοποιώντας τις εισόδους κάθε επιπέδου ανά μικρο-δεσμίδα (mini-batch). Η διαδικασία τυπικά λαμβάνει χώρα αμέσως μετά το συνελκτικό ή το πλήρως συνδεδεμένο επίπεδο, αλλά πριν από τη συνάρτηση ενεργοποίησης (π.χ. ReLU) [79][92].

Για κάθε mini-batch, το BN εκτελεί τα εξής βήματα για κάθε χαρακτηριστικό (κανάλι) ξεχωριστά [92]:

1. **Υπολογισμός Μέσου Όρου & Διακύμανσης:** Υπολογίζει τη μέση τιμή (μ) και τη διακύμανση (σ^2) των ενεργοποιήσεων του τρέχοντος mini-batch .

2. **Κανονικοποίηση:** Κανονικοποιεί κάθε ενεργοποίηση x_i ώστε να έχει μηδενική μέση

τιμή και μοναδιαία διακύμανση
$$\left(\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \right).$$

3. **Κλιμάκωση και Μετατόπιση:** Εισάγει δύο νέες μαθησίσιμες παραμέτρους, τη γάμμα (γ) και τη βήτα (β), οι οποίες κλιμακώνουν και μετατοπίζουν την κανονικοποιημένη τιμή: $y_i = \gamma \hat{x}_i + \beta$. Αυτό το βήμα είναι κρίσιμο, καθώς επιτρέπει στο δίκτυο να μάθει τη βέλτιστη κατανομή για τις εισόδους του επόμενου επιπέδου, αντί να του επιβάλλεται αυστηρά η κατανομή $\mathcal{N}(0,1)$.

Οφέλη του Batch Normalization:

- **Ταχύτερη Εκπαίδευση:** Σταθεροποιώντας την κατανομή των εισόδων, το BN επιτρέπει τη χρήση πολύ υψηλότερων ρυθμών μάθησης (learning rates), μειώνοντας δραματικά τον συνολικό χρόνο εκπαίδευσης [92].
- **Μείωση Ευαισθησίας στην Αρχικοποίηση:** Το δίκτυο γίνεται πολύ λιγότερο ευαίσθητο στην αρχική επιλογή των βαρών [79][92].

- **Επίδραση Κανονικοποίησης:** Το BN εισάγει έναν ελαφρύ θόρυβο (λόγω της εξάρτησης από το mini-batch), ο οποίος λειτουργεί ως τεχνική κανονικοποίησης και συχνά μειώνει ή εξαλείφει την ανάγκη για Dropout [78][92].

4.7.2. Dropout (Απόσυρση)

Το Dropout είναι ίσως η πιο διαδεδομένη και αποτελεσματική τεχνική κανονικοποίησης για την καταπολέμηση της υπερ-προσαρμογής σε νευρωνικά δίκτυα [78][79].

Το Πρόβλημα: "Co-adaptation"

Όταν ένα βαθύ δίκτυο εκπαιδεύεται, οι νευρώνες του τείνουν να αναπτύσσουν πολύπλοκες «συν-προσαρμογές» (co-adaptations). Αυτό σημαίνει ότι ένας νευρώνας μπορεί να βασίζεται υπερβολικά στην παρουσία και την ενεργοποίηση συγκεκριμένων άλλων νευρώνων στο προηγούμενο επίπεδο [78]. Αυτή η αλληλεξάρτηση αποτελεί μια μορφή απομνημόνευσης του θορύβου των δεδομένων εκπαίδευσης και βλάπτει τη γενίκευση [89].

Η Λύση του Dropout

Το Dropout αντιμετωπίζει αυτό το πρόβλημα με έναν απλό τρόπο: κατά τη διάρκεια της εκπαίδευσης, σε κάθε βήμα (iteration), "αποσύρει" (drops) τυχαία ένα ποσοστό p (π.χ., $p=0.5$) των νευρώνων ενός επιπέδου [79]. Αυτό πρακτικά σημαίνει ότι οι έξοδοι αυτών των νευρώνων μηδενίζονται προσωρινά για αυτό το βήμα.

Αυτή η διαδικασία αναγκάζει το δίκτυο να μάθει **πλεονάζουσες αναπαραστάσεις (redundant representations)**. Κανένας νευρώνας δεν μπορεί να βασιστεί στην παρουσία κάποιου άλλου, καθώς οποιοσδήποτε νευρώνας μπορεί να "αποσυρθεί" τυχαία. Το δίκτυο αναγκάζεται να μάθει πιο εύρωστα χαρακτηριστικά που είναι χρήσιμα από μόνα τους [78].

Εννοιολογικά, το Dropout προσεγγίζει την ιδέα της εκπαίδευσης ενός τεράστιου **συνόλου (ensemble)** από διαφορετικά, "λεπτότερα" νευρωνικά δίκτυα, τα οποία μοιράζονται βάρη, και στη συνέχεια υπολογίζει τον μέσο όρο των προβλέψεών τους [79][89].

- **Στην Εκπαίδευση:** Ενεργό. Οι νευρώνες μηδενίζονται με πιθανότητα p .
- **Στη Δοκιμή (Inference):** Ανενεργό. Όλοι οι νευρώνες χρησιμοποιούνται, αλλά οι ενεργοποιήσεις τους κλιμακώνονται (πολλαπλασιάζονται) με τον παράγοντα $1-p$ (ή

ισοδύναμα, γίνεται "inverted dropout" κατά την εκπαίδευση), ώστε η συνολική αναμενόμενη έξοδος να παραμένει ίδια με αυτή της εκπαίδευσης.

Το Dropout εφαρμόζεται συνήθως στα πλήρως συνδεδεμένα επίπεδα, αν και υπάρχουν παραλλαγές του (όπως το 2D Dropout) και για συνελκτικά επίπεδα [78][79].

4.7.3. Data Augmentation (Επαύξηση Δεδομένων)

Μία από τις πιο αποτελεσματικές στρατηγικές για την καταπολέμηση της υπερπροσαρμογής είναι απλώς η χρήση περισσότερων δεδομένων. Ωστόσο, η συλλογή και η επισημείωση ιατρικών εικόνων είναι μια διαδικασία δαπανηρή, χρονοβόρα και απαιτεί εξειδικευμένους παθολογοανατόμους [88][86][87].

Η **Επαύξηση Δεδομένων (Data Augmentation)** είναι μια τεχνική που αντιμετωπίζει αυτό το πρόβλημα, δημιουργώντας τεχνητά νέα δείγματα εκπαίδευσης από τα ήδη υπάρχοντα, εφαρμόζοντας σε αυτά μια σειρά από ρεαλιστικούς μετασχηματισμούς [79][81].

Αντί να τροφοδοτούμε το δίκτυο με την ίδια ακριβώς εικόνα σε κάθε εποχή, του δίνουμε μια ελαφρώς τροποποιημένη εκδοχή της. Αυτό διδάσκει στο μοντέλο να είναι **αναλλοίωτο (invariant)** σε αυτές τις αλλαγές, αναγκάζοντάς το να μάθει τα βαθύτερα χαρακτηριστικά της κλάσης [84][85].

Στην ιστοπαθολογία, όπου η ανάλυση βασίζεται σε κυτταρικά και αρχιτεκτονικά μοτίβα, οι μετασχηματισμοί αυτοί είναι ιδιαίτερα κρίσιμοι, καθώς το μοντέλο πρέπει να αναγνωρίζει έναν καρκινικό αδένα ανεξάρτητα από [88][93]:

- Τον προσανατολισμό του (orientation)
- Τη θέση του στην εικόνα (location)
- Τις μικρές διαφοροποιήσεις στη χρώση H&E (staining variations)

Οι πιο συνηθισμένοι μετασχηματισμοί επαύξησης περιλαμβάνουν [79][81][85]:

- **Rotation (Περιστροφή):** Τυχαία περιστροφή της εικόνας κατά μερικές μοίρες (π.χ., $\pm 10^\circ$). Στην ιστοπαθολογία, συχνά χρησιμοποιούνται και περιστροφές 90° , 180° , 270° , καθώς η απόλυτη κατεύθυνση του ιστού σπάνια έχει σημασία.
- **Flipping (Αναστροφή):** Οριζόντια και κάθετη αναστροφή (horizontal/vertical flip).

- **Scaling (Κλιμάκωση):** Τυχαίο "ζουμάρισμα" (zoom in/out) στην εικόνα κατά ένα μικρό ποσοστό.
- **Translation (Μετατόπιση):** Μικρή μετακίνηση της εικόνας οριζόντια ή κάθετα.
- **Contrast/Color Jittering (Αλλαγή Αντίθεσης/Χρώματος):** Τυχαίες μικροαλλαγές στη φωτεινότητα, την αντίθεση (contrast) και τον κορεσμό (saturation) της εικόνας. Αυτό είναι εξαιρετικά σημαντικό για την ιστοπαθολογία, καθώς βοηθά το μοντέλο να γενικεύει σε δείγματα που προέρχονται από διαφορετικά εργαστήρια, με διαφορετικά πρωτόκολλα χρώσης [88].

4.7.4. Early Stopping (Πρώιμη Διακοπή)

Η Πρώιμη Διακοπή είναι μια απλή αλλά αποτελεσματική τεχνική κανονικοποίησης που βασίζεται στην παρακολούθηση της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης [78].

Η Διαδικασία Κατά την προετοιμασία των δεδομένων, το σύνολο εκπαίδευσης χωρίζεται περαιτέρω σε ένα **σύνολο εκπαίδευσης (training set)** και ένα **σύνολο επικύρωσης (validation set)**, π.χ. σε αναλογία 80%-20% [79].

Η εκπαίδευση (υπολογισμός κλίσης και ενημέρωση βαρών) γίνεται *αποκλειστικά* στο σύνολο εκπαίδευσης. Ωστόσο, στο τέλος κάθε εποχής (epoch), η απόδοση του μοντέλου (συνήθως η συνάρτηση απώλειας - loss) υπολογίζεται ξεχωριστά και για τα δύο σύνολα.

Η Παρατήρηση

Τυπικά, παρατηρείται η ακόλουθη συμπεριφορά :

1. **Αρχικά:** Τόσο το σφάλμα εκπαίδευσης (training loss) όσο και το σφάλμα επικύρωσης (validation loss) μειώνονται. Το μοντέλο μαθαίνει τα θεμελιώδη μοτίβα των δεδομένων.
2. **Σημείο Κορεσμού:** Μετά από κάποιο σημείο, το training loss συνεχίζει να μειώνεται (καθώς το μοντέλο αρχίζει να απομνημονεύει τις ιδιαιτερότητες του training set), αλλά το **validation loss σταματά να μειώνεται και αρχίζει να αυξάνεται** . Αυτό είναι το ακριβές σημείο όπου ξεκινά η υπερ-προσαρμογή [77][79].

Η Τεχνική

Το **Early Stopping** απλώς παρακολουθεί το validation loss. Η εκπαίδευση τερματίζεται αυτόματα όταν το validation loss δεν παρουσιάζει βελτίωση για έναν προκαθορισμένο αριθμό εποχών (που ονομάζεται "**patience**" - "υπομονή") . Το μοντέλο που αποθηκεύεται ως τελικό δεν είναι αυτό της τελευταίας εποχής, αλλά το μοντέλο από την εποχή που είχε το **χαμηλότερο validation loss** . Αυτή η τεχνική διασφαλίζει ότι επιλέγεται το μοντέλο με την καλύτερη ικανότητα γενίκευσης [79].

4.8. Εξειδικευμένες Αρχιτεκτονικές

Μετά την πρωτοποριακή επιτυχία του AlexNet το 2012, η ερευνητική κοινότητα εστίασε στη διερεύνηση των παραγόντων που οδηγούν σε καλύτερη απόδοση των CNNs. Η βασική υπόθεση ήταν ότι τα “βαθύτερα” δίκτυα (με περισσότερα επίπεδα) μπορούσαν να μάθουν πιο σύνθετες και ιεραρχικές αναπαραστάσεις, οδηγώντας σε υψηλότερη ακρίβεια. Αυτή η “πορεία προς το βάθος” οδήγησε σε εμβληματικές αρχιτεκτονικές όπως το VGGNet, το GoogLeNet και το ResNet, οι οποίες έθεσαν νέα πρότυπα απόδοσης [78][79].

4.8.1. VGGNet

Το VGGNet, που αναπτύχθηκε από το **Visual Geometry Group** του Πανεπιστημίου της Οξφόρδης (Simonyan & Zisserman, 2014), ήταν ο επιλαχών στον διαγωνισμό ILSVRC 2014. Η βασική φιλοσοφία πίσω από το VGGNet ήταν η διερεύνηση του πώς το **βάθος** του δικτύου επηρεάζει την απόδοση, διατηρώντας παράλληλα μια εξαιρετικά **απλή και ομοιόμορφη δομή**[78][79].

4.8.1.1. Αρχιτεκτονική Δομή: Η Δύναμη της Απλότητας

Σε αντίθεση με το AlexNet που χρησιμοποιούσε ποικιλία μεγεθών φίλτρων (π.χ., 11 X 11, 5 X 5) ή το GoogLeNet (νικητής του 2014) που εισήγαγε πολύπλοκες "Inception modules" , η αρχιτεκτονική VGG χαρακτηρίζεται από μια κομψή ομοιομορφία [79][81].

Τα βασικά δομικά στοιχεία του είναι:

1. **Αποκλειστική Χρήση 3 X 3 Συνελίξεων:** Η πιο σημαντική σχεδιαστική απόφαση του VGG ήταν η αποκλειστική χρήση των μικρότερων δυνατών συνελκτικών πυρήνων (3

X 3) σε όλα τα συνελκτικά επίπεδα. Οι συγγραφείς απέδειξαν ότι μια στοίβα από δύο διαδοχικά φίλτρα 3 X 3 έχει το ίδιο αποτελεσματικό «πεδίο αντίληψης» (receptive field) με ένα φίλτρο 5 X 5. Ομοίως, τρία φίλτρα 3 X 3 μμμούνται ένα φίλτρο 7 X 7 [78][79].

Πλεονεκτήματα: Αυτή η προσέγγιση είναι ανώτερη γιατί:

- **Εισάγει περισσότερη μη-γραμμικότητα:** Μια στοίβα από τρία 3 X 3 επίπεδα περιλαμβάνει τρεις συναρτήσεις ενεργοποίησης (ReLU), ενώ ένα επίπεδο 7 X 7 περιλαμβάνει μόνο μία. Αυτό καθιστά τη συνάρτηση απόφασης του δικτύου πιο διακριτική (discriminative).
 - **Μειώνει τις Παραμέτρους:** Μια στοίβα από τρία 3 X 3 συνελκτικά επίπεδα με C κανάλια έχει $3 \times (3^2 \times C^2) = 27C^2$ βάρη. Ένα μεμονωμένο 7 X 7 επίπεδο θα είχε $7^2 \times C^2 = 49C^2$ βάρη, δηλαδή σημαντικά περισσότερα.
2. **2 X 2 Max Pooling:** Για τη μείωση των χωρικών διαστάσεων (downsampling), το VGG χρησιμοποιεί αποκλειστικά επίπεδα Max Pooling με παράθυρο 2 X 2 και βήμα (stride) 2. Αυτά τοποθετούνται μετά από κάθε "μπλοκ" συνελκτικών επιπέδων [79] [80].
 3. **Ομοιόμορφα Μπλοκ:** Η τυπική αρχιτεκτονική VGG (όπως το VGG-16 ή το VGG-19, που αναφέρονται στο συνολικό αριθμό των επιπέδων με βάρη) αποτελείται από διαδοχικά μπλοκ 2 ή 3 συνελκτικών επιπέδων 3 X 3, ακολουθούμενα από ένα επίπεδο Max Pooling 2 X 2.
 4. **Βαριά Πλήρως Συνδεδεμένα Επίπεδα:** Στο τέλος του δικτύου, μετά την επιπέδωση (flattening), το VGG χρησιμοποιεί τρία πλήρως συνδεδεμένα (FC) επίπεδα. Αυτά τα επίπεδα, ειδικά τα δύο πρώτα με 4096 νευρώνες το καθένα, είναι υπολογιστικά "βαριά" και περιέχουν τη συντριπτική πλειοψηφία των παραμέτρων του δικτύου (π.χ., το VGG-16 έχει συνολικά ~138 εκατομμύρια παραμέτρους, με τις περισσότερες να βρίσκονται στα FC επίπεδα) [78][79][81].

4.8.1.2. Χρήση ως Βάση για Transfer Learning (Μάθηση Μεταφοράς)

Παρά το μεγάλο του μέγεθος και το υπολογιστικό του κόστος, το VGGNet παραμένει μία από τις πιο δημοφιλείς αρχιτεκτονικές-βάσεις για εφαρμογές **Μάθησης Μεταφοράς (Transfer Learning)**, ειδικά στην ιατρική απεικόνιση [82][83].

Η Μάθηση Μεταφοράς είναι η διαδικασία κατά την οποία ένα μοντέλο που έχει **προ-εκπαιδευτεί (pre-trained)** σε ένα πολύ μεγάλο, γενικό σύνολο δεδομένων (όπως το

ImageNet), χρησιμοποιείται ως σημείο εκκίνησης για την εκπαίδευση σε ένα νέο, συνήθως μικρότερο και πιο εξειδικευμένο σύνολο δεδομένων [94].

Το VGG είναι ιδανικό για αυτό το σκοπό για τους εξής λόγους:

1. **Ισχυρά Γενικευμένα Χαρακτηριστικά:** Λόγω της εκπαίδευσής του σε εκατομμύρια εικόνες, τα συνελκτικά επίπεδα του VGG έχουν μάθει να εξάγουν ένα πλούσιο και ιεραρχικό σύνολο χαρακτηριστικών[78][79][94]. Τα πρώτα επίπεδα μαθαίνουν γενικές έννοιες (ακμές, γωνίες, χρώματα), ενώ τα βαθύτερα επίπεδα μαθαίνουν πιο σύνθετες δομές (υφές, μοτίβα, τμήματα αντικειμένων)[81]. Αυτά τα χαρακτηριστικά είναι χρήσιμα για σχεδόν οποιαδήποτε εργασία υπολογιστικής όρασης.
2. **Απλότητα Αρχιτεκτονικής:** Η ομοιόμορφη δομή του καθιστά εύκολη την τροποποίησή του.

Η τυπική διαδικασία Μάθησης Μεταφοράς με VGG (π.χ. για ταξινόμηση βιοψιών παχέος εντέρου) περιλαμβάνει: [88][82]

- **Φόρτωση Προ-εκπαιδευμένου VGG:** Το μοντέλο φορτώνεται με τα βάρη που έμαθε από το ImageNet.
- **Αντικατάσταση της "Κεφαλής":** Τα τρία τελευταία πλήρως συνδεδεμένα επίπεδα, που ήταν σχεδιασμένα για τις 1000 κλάσεις του ImageNet, αφαιρούνται.
- **Προσθήκη Νέας "Κεφαλής":** Στη θέση τους, προστίθενται νέα, συνήθως ελαφρύτερα, πλήρως συνδεδεμένα επίπεδα, με το τελικό επίπεδο εξόδου να ταιριάζει με τον αριθμό των κλάσεων-στόχων (π.χ., 2 κλάσεις: "καλοήθες" και "κακοήθες")[94].
- **Επανεκπαίδευση (Fine-tuning):** Το νέο μοντέλο εκπαιδεύεται στο εξειδικευμένο ιατρικό σύνολο δεδομένων. Συχνά, τα βάρη των αρχικών συνελκτικών επιπέδων "παγώνουν" (frozen) τουλάχιστον για τις πρώτες εποχές, και εκπαιδεύονται μόνο τα νέα FC επίπεδα, καθιστώντας την εκπαίδευση εξαιρετικά γρήγορη [82].

Αυτή η προσέγγιση επιτρέπει την επίτευξη υψηλής ακρίβειας σε εξειδικευμένα πεδία όπως η ιστοπαθολογία, ακόμη και με περιορισμένο αριθμό επισημειωμένων ιατρικών εικόνων, αξιοποιώντας τη "γνώση" που έχει ήδη αποκτηθεί από εκατομμύρια γενικές εικόνες.

4.9. ResNet (Residual Networks)

Η αρχιτεκτονική VGG απέδειξε ότι το βάθος του δικτύου αποτελεί κρίσιμο παράγοντα για την επίτευξη υψηλής απόδοσης. Ωστόσο, η απλή “στοίβαξη” επιπέδων (stacking layers) δημιουργεί ένα θεμελιώδες εμπόδιο: καθώς τα δίκτυα γίνονται εξαιρετικά βαθιά (π.χ. 30, 40, 50 επίπεδα), η απόδοσή τους κορεσμεύεται και στη συνέχεια αρχίζει να **υποβαθμίζεται (degradation)**[78][79]. Παραδόξως, ένα βαθύτερο μοντέλο κατέληγε να έχει υψηλότερο σφάλμα εκπαίδευσης (training error) από το ρηχότερο αντίστοιχό του. Αυτό το φαινόμενο δεν οφειλόταν στην υπερ-προσαρμογή (overfitting), αλλά σε ένα υπολογιστικό εμπόδιο [95].

Το **ResNet (Residual Network)**, που παρουσιάστηκε από τους Kaiming He κ.ά. (2015) και κέρδισε τον διαγωνισμό ILSVRC 2015, σχεδιάστηκε για να επιλύσει ακριβώς αυτό το πρόβλημα [85][95].

4.9.1. Θεωρητική Βάση: Το Πρόβλημα της Εξαφανιζόμενης Κλίσης (Vanishing Gradient Problem)

Η δυσκολία στην εκπαίδευση εξαιρετικά βαθιών δικτύων συνδέεται άμεσα με το πρόβλημα της **εξαφανιζόμενης κλίσης (vanishing gradient problem)**[78]. Κατά τον αλγόριθμο της οπίσθιας διάδοσης (Backpropagation), η κλίση του σφάλματος πολλαπλασιάζεται διαδοχικά καθώς διαδίδεται προς τα πίσω, από τα τελευταία επίπεδα προς τα πρώτα [89].

Σε ένα πολύ βαθύ δίκτυο, το γινόμενο αυτών των διαδοχικών παραγώγων (ιδίως όταν χρησιμοποιούνται συναρτήσεις ενεργοποίησης όπως η sigmoid ή η tanh που έχουν μικρές παραγώγους) μπορεί να τείνει εκθετικά προς το μηδέν. Ως αποτέλεσμα, οι κλίσεις που φτάνουν στα πρώτα επίπεδα του δικτύου γίνονται σχεδόν μηδενικές («εξαφανίζονται»). Αυτό σημαίνει ότι τα βάρη των πρώτων επιπέδων, που είναι υπεύθυνα για την εκμάθηση των θεμελιωδών χαρακτηριστικών (π.χ. ακμές, υφές), σταματούν να ενημερώνονται και η εκπαίδευση λιμνάζει [78][79][81].

4.9.2. Η Καινοτομία: Παρακαμπτήριες Συνδέσεις (Skip Connections / Identity Shortcuts)

Η λύση που πρότεινε το ResNet ήταν ταυτόχρονα απλή και επαναστατική: η εισαγωγή των **παρακαμπτήριων συνδέσεων (skip connections)**, επίσης γνωστών ως **συνδέσεων ταυτότητας (identity shortcuts)** [79][95].

Αντί να περιμένουμε ότι μια στοίβα επιπέδων θα μάθει απευθείας μια επιθυμητή υποκείμενη αντιστοίχιση $H(x)$, το ResNet επαναδιατυπώνει το πρόβλημα. Αναγκάζει τα επίπεδα να μάθουν μια **υπολειμματική συνάρτηση (residual function)** $F(x)$, η οποία ορίζεται ως η διαφορά μεταξύ της επιθυμητής εξόδου και της εισόδου: $F(x) = H(x) - x$.

Στη συνέχεια, η αρχική είσοδος x προστίθεται στην έξοδο $F(x)$ μέσω μιας παρακαμπτήριας σύνδεσης. Επομένως, η τελική έξοδος y (ή $H(x)$) του μπλοκ είναι:

$$y = F(x, \{W_i\}) + x$$

όπου $F(x, \{W_i\})$ είναι ο μη γραμμικός μετασχηματισμός που εκτελείται από τα επίπεδα (π.χ. δύο ή τρία συνελκτικά επίπεδα) και x είναι η είσοδος που παρακάμπει αυτά τα επίπεδα [79] [95].

Η σημασία αυτής της πρόσθεσης είναι τεράστια για τη ροή της κλίσης. Κατά την οπίσθια διάδοση, η κλίση μπορεί πλέον να ρέει απρόσκοπτα μέσω της «διαδρομής ταυτότητας» ($+x$) απευθείας στα προηγούμενα επίπεδα. Ακόμα κι αν τα βάρη στα επίπεδα $F(x)$ είναι τέτοια που η κλίση τους μηδενίζεται (vanishing gradient), η κλίση από την ταυτότητα x (που είναι ίση με 1) εξακολουθεί να διαδίδεται προς τα πίσω [79].

Αυτό διασφαλίζει ότι ακόμη και τα πιο πρώιμα επίπεδα ενός εξαιρετικά βαθιού δικτύου λαμβάνουν πάντα ένα ισχυρό σήμα σφάλματος, επιτρέποντας στη σταθερή εκπαίδευση να συνεχιστεί.

4.9.3. Εκδοχές και Πλεονεκτήματα

Το βασικό πλεονέκτημα του ResNet είναι ότι επιτρέπει την επιτυχή και **σταθερή εκπαίδευση** δικτύων με πρωτοφανές **βάθος**. Η αρχιτεκτονική αυτή παρουσιάστηκε σε διάφορες εκδοχές, οι οποίες ονομάζονται με βάση τον συνολικό αριθμό των επιπέδων με βάρη [78][79]:

- **ResNet-18**
- **ResNet-34**
- **ResNet-50**
- **ResNet-101**
- **ResNet-152**

Οι βαθύτερες εκδοχές (50, 101, 152) χρησιμοποιούν επίσης μια βελτιστοποίηση που ονομάζεται "bottleneck" (λαιμός μπουκαλιού). Σε αυτήν, το $F(x)$ δεν αποτελείται από δύο συνελίξεις 3×3 , αλλά από τρεις: μια συνέλιξη 1×1 (για μείωση των καναλιών), ακολουθούμενη από τη συνέλιξη 3×3 , και τέλος μια συνέλιξη 1×1 (για την επαναφορά των καναλιών). Αυτή η τεχνική μειώνει κατά πολύ το υπολογιστικό κόστος και τον αριθμό των παραμέτρων, επιτρέποντας ακόμη μεγαλύτερο βάθος.

Ως αποτέλεσμα, το ResNet-50, παρόλο που είναι τρεις φορές βαθύτερο από το VGG-16 (50 επίπεδα έναντι 16), έχει στην πραγματικότητα πολύ λιγότερες παραμέτρους (περίπου 25.5 εκατομμύρια έναντι ~138 εκατομμυρίων του VGG), καθιστώντας το ταυτόχρονα βαθύτερο και πιο αποδοτικό υπολογιστικά [78] [79][96].

4.9.4. Εφαρμογές στην Ιστοπαθολογία (Ιατρικές Εικόνες)

Η ικανότητα του ResNet να μαθαίνει εξαιρετικά σύνθετες ιεραρχίες χαρακτηριστικών, χωρίς να υποφέρει από προβλήματα σύγκλισης, το έχει καταστήσει πρότυπο για πολλές εργασίες υπολογιστικής όρασης, συμπεριλαμβανομένης της ιατρικής απεικόνισης. Στην ψηφιακή παθολογία, όπου η ανάλυση των εικόνων WSIs απαιτεί την αναγνώριση εξαιρετικά πολύπλοκων και συχνά ανεπαίσθητων μορφολογικών μοτίβων, το ResNet κυριαρχεί [88][84][85].

- **Ανάλυση Καρκίνου του Μαστού (BreakHis dataset):** Πολλές ερευνητικές εργασίες έχουν αξιοποιήσει αρχιτεκτονικές ResNet (ιδίως το ResNet-50) για την ταξινόμηση ιστοπαθολογικών εικόνων καρκίνου του μαστού από το σύνολο δεδομένων **BreakHis**. Το BreakHis περιέχει χιλιάδες εικόνες βιοψιών μαστού, ταξινομημένες σε καλοήθειες και κακοήθειες όγκους, σε διαφορετικούς παράγοντες μεγέθυνσης (40x, 100x, 200x, 400x). Μελέτες που χρησιμοποιούν ResNet, συχνά σε συνδυασμό με Μάθηση Μεταφοράς (Transfer Learning), έχουν επιτύχει κορυφαία αποτελέσματα ακρίβειας στην αυτόματη διάκριση μεταξύ των τύπων όγκων σε αυτό το dataset [95].

- **Ανάλυση Ιστοπαθολογικών Βιοψιών Παχέος Εντέρου (CRC dataset):** Όπως τεκμηριώνεται εκτενώς στη βιβλιογραφία, οι αρχιτεκτονικές ResNet (ResNet-18, -34, -50) χρησιμοποιούνται ευρέως για την ανάλυση βιοψιών παχέος εντέρου. Εφαρμόζονται με επιτυχία σε σύνολα δεδομένων όπως το **NCT-CRC-HE-100K** (που αναφέρεται στο περίγραμμα της εργασίας) για την αυτόματη ταξινόμηση εννέα διαφορετικών τύπων ιστού (π.χ. στρώμα, καρκίνωμα, λεμφικός ιστός κ.λπ.). Επιπλέον, μοντέλα βασισμένα σε ResNet έχουν επιδείξει την ικανότητα να προβλέπουν προγνωστικούς βιοδείκτες, όπως η Μικροδορυφορική Αστάθεια (MSI) και η πενταετής επιβίωση, απευθείας από τις τυπικές εικόνες H&E [88][84][85].

4.10. Transfer Learning και Fine-tuning (Μάθηση Μεταφοράς και Λεπτομερής Ρύθμιση)

Η εκπαίδευση ενός βαθιού Συνελκτικού Νευρωνικού Δικτύου (CNN), όπως οι αρχιτεκτονικές VGG ή ResNet, από το μηδέν (δηλαδή με τυχαία αρχικοποιημένα βάρη) είναι μια διαδικασία υπολογιστικά δαπανηρή και, κυρίως, εξαιρετικά “λαίμαργη” σε δεδομένα. Τα δίκτυα αυτά, που διαθέτουν δεκάδες ή και εκατοντάδες εκατομμύρια παραμέτρους, απαιτούν τεράστιους όγκους επισημειωμένων δεδομένων εκπαίδευσης για να μπορέσουν να συγκλίνουν σωστά και να αποφύγουν την έντονη υπερ-προσαρμογή (overfitting) [78][79].

Ενώ σε γενικούς τομείς της υπολογιστικής όρασης υπάρχουν διαθέσιμα τεράστια σύνολα δεδομένων (όπως το ImageNet), στο πεδίο της ιατρικής απεικόνισης και ειδικότερα της ιστοπαθολογίας, η συλλογή και επισημείωση δεδομένων είναι μια διαδικασία εξαιρετικά δύσκολη, χρονοβόρα και δαπανηρή, καθώς απαιτεί την εμπλοκή εξειδικευμένων παθολόγων [82][94].

Για την επίλυση αυτού του προβλήματος, η κυρίαρχη προσέγγιση είναι η **Μάθηση Μεταφοράς (Transfer Learning)** .

4.10.1. Προεκπαίδευση σε ImageNet και Επαναχρησιμοποίηση Βαρών

Η Μάθηση Μεταφοράς είναι μια μεθοδολογία της Μηχανικής Μάθησης όπου ένα μοντέλο που αναπτύχθηκε για μια συγκεκριμένη εργασία (source task) επαναχρησιμοποιείται ως σημείο εκκίνησης για ένα μοντέλο σε μια δεύτερη, διαφορετική αλλά σχετική, εργασία (target task) [94].

Στην πράξη των CNNs, η διαδικασία αυτή αξιοποιεί την ικανότητα των βαθιών δικτύων να μαθαίνουν **ιεραρχικές αναπαραστάσεις χαρακτηριστικών** [79] [94]. Η διαδικασία έχει ως εξής:

1. **Προεκπαίδευση (Pre-training):** Μια πολύ βαθιά αρχιτεκτονική (π.χ. ResNet-50, VGG-16) εκπαιδεύεται σε ένα τεράστιο, γενικού σκοπού σύνολο δεδομένων. Το πιο συνηθισμένο είναι το **ImageNet**, το οποίο περιέχει πάνω από 1,2 εκατομμύρια εικόνες ταξινομημένες σε 1000 κατηγορίες (π.χ. “γάτα”, “σκύλος”, “αυτοκίνητο”, “αεροπλάνο”) [78][89].
2. **Μαθησιακά Χαρακτηριστικά:** Κατά τη διάρκεια αυτής της εκπαίδευσης, το δίκτυο μαθαίνει έναν πλούσιο σύνολο οπτικών χαρακτηριστικών . Τα πρώτα επίπεδα (κοντά στην είσοδο) μαθαίνουν να αναγνωρίζουν απλά, γενικά χαρακτηριστικά (π.χ. ακμές, γωνίες, χρώματα, υφές), ενώ τα βαθύτερα επίπεδα συνθέτουν αυτά τα απλά χαρακτηριστικά για να αναγνωρίσουν πιο σύνθετες δομές (π.χ. μάτια, τροχούς, μοτίβα) [81].
3. **Επαναχρησιμοποίηση Βαρών:** Η κεντρική ιδέα είναι ότι τα χαρακτηριστικά που έμαθε το δίκτυο στο ImageNet (ιδίως τα χαρακτηριστικά χαμηλού και μεσαίου επιπέδου, όπως οι ακμές και οι υφές) είναι χρήσιμα και για άλλα προβλήματα υπολογιστικής όρασης, συμπεριλαμβανομένης της ιατρικής απεικόνισης. Έτσι, αντί να εκπαιδεύσουμε ένα δίκτυο από την αρχή, φορτώνουμε το **προ-εκπαιδευμένο μοντέλο** μαζί με τα «παγωμένα» βάρη (weights) που έμαθε στο ImageNet [82][94].

4.10.2. Fine-tuning (Λεπτομερής Ρύθμιση) σε Ιατρικά Datasets

Η απλή επαναχρησιμοποίηση των βαρών δεν αρκεί, καθώς η τελική ταξινόμηση πρέπει να προσαρμοστεί στο νέο πρόβλημα (π.χ. από 1000 κλάσεις ImageNet σε 2 κλάσεις ιστοπαθολογίας). Η διαδικασία προσαρμογής του προ-εκπαιδευμένου μοντέλου στα νέα δεδομένα ονομάζεται **Fine-tuning (Λεπτομερής Ρύθμιση)** [82][83] [94].

Αυτή η προσέγγιση έχει αποδειχθεί εμπειρικά ότι είναι σχεδόν πάντα ανώτερη από την εκπαίδευση από το μηδέν σε ιατρικά σύνολα δεδομένων, ακόμη και αν τα ιατρικά δεδομένα (π.χ. ιστοπαθολογία) φαίνονται οπτικά πολύ διαφορετικά από τα δεδομένα του ImageNet.

Η διαδικασία του fine-tuning ακολουθεί συγκεκριμένα τεχνικά βήματα:

- 1. Αντικατάσταση της "Κεφαλής" (Classifier Head):** Το προ-εκπαιδευμένο δίκτυο αποτελείται από δύο μέρη: τη "βάση" (convolutional base) που εξάγει χαρακτηριστικά, και την «κεφαλή» (classifier head), που είναι συνήθως τα τελευταία πλήρως συνδεδεμένα επίπεδα. Αφαιρούμε την αρχική κεφαλή (π.χ. το FC επίπεδο των 1000 νευρώνων) και την αντικαθιστούμε με μια νέα, τυχαία αρχικοποιημένη κεφαλή που ταιριάζει στο δικό μας πρόβλημα (π.χ. ένα FC επίπεδο με 2 εξόδους για "καλοήθες" vs "κακοήθες", ακολουθούμενο από Softmax ή Sigmoid) [97].
- 2. Freezing Layers (Πάγωμα Επιπέδων):** Στα πρώτα στάδια του fine-tuning, είναι κρίσιμο να «παγώσουμε» τα βάρη ολόκληρης της συνελκτικής βάσης (ή τουλάχιστον των πρώτων επιπέδων της). Αυτό σημαίνει ότι τα βάρη αυτά δεν θα ενημερωθούν κατά την οπίσθια διάδοση. Ο λόγος είναι ότι η νέα κεφαλή είναι τυχαία και θα παράγει πολύ μεγάλες κλίσεις σφάλματος αρχικά. Αν αυτές οι κλίσεις διαδίδονταν ανεξέλεγκτα προς τα πίσω, θα «κατέστρεφαν» τα προσεκτικά ρυθμισμένα προ-εκπαιδευμένα βάρη. Σε αυτό το στάδιο, εκπαιδύουμε μόνο τη νέα κεφαλή, ώστε να μάθει να αντιστοιχίζει τα εξαγόμενα χαρακτηριστικά του ImageNet στις νέες μας ιατρικές κλάσεις [79][82].
- 3. Unfreezing (Απο-πάγωμα) και Ρύθμιση Learning Rate:** Αφού η νέα κεφαλή έχει εκπαιδευτεί για μερικές εποχές και έχει αρχίσει να συγκλίνει (το σφάλμα σταθεροποιείται), μπορούμε να "ξεπαγώσουμε" ολόκληρο το δίκτυο ή τα ανώτερα επίπεδα της συνελκτικής βάσης (π.χ. τα τελευταία μπλοκ ενός ResNet).
 - Αυτό που είναι **κρίσιμο** σε αυτό το στάδιο είναι η χρήση **διαφορικών ρυθμών μάθησης (differential learning rates)** [82].
 - Χρησιμοποιούμε έναν **πολύ μικρό ρυθμό μάθησης (learning rate)** (π.χ. $1e-5$) για τα προ-εκπαιδευμένα συνελκτικά επίπεδα, ώστε να τα «πειράξουμε» ελάχιστα και να τα προσαρμόσουμε στις λεπτές αποχρώσεις των ιατρικών εικόνων.
 - Διατηρούμε έναν (σχετικά) υψηλότερο ρυθμό μάθησης (π.χ. $1e-4$) για τη νέα κεφαλή, η οποία εξακολουθεί να μαθαίνει [94].

Η στρατηγική του Transfer Learning, όταν εφαρμόζεται σωστά, επιτρέπει σε αρχιτεκτονικές όπως το ResNet να επιτυγχάνουν κορυφαία αποτελέσματα σε ιατρικά σύνολα δεδομένων (όπως το BreakHis ή τα CRC datasets) με ένα κλάσμα του χρόνου και των δεδομένων που θα απαιτούνταν για εκπαίδευση από το μηδέν [88][95][93][97].

Κεφάλαιο 5ο

5. Εφαρμογές CNN Στη Βιοιατρική Εικόνα

5.1. Η Μεταμόρφωση της Ιατρικής Διάγνωσης μέσω της Υπολογιστικής Όρασης

Η ιατρική διάγνωση, τόσο στην ακτινολογία όσο και στην παθολογοανατομία, βασίζεται ιστορικά στην οπτική αναγνώριση προτύπων από εξειδικευμένους ιατρούς. Ωστόσο, η διαδικασία αυτή είναι απαιτητική, χρονοβόρα και, σε ορισμένες περιπτώσεις, εμπεριέχει στοιχεία υποκειμενικότητας που οδηγούν σε μεταβλητότητα μεταξύ διαφορετικών παρατηρητών [86][87]. Η ραγδαία αύξηση του όγκου των ιατρικών απεικονιστικών εξετάσεων, σε συνδυασμό με την αυξανόμενη πολυπλοκότητα των διαγνωστικών κριτηρίων, έχει δημιουργήσει μια επιτακτική ανάγκη για αυτοματοποιημένα, αντικειμενικά και αποδοτικά υπολογιστικά εργαλεία [81][85].

Η Βαθιά Μάθηση (Deep Learning - DL), και ειδικότερα τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs), έχουν αναδειχθεί ως η πλέον επαναστατική τεχνολογία για την αντιμετώπιση αυτής της πρόκλησης [89][83]. Αξιοποιώντας την ικανότητά τους να μαθαίνουν αυτόματα ιεραρχικές αναπαραστάσεις χαρακτηριστικών απευθείας από τα ακατέργαστα δεδομένα των εικονοστοιχείων (pixels), τα CNNs προσφέρουν πρωτοφανείς δυνατότητες στην ανάλυση βιοϊατρικών εικόνων [79][84]. Η εφαρμογή τους εκτείνεται σε όλο το φάσμα της ιατρικής απεικόνισης, από την ανάλυση ακτινογραφιών, αξονικών (CT) και μαγνητικών (MRI) τομογραφιών, έως τη μικροσκοπική ανάλυση δειγμάτων αίματος και, κυρίως, την ψηφιακή ιστοπαθολογία [81][85].

Σε αντίθεση με τις παραδοσιακές μεθόδους μηχανικής μάθησης που βασίζονταν σε χειροκίνητα σχεδιασμένα χαρακτηριστικά (hand-crafted features), τα CNNs μαθαίνουν τα βέλτιστα χαρακτηριστικά για μια συγκεκριμένη εργασία, επιτρέποντας την εκτέλεση πολύπλοκων διαγνωστικών και προγνωστικών αναλύσεων [89]. Οι κυριότερες εργασίες

που εκτελούν τα CNNs στη βιοϊατρική εικόνα μπορούν να κατηγοριοποιηθούν ως εξής [78][79][84]:

1. **Ταξινόμηση (Classification):** Η ανάθεση μιας μοναδικής ετικέτας σε ολόκληρη την εικόνα ή σε ένα τμήμα της (patch). Για παράδειγμα, η διάγνωση “καλοήθεις” έναντι “κακοήθεις” [88][95].
2. **Ανίχνευση/Εντοπισμός (Detection/Localization):** Ο προσδιορισμός της θέσης ενός ευρήματος μέσω ενός “πλαισίου οριοθέτησης” (bounding box). Για παράδειγμα, ο εντοπισμός της ακριβούς θέσης μιας μετάστασης σε έναν λεμφαδένα ή μιας ομάδας καρκινικών κυττάρων [88][84].
3. **Τμηματοποίηση (Segmentation):** Η πιο λεπτομερής εργασία, κατά την οποία ανατίθεται μια ετικέτα σε *κάθε* εικονοστοιχείο της εικόνας. Αυτό επιτρέπει την ακριβή οριοθέτηση των ορίων ενός όγκου ή τον διαχωρισμό διαφορετικών τύπων ιστού (π.χ. στρώμα, επιθήλιο, αδένες) [87][84].

Το παρόν κεφάλαιο θα εστιάσει στις εφαρμογές των CNNs στην ιστοπαθολογία, με ιδιαίτερη έμφαση στις τεχνικές που χρησιμοποιούνται για την ανάλυση βιοψιών παχέος εντέρου, αξιοποιώντας τις προηγμένες αρχιτεκτονικές που αναλύθηκαν στο προηγούμενο κεφάλαιο.

5.2. Επισκόπηση Εφαρμογών CNN στην Ιστοπαθολογία

Η ψηφιακή παθολογοανατομία, η οποία βασίζεται στη σάρωση γυάλινων πλακιδίων για τη δημιουργία Εικόνων Ολόκληρης Τομής (Whole Slide Images - WSIs), αποτελεί ιδανικό πεδίο εφαρμογής για τα CNNs. Μια WSI είναι μια εικόνα τεράστιου μεγέθους (συντά σε κλίμακα gigapixel) που περιέχει μορφολογικές πληροφορίες σε πολλαπλά επίπεδα ανάλυσης, από τη συνολική αρχιτεκτονική του ιστού έως τις λεπτομέρειες των μεμονωμένων κυτταρικών πυρήνων. Τα CNNs έχουν εφαρμοστεί με επιτυχία σε ένα ευρύ φάσμα τύπων καρκίνου, όπως του μαστού, του πνεύμονα, του δέρματος και του παχέος εντέρου [86][87].

5.2.1. Καρκίνος του Παχέος Εντέρου (CRC)

Ο καρκίνος του παχέος εντέρου (CRC) είναι ένας από τους πιο εκτενώς μελετημένους τύπους καρκίνου με μεθόδους Βαθιάς Μάθησης, όπως τεκμηριώνεται εκτενώς στη συστηματική ανασκόπηση των Davri et al. (2022)[88]. Οι εφαρμογές των CNNs σε ιστοπαθολογικές εικόνες CRC (συνήθως βαμμένες με Αιματοξυλίνη και Ηωσίνη - H&E) καλύπτουν όλο το διαγνωστικό και προγνωστικό φάσμα:

- **Διάγνωση και Ταξινόμηση Ιστού:** Η βασικότερη εφαρμογή είναι η αυτοματοποιημένη ταξινόμηση τμημάτων (patches) της WSI [88]. Εκπαιδεύονται μοντέλα για να διακρίνουν μεταξύ καρκινικού (π.χ. αδenoκαρκίνωμα) και μη καρκινικού ιστού, ή για να ταξινομήσουν λεπτομερώς τον ιστό σε πολλαπλές κατηγορίες, όπως φυσιολογικός βλεννογόνος, αδένωμα (προ-καρκινική αλλοίωση), και διηθητικό καρκίνωμα. Σύνολα δεδομένων όπως το **NCT-CRC-HE-100K** και τα δεδομένα από το **The Cancer Genome Atlas (TCGA)** χρησιμοποιούνται ευρέως για την εκπαίδευση μοντέλων που ταξινομούν τον ιστό σε κατηγορίες όπως στρώμα (Stroma), καρκίνος (Tumor), λεμφικός ιστός (Lymphocytes), βλέννη (Mucus) κλπ [88].
- **Πρόβλεψη Βιοδεικτών και Πρόγνωσης:** Πέρα από την απλή ταξινόμηση, τα CNNs επιδεικνύουν την αξιοσημείωτη ικανότητα να εξάγουν προγνωστικές πληροφορίες απευθείας από την H&E εικόνα. Έχουν αναπτυχθεί μοντέλα που προβλέπουν την πενταετή επιβίωση ή την πιθανότητα μετάστασης. Μια πρωτοποριακή εφαρμογή είναι η πρόβλεψη της κατάστασης **Μικροδορυφορικής Αστάθειας (Microsatellite Instability - MSI)**. Η MSI είναι ένας κρίσιμος βιοδείκτης που καθορίζει την ανταπόκριση των ασθενών στην ανοσοθεραπεία. Τα CNNs μπορούν να προβλέψουν την κατάσταση MSI/MSS (Microsatellite Stable) με υψηλή ακρίβεια, απλώς αναλύοντας τη μορφολογία του όγκου στην H&E εικόνα, εξοικονομώντας έτσι πόρους από δαπανηρές μοριακές εξετάσεις [88].

5.2.2. Καρκίνος του Μαστού (Breast Cancer)

Αντίστοιχες τεχνικές εφαρμόζονται ευρέως στον καρκίνο του μαστού. Το δημόσια διαθέσιμο σύνολο δεδομένων **BreakHis** έχει αποτελέσει σημείο αναφοράς για την εκπαίδευση μοντέλων CNN που ταξινομούν τις βιοψίες μαστού σε καλοήθειες (π.χ.

ιναδένωμα) και κακοήθειες (π.χ. πορογενές καρκίνωμα) υποτύπους, σε διαφορετικούς παράγοντες μεγέθυνσης (40x, 100x, 200x, 400x) [91][93][97].

5.2.3. Άλλοι Τύποι Ιστών (Πνεύμονας, Δέρμα, Εγκέφαλος)

Αν και το CRC και ο καρκίνος του μαστού κυριαρχούν στη βιβλιογραφία, οι ίδιες αρχές εφαρμόζονται και σε άλλους τύπους καρκίνου (π.χ. πνεύμονα, δέρματος). Μια θεμελιώδης εργασία σε όλους αυτούς τους τομείς είναι η ανάλυση του **μικροπεριβάλλοντος του όγκου (Tumor Microenvironment - TME)**. Τα CNNs χρησιμοποιούνται για την αυτόματη ανίχνευση και τμηματοποίηση κρίσιμων κυτταρικών πληθυσμών, όπως:

- **Ανίχνευση Λεμφοκυττάρων (Lymphocyte Detection):** Ο εντοπισμός και η ποσοτικοποίηση των λεμφοκυττάρων που διηθούν τον όγκο (Tumor-Infiltrating Lymphocytes - TILs) είναι ζωτικής σημασίας προγνωστικός δείκτης.
- **Τμηματοποίηση Πυρήνων (Nuclei Segmentation):** Η ακριβής οριοθέτηση των πυρήνων των καρκινικών κυττάρων επιτρέπει την εξαγωγή μορφομετρικών χαρακτηριστικών (π.χ. μέγεθος, σχήμα, υφή χρωματίνης) που σχετίζονται με τον βαθμό κακοήθειας (grading).

Επιπλέον, συγκεκριμένες αρχιτεκτονικές έχουν χρησιμοποιηθεί για την τμηματοποίηση όγκων εγκεφάλου από εικόνες MRI ή την ταξινόμηση κυττάρων αίματος [81][83].

5.3. ResNet: Η Αρχιτεκτονική Επιλογής για Ταξινόμηση και Πρόγνωση

Όπως αναφέρθηκε στο Κεφάλαιο 4, η οικογένεια αρχιτεκτονικών **ResNet (Residual Networks)** έφερε επανάσταση στη βαθιά μάθηση, επιλύοντας το πρόβλημα της εξαφανιζόμενης κλίσης μέσω των παρακαμπτήριων συνδέσεων (skip connections)[79]. Αυτό επέτρεψε την εκπαίδευση εξαιρετικά βαθιών δικτύων (π.χ. ResNet-50, ResNet-101) [79][78].

Στην ιστοπαθολογία, όπου τα διαγνωστικά μοτίβα είναι συχνά εξαιρετικά ανεπαίσθητα και ιεραρχικά, αυτό το βάθος είναι απαραίτητο[87][84]. Η ικανότητα του ResNet να

μαθαίνει πολύπλοκες συναρτήσεις το έχει καθιερώσει ως την προτιμώμενη αρχιτεκτονική-βάση (backbone) για εργασίες ταξινόμησης και πρόγνωσης [78][79].

5.3.1. Εφαρμογή: Ταξινόμηση Τύπων Ιστού σε Βιοψίες Παχέος Εντέρου

Η πιο άμεση εφαρμογή του ResNet στην ιστοπαθολογία του CRC είναι η ταξινόμηση εικόνων-τμημάτων (patches/tiles). Δεδομένου ότι μια WSI είναι υπερβολικά μεγάλη για να αναλυθεί ολόκληρη, η τυπική προσέγγιση είναι η «τεμαχισμός» της (tiling) σε χιλιάδες μικρότερα, διαχειρίσιμα τετράγωνα patches (π.χ. 224 X 224 pixels) [88][86].

Ένα μοντέλο ResNet, συνήθως ένα ResNet-18 ή ResNet-50 προ-εκπαιδευμένο στο ImageNet (αξιοποιώντας την τεχνική Transfer Learning), εκπαιδεύεται στη συνέχεια στο να ταξινομεί κάθε patch σε μία από τις προκαθορισμένες κλάσεις ιστού [82][97].

Ένα κλασικό παράδειγμα είναι το σύνολο δεδομένων **NCT-CRC-HE-100K**, το οποίο περιέχει 100.000 patches από βιοψίες CRC, ταξινομημένα σε 9 κατηγορίες[88]. Ένα μοντέλο βασισμένο σε ResNet μαθαίνει να διακρίνει μεταξύ:

1. **Tumor (TUM):** Διηθητικό αδενοκαρκίνωμα.
2. **Stroma (STR):** Συνδετικός ιστός που περιβάλλει τον όγκο.
3. **Lymphocytes (LYM):** Περιοχές πυκνής φλεγμονώδους διήθησης.
4. **Debris (DEB):** Νεκρωτικό υλικό.
5. **Mucus (MUC):** Βλέννη.
6. **Smooth Muscle (MUS):** Λείος μυϊκός ιστός.
7. **Adipose (ADI):** Λιπώδης ιστός.
8. **Normal Mucosa (NORM):** Φυσιολογικός βλεννογόνο.
9. **Background (BACK):** Κενές περιοχές του πλακιδίου.

Η εκπαίδευση ενός ResNet επιτρέπει την αυτόματη δημιουργία ενός σημασιολογικού χάρτη ολόκληρης της WSI, ο οποίος με τη σειρά του επιτρέπει ποσοτικές μετρήσεις, όπως ο υπολογισμός του **Λόγου Όγκου-προς-Στρώμα (Tumor-Stroma Ratio - TSR)**, ενός σημαντικού προγνωστικού δείκτη [88] .

5.3.2. Εφαρμογή: Πρόγνωση Μοριακών Βιοδεικτών (π.χ. MSI)

Μια πιο προηγμένη εφαρμογή του ResNet υπερβαίνει την απλή μορφολογική ταξινόμηση και εισέρχεται στο πεδίο της “υπολογιστικής παθολογίας” (computational pathology)[87]. Εδώ, το δίκτυο δεν εκπαιδεύεται να αναγνωρίζει αυτό που βλέπει ο παθολογοανατόμος (π.χ. "αδένες"), αλλά να προβλέπει μια αφηρημένη μοριακή ιδιότητα, όπως η κατάσταση MSI [88].

Η Μικροδορυφορική Αστάθεια (MSI) είναι μια κατάσταση γενετικής υπερμεταλλαξιμότητας που προκύπτει από βλάβη στο σύστημα επιδιόρθωσης αναντιστοιχιών DNA (dMMR) . Οι όγκοι CRC διακρίνονται σε MSI-High (MSI-H) και Microsatellite Stable (MSS). Αυτή η διάκριση είναι κρίσιμη, καθώς οι ασθενείς με MSI-H όγκους συχνά ανταποκρίνονται θεαματικά στην ανοσοθεραπεία, ενώ οι ασθενείς με MSS όχι.

Παραδοσιακά, η διάγνωση MSI απαιτεί είτε ανοσοϊστοχημεία (IHC) είτε μοριακή ανάλυση PCR. Ωστόσο, πολυάριθμες μελέτες έχουν αποδείξει ότι ένα CNN, συνήθως μια παραλλαγή του ResNet ή του Inception, μπορεί να εκπαιδευτεί για να προβλέψει την κατάσταση MSI/MSS **απευθείας από την τυπική H&E εικόνα** [88][84].

Το μοντέλο ResNet τροφοδοτείται με patches από WSIs ασθενών με γνωστή κατάσταση MSI/MSS (από μοριακό έλεγχο). Μέσω της εκπαίδευσης, το δίκτυο μαθαίνει να συσχετίζει ανεπαίσθητα μορφολογικά χαρακτηριστικά—πολλά από τα οποία δεν είναι εμφανή ή αξιόπιστα για τον άνθρωπο-παθολογοανατόμο—με την κατάσταση MSI [88]. Αυτά τα χαρακτηριστικά περιλαμβάνουν την πυκνότητα των TILs, συγκεκριμένους τύπους αδενικής δομής, ή την παρουσία βλέννης. Το αποτέλεσμα είναι ένα πανίσχυρο, χαμηλού κόστους εργαλείο screening που μπορεί να εντοπίσει ασθενείς υποψήφιους για ανοσοθεραπεία.

5.4. R-CNN: Ανίχνευση και Εντοπισμός Καρκινικών Περιοχών

Ενώ η ταξινόμηση patches (με ResNet) είναι χρήσιμη, συχνά δεν απαντά στο πιο κρίσιμο ερώτημα για έναν παθολογοανατόμο που εξετάζει μια WSI: «Πού ακριβώς βρίσκεται ο καρκίνος;». Αυτό δεν είναι πλέον πρόβλημα ταξινόμησης, αλλά **Ανίχνευσης Αντικειμένων (Object Detection)**[79].

Στόχος είναι η σχεδίαση ενός «πλαισίου οριοθέτησης» (bounding box) γύρω από κάθε καρκινική περιοχή. Η οικογένεια αρχιτεκτονικών **R-CNN (Region-based Convolutional Neural Networks)** είναι ειδικά σχεδιασμένη για αυτήν την εργασία[78] [79].

5.4.1. Η Εξέλιξη της Οικογένειας R-CNN

Οι αρχιτεκτονικές R-CNN είναι σχεδόν πάντα **δι-σταδιακές (two-stage)** [79]:

1. **Στάδιο 1: Δημιουργία Προτάσεων Περιοχών (Region Proposal)**. Το δίκτυο εντοπίζει υποψήφιες περιοχές (Regions of Interest - RoIs) που *ενδέχεται* να περιέχουν ένα αντικείμενο.
 2. **Στάδιο 2: Ταξινόμηση Περιοχών**. Οι προτεινόμενες περιοχές ταξινομούνται (και το πλαίσιο τους βελτιώνεται).
- **R-CNN (2014)**: Η αρχική έκδοση χρησιμοποιούσε έναν εξωτερικό αλγόριθμο (Selective Search) για να προτείνει ~2000 RoIs, και στη συνέχεια κάθε μία από αυτές τροφοδοτούνταν ξεχωριστά σε ένα CNN (π.χ. AlexNet) για ταξινόμηση. Ήταν εξαιρετικά αργό [78] [79].
 - **Fast R-CNN (2015)**: Βελτίωσε την ταχύτητα τροφοδοτώντας ολόκληρη την εικόνα στο CNN *μία* φορά για την εξαγωγή χαρακτηριστικών, και στη συνέχεια εφάρμοξε τις προτάσεις RoIs (που παράγονταν ακόμα εξωτερικά) πάνω στον χάρτη χαρακτηριστικών [79].
 - **Faster R-CNN (2015)**: Αυτή ήταν η μεγάλη τομή. Ενσωμάτωσε το πρώτο στάδιο μέσα στο δίκτυο. Εισηγάγε το **Region Proposal Network (RPN)**, ένα μικρό δίκτυο που «γλιστράει» πάνω στον χάρτη χαρακτηριστικών και μαθαίνει να προτείνει τα ίδια τα RoIs. Χρησιμοποιώντας μια κοινή συνελκτική βάση (π.χ. VGG ή ResNet) τόσο για το RPN όσο και για τον τελικό ταξινομητή, το Faster R-CNN πέτυχε ανίχνευση σε πραγματικό χρόνο [79].

- **Mask R-CNN (2017):** Αποτελεί την εξέλιξη του Faster R-CNN. Πέρα από την ταξινόμηση και το bounding box, το Mask R-CNN προσθέτει έναν τρίτο, παράλληλο κλάδο (branch) ο οποίος παράγει μια υψηλής ποιότητας **μάσκα τμηματοποίησης (segmentation mask)** για κάθε αντικείμενο που ανιχνεύεται [79][84].

5.4.2. Εφαρμογή: Εντοπισμός Ευρυμάτων σε Βιοψίες Παχέος Εντέρου

Στο πλαίσιο της διπλωματικής εργασίας, η ανίχνευση περιοχών ενδιαφέροντος σε βιοψίες παχέος εντέρου είναι μια κλασική εφαρμογή για τις αρχιτεκτονικές Faster R-CNN και Mask R-CNN [88][84].

Ροή Εργασίας:

1. **Backbone:** Χρησιμοποιείται μια ισχυρή προ-εκπαιδευμένη συνελκτική βάση, σχεδόν πάντα ένα **ResNet** (π.χ. ResNet-50 ή ResNet-101), για την εξαγωγή χαρακτηριστικών από την WSI (ή μεγάλα τμήματά της) [79][95].
2. **Εκπαίδευση:** Το μοντέλο (π.χ. Faster R-CNN) εκπαιδεύεται σε ένα σύνολο δεδομένων ιστοπαθολογίας όπου οι παθολογοανατόμοι έχουν επισημειώσει χειροκίνητα τις περιοχές ενδιαφέροντος με bounding boxes (π.χ. δεδομένα από το TCGA) [88][86].
3. **Εξαγωγή (Inference):** Όταν το εκπαιδευμένο μοντέλο τροφοδοτείται με μια νέα, αθέατη WSI, το RPN παράγει χιλιάδες προτάσεις περιοχών. Κάθε περιοχή στη συνέχεια ταξινομείται από το δίκτυο ως «υγιής» ή «μη-υγιής». Το τελικό αποτέλεσμα είναι μια λίστα συντεταγμένων (bounding boxes) που περιγράφουν με ακρίβεια τις προβληματικές περιοχές μέσα στον ιστό [84][85].

Αυτή η προσέγγιση είναι θεμελιώδης για την κλινική πράξη, καθώς επιτρέπει στον παθολογοανατόμο να εστιάσει αμέσως στις κρίσιμες περιοχές, επιταχύνοντας τη διάγνωση και επιτρέποντας την αυτόματη ποσοτικοποίηση του ποσοστού κολλαγόνου [86].

Κεφάλαιο 6ο

6. Ανάλυση ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου με χρήση Mask R-CNN για μέτρηση ποσοστού κολλαγόνου

Στο παρόν κεφάλαιο παρουσιάζεται η πειραματική μεθοδολογία που χρησιμοποιήθηκε για την τμηματοποίηση ιστολογικών περιοχών και στη συνέχεια ο υπολογισμός του ποσοστού κολλαγόνου, σε ιστοπαθολογικές εικόνες από βιοψίες του παχέος εντέρου. Αρχικά, περιγράφεται το σύνολο των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Στη συνέχεια, παρουσιάζεται η εφαρμογή του μοντέλου Mask R-CNN για την τμηματοποίηση ιστολογικών κλάσεων, καθώς και η διαδικασία παραγωγής των τελικών масκών. Έπειτα, γίνεται ανάλυση του κολλαγόνου μέσω χρωματικής ομαδοποίησης, με την εφαρμογή του αλγορίθμου $k - means$, και υπολογισμός του δείκτη CPA για κάθε ιστολογική περιοχή. Τέλος, παρουσιάζονται τα πειραματικά αποτελέσματα, καθώς και η διαδικασία αξιολόγησης της τμηματοποίησης.

6.1. Ανάλυση Ιστοπαθολογικών Εικόνων

Ένα από τα πεδία της Υπολογιστικής Όρασης είναι η ανάλυση ιστοπαθολογικών εικόνων, μιας και στα χαρακτηριστικά τους συμπεριλαμβάνονται η υψηλή ανάλυση, η σύνθετη μορφολογία και η διαφοροποίηση σε επίπεδο υφής και χρωμάτων. Τα τελευταία χρόνια παρατηρείται συνεχόμενη αύξηση του ενδιαφέροντος και της ανάγκης για την ανάπτυξη αυτοματοποιημένων υπολογιστικών μεθόδων, τέτοιων ώστε να μπορούν να υποστηρίξουν και να λειτουργήσουν συμπληρωματικά στην αξιολόγηση από τους ειδικούς [98].

Η βαθιά μάθηση και ειδικότερα τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) έχουν συμβάλει στην ανάπτυξη μοντέλων τα οποία είναι σε θέση να εξάγουν σύνθετα χαρακτηριστικά από δεδομένα εικόνας. Τα μοντέλα αυτά έχουν, μέχρι σήμερα,

χρησιμοποιηθεί με επιτυχία σε διάφορες ιατρικές εφαρμογές όπως είναι η ανίχνευση όγκων, η τμηματοποίηση ιστών καθώς και η ανάλυση ιστολογικών δομών [99].

Στο πεδίο της υπολογιστικής όρασης (Computer Vision) η ανάλυση ψηφιακών εικόνων αφορά στην εξαγωγή πληροφορίας από δεδομένα υψηλής διάστασης, όπου κάθε εικόνα μπορεί να αναλυθεί ως δισδιάστατο σήμα ή ως πίνακας τιμών της έντασης και του χρώματος [98].

Η ανάλυση της εικόνας μπορεί να διαχειριστεί είτε ως πρόβλημα Ταξινόμησης (Classification), είτε ως πρόβλημα Τμηματοποίησης (Segmentation), ανάλογα με το επιδιωκόμενο αποτέλεσμα. Η ταξινόμηση αφορά στη χαρτογράφηση μιας εικόνας σε ένα σύνολο κλάσεων υλοποιώντας την κατάλληλη συνάρτηση από τον χώρο των εικόνων, στον χώρο των ετικετών [100]. Η τμηματοποίηση από την άλλη, αφορά σε πιο σύνθετο υπολογιστικό πρόβλημα καθώς προϋποθέτει την εκτίμηση μιας συνάρτησης που αντιστοιχίζει κάθε pixel της εικόνας σε μια περιοχή ενδιαφέροντος, διατηρώντας παράλληλα την χωρική πληροφορία [99].

Ο διαχωρισμός αυτός είναι ιδιαίτερα σημαντικός σε εφαρμογές ιατρικής απεικόνισης, λόγω του γεγονότος ότι η αξιόπιστη εξαγωγή ποσοτικών χαρακτηριστικών θέτει ως προϋπόθεση την ακριβή οριοθέτηση περιοχών ιστού σε επίπεδο pixel [98]. Ως εκ τούτου, η παρούσα εργασία αντιμετωπίζει το πρόβλημα ως πρόβλημα τμηματοποίησης (segmentation), όπου η απομόνωση πολλαπλών περιοχών ενδιαφέροντος κρίνεται απαραίτητη, μιας και ο στόχος δεν είναι η συνολική κατηγοριοποίηση των εικόνων αλλά ο ακριβής εντοπισμός επιμέρους ιστολογικών περιοχών, η ποσοτική ανάλυση κάθε περιοχής καθώς επίσης και η εξαγωγή μετρήσεων. Συνεπώς, απαιτείται η χρήση και αξιοποίηση σύγχρονων μοντέλων βαθιάς μάθησης για τμηματοποίηση, και συγκεκριμένα η χρήση του Mask R-CNN ως βασικό υπολογιστικό εργαλείο.

6.2. Ιστολογικές Περιοχές ως Κλάσεις Segmentation

Στο πλαίσιο της παρούσας εργασίας, οι ιστοπαθολογικές εικόνες βιοψίας του παχέος εντέρου εξετάζονται, από υπολογιστικής άποψης, ως δεδομένα για τμηματοποίηση (segmentation), όπου στόχος είναι ο διαχωρισμός της εικόνας σε διακριτές περιοχές ενδιαφέροντος. Οι περιοχές αυτές μοντελοποιούνται ως κλάσεις τμηματοποίησης στις οποίες αντιστοιχίζονται τα pixels της εικόνας που ανήκουν σε αυτές. Η πρακτική αυτή είναι ιδιαίτερα διαδεδομένη στην βιοϊατρική ανάλυση εικόνας, λόγω του ότι συμβάλλει

στην εξαγωγή ποσοτικών χαρακτηριστικών και δεν περιορίζεται μόνο στην εξαγωγή συνολικών περιγραφών [98].

Η συγκεκριμένη προσέγγιση επιτρέπει την αντιμετώπιση του προβλήματος ως πρόβλημα πολυκλασικής τμηματοποίησης σε επίπεδο pixel. Στόχος είναι να αντιστοιχιστεί σε κάθε pixel της εικόνας μια (προκαθορισμένη) κλάση έτσι ώστε να δημιουργηθούν μάσκες που επιτρέπουν την ανάλυση κάθε περιοχής. Υπολογιστικά, η τμηματοποίηση αναπαρίσταται ως πίνακας ετικετών, διαστάσεων $H \times W$ (ίδιο μέγεθος με την εικόνα), όπου κάθε θέση του αντιστοιχεί σε ένα pixel και περιλαμβάνει την κλάση στην οποία ανήκει (το pixel). Εναλλακτικά, μπορεί να αναπαρασταθεί και ως σύνολο από δυαδικές μάσκες, μια για κάθε κλάση διαστάσεων $H \times W \times C$, όπου C είναι ο αριθμός των κλάσεων, και οι οποίες ορίζουν τα pixels που ανήκουν στην αντίστοιχη περιοχή. Αυτού του είδους η τμηματοποίηση αποτελεί βασική αρχή βαθιάς μάθησης για χρήση σε μοντέλα βιοϊατρικών εφαρμογών [99].

Οι κλάσεις τμηματοποίησης που χρησιμοποιήθηκαν στο πλαίσιο του πειράματος αφορούν στα βασικά ιστολογικά στρώματα του εντερικού τοιχώματος. Επιλέχθηκαν με βάση τα μορφολογικά τους γνωρίσματα, όσον αφορά την εικόνα, σύμφωνα με την ιστολογία του γαστρεντερικού συστήματος [101]. Συγκεκριμένα, η βλεννογόνος (mucosa) η οποία αποτελεί το εσωτερικό στρώμα του τοιχώματος και το βασικό χαρακτηριστικό της είναι η υψηλή κυτταρική και η εμφάνιση εντερικών κρυπτών, γνωρίσματα τα οποία αποτυπώνονται στις ιστοπαθολογικές εικόνες ως έντονη υφή και διαφορετική χρωματική κατανομή. Η υποβλεννογόνος (submucosa) βρίσκεται ακριβώς κάτω από την βλεννογόνο, όπου τα κύρια συστατικά της είναι συνδετικός ιστός και αγγειακές δομές, και η οποία εμφανίζει διαφορετικά μορφολογικά και χρωματικά γνωρίσματα στις ιστοπαθολογικές εικόνες. Τέλος, ο μυϊκός χιτώνας (muscularis externa), αντιστοιχεί στο εξωτερικό στρώμα και εμφανίζει πιο ομοιογενή μορφολογία (σε σύγκριση με τα προηγούμενα), λόγω της ύπαρξης λείου μυϊκού ιστού, χαρακτηριστικό το οποίο διευκολύνει τον εντοπισμό του από υπολογιστικές μεθόδους τμηματοποίησης σε συστήματα βαθιάς μάθησης [102].

Η οριοθέτηση με ακριβή και σαφή τρόπο των παραπάνω κλάσεων επιτρέπει την επεξεργασία κάθε κλάσης ξεχωριστά και την εξαγωγή χρήσιμων ποσοτικών χαρακτηριστικών όπως είναι ο υπολογισμός του ποσοστού κολλαγόνου ανά περιοχή [98].

Το πρόβλημα μπορεί να περιγραφεί ως η εκτίμηση μιας συνάρτησης ταξινόμησης ανά pixel ως εξής:

$$f(x, y) \rightarrow \{mucosa, submucosa, muscularis\}$$

Όπου, x, y είναι οι συντεταγμένες του pixel και η έξοδος δίνει την κλάση στην οποία ανήκει.

Ο ορισμός του προβλήματος με αυτό τον τρόπο δείχνει ότι απαιτείται όχι ταξινόμηση αλλά τμηματοποίηση (segmentation) σε επίπεδο pixel. Ως βασικό μοντέλο τμηματοποίησης στην παρούσα εργασία επιλέχθηκε η χρήση του **Mask R-CNN**, το οποίο έχει τη δυνατότητα να παράγει μάσκες υψηλής ακρίβειας για κάθε περιοχή ενδιαφέροντος (κλάση), κάνοντας εφικτή στη συνέχεια την ποσοτική ανάλυση στις οριοθετημένες περιοχές [103].

6.3. Δεδομένα και Επισημάνσεις Αναφοράς

Το παρόν πείραμα βασίστηκε σε ένα σύνολο 500 ιστοπαθολογικών εικόνων από βιοψίες του παχέος εντέρου. Η κάθε εικόνα είναι ψηφιακή απεικόνιση υψηλής ανάλυσης, πράγμα που συμβάλλει άμεσα στην υπολογιστική πολυπλοκότητα της τμηματοποίησης αλλά και στην ανάλυση των χαρακτηριστικών ώστε να είναι διαθέσιμα για εξαγωγή συμπερασμάτων. Οι πιο κοινές διαστάσεις των εικόνων είναι 768 x 1368 pixels ενώ κυμαίνονται και ως 3600 x 5984 pixels.

Στόχος του πειράματος είναι η τμηματοποίηση τριών ιστολογικών περιοχών (βλεννογόνος, υποβλεννογόνος και μυϊκός χιτώνας) και η δημιουργία των αντίστοιχων μασκών σε επίπεδο pixel. Η προσέγγιση αυτή θεωρείται κατάλληλη για ανάλυση ιστοπαθολογικών εικόνων, μιας και η περιοχή ενδιαφέροντος είναι χωρικά καθορισμένη και απαιτεί ακριβή οριοθέτηση των περιοχών ιστού συγκεκριμένου τύπου, με αποτέλεσμα να εξασφαλίζεται ο αξιόπιστος υπολογισμός ποσοτικών χαρακτηριστικών ανά περιοχή.

Η τμηματοποίηση σε επίπεδο pixel αποτελεί πεδίο της μάθησης με επίβλεψη, κατά την οποία το μοντέλο εκπαιδεύεται με σκοπό να προσεγγίσει μια επιθυμητή έξοδο για κάθε εικόνα εισόδου. Για να επιτευχθεί αυτό, χρειάζεται η ύπαρξη ενός συνόλου επισημάνσεων (ground truth) που λειτουργούν ως αναφορά σωστού αποτελέσματος για το οποίο βελτιστοποιείται το μοντέλο κατά τη διάρκεια της εκπαίδευσης και αξιολογείται

κατά την εξαγωγή των αποτελεσμάτων. Στην ιστοπαθολογική απεικόνιση, οι επισημάνσεις παράγονται από ειδικούς μιας και η οριοθέτηση των περιοχών απαιτεί εξειδικευμένη γνώση [98].

Στην παρούσα εργασία οι επισημάνσεις (annotations) αφορούν σε τμηματοποίηση σε επίπεδο pixel τριών κλάσεων, βλεννογόνος (mucosa), υποβλεννογόνος (submucosa), μυϊκός χιτώνας (muscularis), όπου κάθε pixel της εικόνας αντιστοιχίζεται σε μια από τις κλάσεις αυτές. Η συγκεκριμένη προσέγγιση είναι συμβατή με μοντέλα βαθιάς μάθησης για βιοϊατρική απεικόνιση [99].

Οι επισημάνσεις (ground truth) αναπαρίστανται είτε ως label map $Y \in \{0,1,2\}^{H \times W}$, όπου κάθε pixel αναπαρίσταται με τον κωδικό της κλάσης, είτε ως σύνολο δυαδικών μασκών $\{M_c\}_{c=1}^3$, μια μάσκα ανά κλάση όπου:

$M_c(x, y) = 1$, όταν το pixel ανήκει στην κλάση

$M_c(x, y) = 0$, όταν το pixel δεν ανήκει στην κλάση.

Και οι δυο αναπαραστάσεις περιγράφουν την ίδια πληροφορία και χρησιμοποιούνται ισοδύναμα σε μοντέλα τμηματοποίησης.

Στην παρούσα εργασία η σειρά των κλάσεων είναι: [mucosa, muscularis, submucosa].

Στο σύνολο δεδομένων, οι 117 εικόνες διαθέτουν επισημάνσεις και χρησιμοποιούνται ως σύνολο εκπαίδευσης (Training Set). Έπειτα από την εκπαίδευση γίνεται εξαγωγή προβλέψεων (inference) για το σύνολο των εικόνων. Στην παρούσα εργασία ο όρος Training αφορά σε μετρήσεις που υπολογίζονται από τις μάσκες αναφοράς (ground truth), ενώ ο όρος Testing αφορά σε μετρήσεις που προκύπτουν από τις προβλεπόμενες μάσκες του μοντέλου.

Στα προβλήματα τμηματοποίησης η ποιότητα των επισημάνσεων αποτελεί σημαντικό παράγοντα, αφού τα δεδομένα αναφοράς είναι αυτά που καθορίζουν τον στόχο που προσπαθεί να προσεγγίσει το μοντέλο κατά τη διάρκεια της εκπαίδευσης. [104]. Στην ιστοπαθολογική απεικόνιση ενδέχεται να υπάρχει ασάφεια μεταξύ των ιστολογικών στρωμάτων, γεγονός που μπορεί να οδηγήσει σε διαφορές μεταξύ επισημαντών. Ως εκ τούτου, οι επισημάνσεις αποτελούν την καλύτερη διαθέσιμη αναφορά, αλλά ενδέχεται να περιέχουν αβεβαιότητα που πρέπει να λαμβάνεται υπόψη κατά την ερμηνεία των αποτελεσμάτων [104].

6.4. Τμηματοποίηση Ιστολογικών Περιοχών με Mask R-CNN

Στην παρούσα εργασία, η τμηματοποίηση των τριών ιστολογικών περιοχών (mucosa, muscularis, submucosa) γίνεται με τη χρήση του μοντέλου Mask R-CNN. Το συγκεκριμένο μοντέλο ανήκει στην κατηγορία instance segmentation όπου έχει την δυνατότητα να εντοπίζει περιοχές ενδιαφέροντος και να παράγει, για κάθε μια από αυτές, μάσκες τμηματοποίησης σε επίπεδο pixel [103]. Οι μάσκες που παράγονται, αποτελούν το βασικό αποτέλεσμα της διαδικασίας και στη συνέχεια χρησιμοποιούνται ως ποσοτική ανάλυση του κολλαγόνου.

Η βασική λειτουργία του Mask R-CNN βασίζεται σε διακριτά στάδια επεξεργασίας της εικόνας εισόδου. Αρχικά, εισάγεται η εικόνα σε ένα Συνελκτικό Νευρωνικό Δίκτυο κορμού (backbone), το οποίο δημιουργεί χάρτες χαρακτηριστικών (feature maps) που αφορούν τη μορφολογία και τα χρωματικά χαρακτηριστικά της εικόνας. Για να γίνει αποτελεσματικότερα η ανάλυση δομών διαφορετικής κλίμακας χρησιμοποιείται το Feature Pyramid Network (FPN), το οποίο έχει τη δυνατότητα να συνδυάζει πληροφορία από πολλαπλά επίπεδα ανάλυσης [105].

Έπειτα, εφαρμόζεται το Δίκτυο Περιοχών Προτάσεων (Region Proposal Network – RPN), για τον εντοπισμό υποψηφίων περιοχών στην εικόνα, οι οποίες αντιστοιχούν σε τμήματα της εικόνας που πιθανόν να περιέχουν ιστολογικές δομές. Για κάθε μια τέτοια περιοχή εξάγονται χαρακτηριστικά σταθερού μεγέθους, μέσω της διαδικασίας Roi Align, που διατηρεί με ακρίβεια τη χωρική πληροφορία.

Το μοντέλο στη συνέχεια παράγει τρεις τύπους εξόδου, για κάθε περιοχή ενδιαφέροντος, μέσω των κατάλληλων κεφαλών:

- α. Την πρόβλεψη της κλάσης.
- β. την βελτίωση του πλαισίου οριοθέτησης.
- γ. Τη Μάσκα Τμηματοποίησης.

Η Κεφαλή Μάσκας (Mask Head) παράγει δυαδικές μάσκες σε επίπεδο pixel για κάθε κλάση. Στην παρούσα εργασία δημιουργούνται μάσκες για τις κλάσεις mucosa, muscularis και submucosa, για κάθε περιοχή ενδιαφέροντος, από τις οποίες επιλέγεται η μάσκα προβλεπόμενης κλάσης.

Οι συγκεκριμένες μάσκες είναι η τελική έξοδος της διαδικασίας τμηματοποίησης οι οποίες χρησιμοποιούνται στην επόμενη φάση επεξεργασίας ως χωρικά φίλτρα. Αυτό που

κάνουν συγκεκριμένα, είναι ότι περιορίζουν την ανάλυση μόνο στα pixels που ανήκουν σε κάθε περιοχή, καθιστώντας έτσι δυνατό τον υπολογισμό του ποσοστού κολλαγόνου (CPA) ξεχωριστά για κάθε κλάση. Ως εκ τούτου, η ακρίβεια της τμηματοποίησης καθορίζει άμεσα την αξιοπιστία των ποσοτικών μετρήσεων που ακολουθούν [103].

6.5. Mask R-CNN: Ροή Επεξεργασίας και παραγωγή τελικών μασκών

Στην παρούσα ενότητα παρουσιάζεται η ροή επεξεργασίας μέσω της οποίας η έξοδος του μοντέλου μετατρέπεται σε τελικές μάσκες τμηματοποίησης ανά κλάση. Στόχος είναι η παραγωγή ενός ενιαίου πολυκλασικού αποτελέσματος που αν προκύπτουν διακριτές περιοχές για τις τρεις κλάσεις (mucosa, muscularis, submucosa) έτσι ώστε να μπορεί να γίνει η ποσοτική ανάλυση του κολλαγόνου στην επόμενη φάση.

Το Mask R-CNN παράγει ένα σύνολο προβλέψεων για κάθε εικόνα εισόδου [103]. Κάθε πρόβλεψη αντιστοιχεί σε μια υποψήφια περιοχή ενδιαφέροντος και περιλαμβάνει: την **προβλεπόμενη κλάση**, το **πλαίσιο οριοθέτησης** (Bounding Box), το **βαθμό εμπιστοσύνης** και την **δυναδική μάσκα** σε επίπεδο pixel. Οι μάσκες που παράγονται, ενδεχομένως να μην αποτελούν το τελικό αποτέλεσμα τμηματοποίησης, αφού υπάρχει η πιθανότητα να υπάρχουν πολλές επιμέρους προβλέψεις για την ίδια εικόνα, ακόμη και με χωρική επικάλυψη. Για τον λόγο αυτό, οι επιμέρους προβλέψεις ενοποιούνται έτσι ώστε να παραχθεί ένας πολυκλασικός χάρτης ετικετών, στον οποίο κάθε pixel αντιστοιχίζεται σε μια και μόνο κλάση ιστού. Σε περιπτώσεις όπου επικαλύπτονται χωρικά διαφορετικές περιοχές, εφαρμόζεται ο κανόνας επιλογής με βάση τον μέγιστο βαθμό εμπιστοσύνης, όπου το pixel αντιστοιχίζεται στην κλάση της πρόβλεψης με την μεγαλύτερη τιμή βαθμού εμπιστοσύνης στην θέση αυτή. Έτσι αποφεύγεται η ανάθεση της ίδιας θέσης σε περισσότερες από μια κλάσεις.

Από τον τελικό χάρτη ετικετών που προκύπτει, εξάγονται τρεις δυναδικές μάσκες, μια για κάθε ιστολογική περιοχή:

- M_{mucosa}
- $M_{\text{muscularis}}$
- $M_{\text{submucosa}}$

Σε κάθε κλάση, τα pixels που ανήκουν στην κλάση αναπαρίστανται με την τιμή 1, ενώ όλα τα υπόλοιπα με την τιμή 0. Οι συγκεκριμένες μάσκες αποτελούν την τελική έξοδο της τμηματοποίησης, χρησιμοποιούνται δηλαδή ως χωρικά φίλτρα για τον εντοπισμό των pixels κάθε ιστολογικής περιοχής.

Με βάση τις τελικές μάσκες που παράγονται, υπολογίζεται επίσης και το εμβαδό κάθε περιοχής σε pixels, το πλήθος δηλαδή των pixels που ανήκουν στην αντίστοιχη κλάση (Testing Area Class mucosa, Testing Area Class muscularis, Testing Area class submucosa). Οι τιμές αυτές που προκύπτουν, χρησιμοποιούνται στην επόμενη φάση για την χρωματική ανάλυση με χρήση $k - means$ για τον υπολογισμό του δείκτη CPA ανά ιστολογική περιοχή. Ως εκ τούτου, το παρόν στάδιο λειτουργεί συνδυαστικά μεταξύ της τμηματοποίησης και της ποσοτικής ανάλυσης του κολλαγόνου. Το τελικό label map που παράγεται, καθώς και οι δυαδικές μάσκες επηρεάζουν άμεσα την αξιοπιστία του υπολογισμού του ποσοστού κολλαγόνου (CPA) αφού καθορίζουν σε ποια pixels θα γίνει η επόμενη φάση επεξεργασίας.

Χρήση $k - means$ για ανάλυση κολλαγόνου και υπολογισμός CPA.

Έπειτα από την παραγωγή των τελικών μαस्कών τμηματοποίησης για τις κλάσεις mucosa, muscularis, submucosa, ακολουθεί η ποσοτική ανάλυση του κολλαγόνου για κάθε ιστολογική περιοχή. Η ανάλυση γίνεται με την χρήση του αλγορίθμου $k - means$ clustering σε επίπεδο pixel και βασίζεται στην χρωματική πληροφορία των εικόνων. Αρχικά, για κάθε ιστολογική περιοχή (mucosa, muscularis, submucosa) χρησιμοποιείται η αντίστοιχη δυαδική μάσκα που προέκυψε κατά το στάδιο της τμηματοποίησης, για να περιοριστεί η ανάλυση στα pixels της συγκεκριμένης περιοχής. Έτσι, διασφαλίζεται ότι ο υπολογισμός του ποσοστού κολλαγόνου θα γίνει ανεξάρτητα για κάθε στρώμα ιστού.

Στη συνέχεια, τα pixels της περιοχής μετατρέπονται σε διανύσματα χρώματος στον χώρο RGB όπου και εφαρμόζεται ο αλγόριθμος $k - means$ με πλήθος συστάδων $k = 3$. Ο αλγόριθμος ομαδοποιεί τα pixels της περιοχής σε τρεις ομάδες με βάση την χρωματική τους ομοιότητα και υπολογίζει τα αντίστοιχα κέντρα (centroids) κάθε συστάδας. Από τις τρεις συστάδες, επιλέγεται ως “κόκκινη” αυτή της οποίας το κέντρο είναι πλησιέστερο σε ένα χρώμα – στόχο στον χώρο RGB. Η επιλογή γίνεται με βάση τον υπολογισμό της Ευκλείδειας απόστασης μεταξύ των κέντρων των συστάδων και του χρώματος – στόχου, εντοπίζοντας έτσι την συστάδα που αντιστοιχεί στο κόκκινο χρώμα της εικόνας[106]. Έπειτα, δημιουργείται δυαδική μάσκα κολλαγόνου, όπου τα pixels που ανήκουν στην

κόκκινη συστάδα αναπαρίστανται με την τιμή 1, ενώ τα υπόλοιπα με την τιμή 0. Η μάσκα αυτή αφορά στην περιοχή της εικόνας που περιέχει κολλαγόνο. Με βάση αυτή τη δυαδική μάσκα υπολογίζεται το πλήθος των pixels που αντιστοιχούν σε κολλαγόνο και ορίζεται ως Fibrosis Area. Επιπλέον, από τη μάσκα της ιστολογικής περιοχής έχει προκύψει και το συνολικό πλήθος των pixels της περιοχής, το οποίο ορίζεται ως Area.

Ο δείκτης CPA (Collagen Proportionate Area) για κάθε ιστολογική περιοχή υπολογίζεται ως:

$$CPA = \frac{Fibrosis\ Area}{Area} \times 100$$

Η διαδικασία αυτή εφαρμόζεται ανεξάρτητα για κάθε κλάση, κάνοντας εφικτή τον υπολογισμό ποσοστού κολλαγόνου ανά ιστολογική περιοχή.

Η χρήση των масκών τμηματοποίησης καθώς και η χρωματική ομαδοποίηση μέσω k – means, συμβάλλουν έτσι ώστε ο υπολογισμός του κολλαγόνου να γίνεται στοχευμένα και μόνο στις περιοχές ενδιαφέροντος, χωρίς να επιδρούν σε αυτό άσχετες δομές ιστού.

6.6. Αξιολόγηση Τμηματοποίησης

Η αξιολόγηση της απόδοσης τμηματοποίησης έγινε σε επίπεδο pixel με σύγκριση των προβλεπόμενων масκών που παράγει το μοντέλο, με τις αντίστοιχες μάσκες αναφοράς (ground truth). Αυτή η διαδικασία εφαρμόζεται στο επισημασμένο σύνολο δεδομένων, που διαθέτουν δηλαδή διαθέσιμες χειροκίνητες επισημάνσεις.

6.6.1. Πίνακας σύγχυσης

Για την ποσοτική αξιολόγηση της τμηματοποίησης χρησιμοποιείται ο Πίνακας Σύγχυσης (Confusion Matrix), στον οποίο αποτυπώνεται η αντιστοίχιση μεταξύ των πραγματικών και των προβλεπόμενων κλάσεων σε επίπεδο pixel με σειρά κλάσεων:

[mucosa, muscularis, submucosa]

Κάθε στοιχείο του πίνακα ορίζεται ως:

$$CM_{i,j} = \{pixels\ πραγματικής\ κλάσης\ i\ και\ προβλεπόμενη\ κλάση\ j\}$$

Κάθε γραμμή του Πίνακα Σύγχυσης αντιστοιχεί στην **πραγματική κλάση** (ground truth), ενώ κάθε **στήλη** του αντιστοιχεί στην **προβλεπόμενη κλάση** από το μοντέλο. Τα

στοιχεία δηλαδή του πίνακα αποτελούν το πλήθος των pixels που ταξινομήθηκαν σε κάθε συνδυασμό πραγματικής και προβλεπόμενης κλάσης. Με τον τρόπο αυτό, γίνεται ο εντοπισμός τόσο των σωστών ταξινομήσεων, όσο και των σφαλμάτων μεταξύ των διαφορετικών ιστολογικών περιοχών.

6.6.2. Μετρικές ανά κλάση (Sensitivity, Specificity, F1 Score) και Συνολικές Μετρικές (Total Accuracy).

Σε πολυκλασικό πρόβλημα τα **TP, FP, FN, TN** για κάθε κλάση c ορίζονται με την λογική η κλάση c έναντι όλων των άλλων.

Για κάθε κλάση c ορίζονται:

TP_c: pixels με πραγματική κλάση (GT) = c και προβλεπόμενη κλάση (Pred) = c .

FN_c: pixels με πραγματική κλάση (GT) = c και προβλεπόμενη κλάση (Pred) $\neq c$.

FP_c: pixels με πραγματική κλάση (GT) $\neq c$ και προβλεπόμενη κλάση (Pred) = c .

TN_c: pixels με πραγματική κλάση (GT) $\neq c$ και προβλεπόμενη κλάση (Pred) $\neq c$.

Από τον πίνακα σύγχυσης υπολογίζονται οι βασικές μετρικές αξιολόγησης της τμηματοποίησης ανά εικόνα και για κάθε κλάση ξεχωριστά.

Η **Ευαισθησία (Sensitivity)** εκφράζει το κατά πόσο το μοντέλο εντοπίζει σωστά τα pixels που ανήκουν σε μια συγκεκριμένη κλάση και υπολογίζεται ως:

$$Sensitivity = \frac{TP}{TP + FN}$$

Η **Ειδικότητα (Specificity)** εκφράζει την ικανότητα του μοντέλου να μην κάνει εσφαλμένη ταξινόμηση pixels που ανήκουν σε άλλες κλάσεις, στην εξεταζόμενη κλάση και υπολογίζεται ως:

$$Speicificity = \frac{TN}{TN + FP}$$

Ο δείκτης **F1-Score** είναι μετρική που ισορροπεί την επίδραση των σφαλμάτων (FP και FN). Υψηλό F1-Score δείχνει υψηλό ποσοστό αναγνώρισης της κλάσης (χαμηλό FN) και περιορισμό της λανθασμένης απόδοσης της κλάσης σε pixels άλλων κλάσεων (χαμηλό FP). Ο δείκτης F1-Score υπολογίζεται ως:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Επίσης, ως **Συνολική Ακρίβεια (Total Accuracy)**, ανά εικόνα, ορίζεται το ποσοστό των pixels που ταξινομήθηκαν σωστά και υπολογίζεται ως εξής:

$$Total Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Οι συγκεκριμένες μετρικές υπολογίζονται ανά εικόνα και ανά κλάση και συνοψίζονται στη συνέχεια μέσω της μέσης τιμής και της τυπικής απόκλισης.

Σε κάποιες από τις εικόνες μια ή περισσότερες από τις τρεις κλάσεις ενδέχεται να απουσιάζουν από τις επισημάνσεις αναφοράς (ground truth), να έχουν δηλαδή μηδενικό εμβαδό, γεγονός που κάνει αδύνατο τον ορισμό κάποιων από τις μετρικές. Για τον λόγο αυτό, στις περιπτώσεις όπου μια κλάση απουσιάζει, στον υπολογισμό των συνολικών στατιστικών εξαιρούνται από τον υπολογισμό της μέσης τιμής. Έτσι εξασφαλίζεται ότι οι μετρικές αξιολόγησης αφορούν μόνο τις περιπτώσεις όπου η συγκεκριμένη κλάση υπάρχει.

6.7. Αποτελέσματα

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα τμηματοποίησης των τριών ιστολογικών περιοχών και της ανάλυσης του ποσοστού κολλαγόνου εντός αυτών. Τα αποτελέσματα προέρχονται από την εφαρμογή της προτεινόμενης μεθοδολογίας στο σύνολο των δεδομένων και παρουσιάζονται μέσω στατιστικών μεγεθών. Στους πίνακες αποτελεσμάτων οι συνεχείς μεταβλητές συνοψίζονται σε μέση τιμή \pm τυπική απόκλιση. Επίσης, σε κάποιες περιπτώσεις, όπου παρατηρείται ασυμμετρία της κατανομής ή ακόμη και ύπαρξη ακραίων τιμών, αναφέρονται και η διάμεσος με τα τεταρτημόρια $Q1 - Q3$, το εύρος δηλαδή μέσα στο οποίο βρίσκεται το 50% των τιμών και $IQR = Q1 - Q3$.

6.7.1. Αποτελέσματα Τμηματοποίησης

Για την απόδοση της τμηματοποίησης χρησιμοποιήθηκε το επισημασμένο υποσύνολο δεδομένων, μέσω του πίνακα σύγχυσης, σε επίπεδο pixel. Ο πίνακας σύγχυσης εμφανίζει το συνολικό πλήθος των pixels για κάθε συνδυασμό πραγματικής και προβλεπόμενης κλάσης. Έτσι, μπορούμε να έχουμε μια συνολική εικόνα της συμπεριφοράς του μοντέλου. Ο πίνακας σύγχυσης κανονικοποιείται ανά γραμμή, ώστε να εκφράζονται οι τιμές σε ποσοστά επί τοις εκατό, επιτρέποντας έτσι την αξιολόγηση του μοντέλου για κάθε πραγματική κλάση ξεχωριστά.

GT /Pred	mucosa	muscularis	submucosa
mucosa	98.68%	0,97%	0,36%
muscularis	9,30%	84,17%	6,54%
submucosa	5,28%	5,48%	89,24%

Πίνακας 1 Κανονικοποιημένος πίνακας σε ποσοστά (Confussion Matrix Percentages)

Γραμμές = Πραγματική κλάση (GT)

Στήλες = Πρόβλεψη (Pred)

Σύμφωνα με τα αποτελέσματα, παρατηρείται υψηλή ακρίβεια στην αναγνώριση της κλάσης mucosa, ενώ στις κλάσεις muscularis και submucosa παρατηρούνται μεγαλύτερες αποκλίσεις.

Μετρικές Τμηματοποίησης Sensitivity, Specificity, F1 – Score, Total Accuracy

Οι τιμές των μετρικών αξιολόγησης (Sensitivity, Specificity, F1-score και Total Accuracy) υπολογίστηκαν ανά εικόνα και ανά κλάση και τα αποτελέσματα παρουσιάζονται μέσω της μέσης τιμής και τυπικής απόκλισης.

Σύνοψη της απόδοσης τμηματοποίησης του μοντέλου ανά ιστολογική κλάση στο επισημασμένο υποσύνολο σε επίπεδο pixel:

Κλάση	Sensitivity			Specificity			F1 - Score		
	n	Mean	± SD	n	Mean	± SD	n	Mean	± SD
mucosa	116	0,985	± 0,33	110	0,894	± 0,177	114	0,984	±0,032
muscularis	105	0,731	± 0,238	105	0,979	± 0,021	103	0,642	±0,224
submucosa	102	0,813	± 0,213	102	0,989	± 0,33	102	0,849	±0,185

Πίνακας 2 Μετρικές ανά κλάση - Μέση τιμή ± SD τυπική απόκλιση

Συνολική ακρίβεια (Total Accuracy) για επισημασμένο υποσύνολο:

Για n = 116 έχουμε:

Total Accuracy						
Mean	± SD	Διάμεσος (Median)	Q1	Q3	IQR (Q3 – Q1)	Ελάχιστο – Μέγιστο (Minn – Max)
0,961	± 0,047	0,972 [0,954 – 0,986]	0,954	0,986	0,032	0,618 – 1,000

Πίνακας 3 Συνολική Ακρίβεια (Total Accuracy)

Τα αποτελέσματα των παραπάνω μετρικών επιβεβαιώνουν την σταθερή απόδοση του μοντέλου, με μικρή διακύμανση να παρατηρείται μεταξύ των εικόνων.



Σχήμα 8 Παραδείγματα αποτελεσμάτων Τμηματοποίησης στο σύνολο *Testing*

Στις εικόνες του Σχήματος 8, απεικονίζονται ενδεικτικά παραδείγματα εικόνων εισόδου με τις αντίστοιχες προβλεπόμενες μάσκες για τις περιοχές *mucosa* (πράσινη απόχρωση) και *muscularis* (ροζ απόχρωση). Παρατηρούμε ότι το μοντέλο έχει εντοπίσει με μεγάλη ακρίβεια τις κύριες περιοχές ιστού, μόνο που σε κάποιες περιπτώσεις εμφανίζονται μικρές αποκλίσεις στα όρια των περιοχών. Παρατηρούμε επίσης ύπαρξη μικρών απομονωμένων περιοχών, που πιθανότατα να σχετίζονται με την ύπαρξη θορύβου ή με την δυσκολία διαφοροποίησης μεταξύ διαφορετικών ιστολογικών στρωμάτων.



Σχήμα 9 Ενδεικτικές επισημάνσεις σε εικόνες του συνόλου Training

Στις εικόνες του Σχήματος 9 παρουσιάζονται ενδεικτικά παραδείγματα εικόνων στις οποίες απεικονίζονται οι περιοχές των κλάσεων, όπως έχουν οριστεί μέσω χειροκίνητων επισημάνσεων. Οι συγκεκριμένες επισημάνσεις δείχνουν τις ιστολογικές περιοχές ενδιαφέροντος και χρησιμοποιήθηκαν για την εκπαίδευση αλλά και για την αξιολόγηση της απόδοσης του μοντέλου.

6.7.2. Αποτελέσματα Ποσοστού Κολλαγόνου (CPA) ανά ιστολογική περιοχή

Το κύριο αποτέλεσμα στο οποίο εστιάζουμε στην παρούσα εργασία, είναι ο δείκτης CPA (%) ανά ιστολογική περιοχή, έτσι όπως έχει υπολογιστεί από τις προβλεπόμενες μάσκες του μοντέλου (CPA Testing Class mucosa / muscularis / submucosa). Για κάθε εικόνα και για κάθε κλάση υπολογίστηκε το CPA, που αποτελεί υπολογιστικό δείκτη της ύπαρξης κόκκινου χρώματος μέσα σε κάθε ιστολογική περιοχή και χρησιμοποιείται ως ένδειξη ύπαρξης κολλαγόνου, ποσοτικός δείκτης δηλαδή μορφολογικών αλλαγών του ιστού. Υπολογίζεται δηλαδή το ποσοστό των pixels που αντιστοιχούν σε κολλαγόνο, ως προς το συνολικό εμβαδό της περιοχής.

Για την στατιστική ανάλυση του CPA (%) έγινε περιγραφική ανάλυση των τιμών CPA Testing Class mucosa, muscularis και submucosa, για όλες τις εικόνες των τιμών Cpa Training Class των ίδιων κλάσεων για το επισημασμένο υποσύνολο. Για κάθε κλάση υπολογίζεται η μέση τιμή \pm τυπική απόκλιση (mean \pm SD) η διάμεσος (median) το 1ο τεταρτημόριο Q1 (25ο εκατοστημόριο) το 3ο τεταρτημόριο Q3 (75ο εκατοστημόριο) το ενδοτεταρτημοριακό εύρος (IQR = Q3 – Q1) όπως και τα ελάχιστο – μέγιστο (min – max) και παρουσιάζονται τα αποτελέσματα στους παρακάτω πίνακες.

Ποσοστό CPA Testing στο σύνολο των εικόνων

Κλάση	Μέση Τιμή (%)	\pm Τυπική Απόκλιση (%)	Διάμεσος (%)	Q1 (%)	Q3(%)	IQR (=Q3 – Q1)	Ελάχιστο – Μέγιστο (%)
Mucosa	24,53	20,27	18,40	11,37	29,79	18,43	0 - 99,77
Muscularis	38,26	30,94	44,32	0,00	63,19	63,19	0 - 102,78
Submucosa	36,42	28,51	38,19	0,00	58,69	58,69	0 - 108,86

Πίνακας 4 CPA Testing % στο σύνολο των εικόνων N = 500

Ποσοστό CPA Testing όπου Testing Area > 0

Κλάση	Μέση Τιμή (%)	\pm Τυπική Απόκλιση (%)	Διάμεσος (%)	Q1 (%)	Q3(%)	IQR (=Q3 – Q1)	Ελάχιστο – Μέγιστο (%)
Mucosa	24,78	20,22	18,70	11,61	29,87	18,26	0 - 99,77
Muscularis	55,82	20,39	56,39	42,97	69,85	26,88	0 - 102,78
Submucosa	49,45	21,42	49,91	32,36	64,28	31,93	0 - 108,86

Πίνακας 5 CPA Testing % για Testing Area > 0

Σύγκριση CPA μεταξύ μασκών αναφοράς και προβλεπόμενων μασκών για Training Area > 0

Κλάση	n	MAE	Pearson r
Mucosa	116	21,38	0,35
Muscularis	105	32,90	0,27
Submucosa	102	31,01	0,30

Πίνακας 6 CPA μεταξύ μασκών αναφοράς και προβλεπόμενων μασκών

Σύμφωνα με τα αποτελέσματα της ανάλυσης κολλαγόνου, παρατηρείται ότι η κατανομή των τιμών του δείκτη CPA διαφέρει μεταξύ των περιοχών ιστού, και αυτό εξηγεί την διαφορετική παρουσία κολλαγόνου σε κάθε περιοχή ιστού.

6.8. Συμπεράσματα

Σύμφωνα με την αξιολόγηση της τμηματοποίησης, παρατηρείται ότι το μοντέλο παρουσιάζει υψηλή απόδοση στην αναγνώριση βασικών ιστολογικών περιοχών, ειδικά για την κλάση mucosa. Επιπλέον, η υψηλή τιμή της Ειδικότητας (Specificity) για τις κλάσεις muscularis και submucosa, δείχνει ότι σπάνια τα pixels άλλων κλάσεων ταξινομούνται λανθασμένα ως μια από αυτές, που σημαίνει ότι τα σφάλματα εντοπίζονται κυρίως στις μεταβατικές περιοχές μεταξύ ιστών και όχι τόσο σε εκτεταμένες λανθασμένες ταξινομήσεις.

Ο υπολογισμός του δείκτη CPA (Collagen Proportionate Area) καθιστά δυνατή την μετατροπή της χρωματικής πληροφορίας της εικόνας σε ποσοτικό δείκτη, επιτρέποντας έτσι την συστηματική εκτίμηση της ύπαρξης κολλαγόνου εντός των ιστολογικών περιοχών. Παρόλα αυτά, ο δείκτης CPA είναι υπολογιστικό μέγεθος, που εξαρτάται από την **ακρίβεια της τμηματοποίησης** αλλά και από την **αποτελεσματικότητα διαχωρισμού των χρωμάτων**, μέσω του αλγορίθμου k – means. Για τον λόγο αυτό, τα σχετικά αποτελέσματα ερμηνεύονται ως ποσοτικός δείκτης της παρουσίας κόκκινου χρώματος και όχι ως αυτόνομο διαγνωστικό εργαλείο. Συγκεκριμένα, η ανάλυση των τιμών του δείκτη CPA έδειξε ότι υπάρχει **διαφοροποίηση μεταξύ διαφορετικών στρωμάτων ιστού**, με την ύπαρξή του σε υψηλότερα ποσοστά στις κλάσεις muscularis και submucosa, σε σχέση με την κλάση mucosa. Η διαφορά αυτή οφείλεται κυρίως στην μορφολογική ετερογένεια που παρουσιάζουν οι δομές ιστού και αναδεικνύει την παρουσία διαφορετικών βαθμών ίνωσης εντός αυτών.

Είναι επίσης σημαντικό να διαχωριστούν οι περιπτώσεις όπου, δεν εντοπίζεται ιστολογική περιοχή (Area = 0), από τις περιπτώσεις όπου υπάρχει η περιοχή αλλά δεν παρατηρείται ύπρηξη κολλαγόνου. Οι μηδενικές δηλαδή τιμές του δείκτη CPA ενδεχομένως να οφείλονται στην απουσία της συγκεκριμένης κλάσης από την πρόβλεψη ή στην απουσία κόκκινου χρώματος εντός της κλάσης. Ως εκ τούτου, και στις δυο αυτές περιπτώσεις το CPA θα έχει τιμή μηδέν, θα είναι όμως διαφορετική η ερμηνεία.

Σχετικά με τον έλεγχο αξιοπιστίας του CPA, επιτυγχάνεται μέσω της σύγκρισης των τιμών που υπολογίζονται στις μάσκες αναφοράς και στις προβλεπόμενες μάσκες, δείχνει υψηλό βαθμό συμφωνίας για τις περισσότερες κλάσεις (mucosa, submucosa). Ο υψηλός αυτός βαθμός συμφωνίας, δείχνει ότι η προτεινόμενη ροή επεξεργασίας (Τμηματοποίηση, Εφαρμογή k – means και υπολογισμός δείκτη CPA), δίνει αξιόπιστες

και συνεπείς μετρήσεις όταν υπάρχουν διαθέσιμες επισημάνσεις αναφοράς. Από την άλλη, η χαμηλότερη συμφωνία που παρατηρείται στην κλάση muscularis, οφείλεται στο γεγονός ότι υπάρχει μεγαλύτερη δυσκολία οριοθέτησης της συγκεκριμένης περιοχής. Ο υπολογισμός του δείκτη CPA μπορεί να επηρεαστεί σημαντικά, ακόμη και από μικρές αποκλίσεις στα όρια της περιοχής και αυτό εξηγεί την μεγαλύτερη διακύμανση στις μετρήσεις.

Συμπερασματικά, τα αποτελέσματα δείχνουν ότι η προτεινόμενη μεθοδολογία παρέχει έναν αξιόπιστο τρόπο ποσοτικής εκτίμησης της παρουσίας κόκκινου χρώματος ως ένδειξη ίνωσης, στις ιστολογικές περιοχές, προσφέροντας την δυνατότητα αντικειμενικής ανάλυσης ιστολογικών εικόνων.

6.9. Προτάσεις για μελλοντική έρευνα

Η προτεινόμενη μεθοδολογία μπορεί να αποτελέσει βάση για περαιτέρω ανάπτυξη προηγμένων συστημάτων ανάλυσης ιστοπαθολογικών εικόνων. Αρχικά, μια κατεύθυνση στην οποία θα μπορούσε να χρησιμοποιηθεί αφορά στη βελτίωση της τμηματοποίησης με την αξιοποίηση πιο σύγχρονων αρχιτεκτονικών βαθιάς μάθησης, όπως είναι μοντέλα βασισμένα σε Attention Mechanisms ή Transformer–Based προσεγγίσεις μιας και έχουν δείξει δυνατότητα κατανόησης πολύπλοκων χωρικών σχέσεων [107]. Επιπλέον, η αύξηση του μεγέθους του επισημασμένου συνόλου δεδομένων θα μπορούσε ενδεχομένως να επιφέρει καλύτερη γενίκευση του μοντέλου και βελτιωμένη απόδοση, ιδιαίτερα σε δύσκολες περιπτώσεις όπως είναι η διάκριση μεταξύ των κλάσεων *muscularis* και *submucosa*.

Μια επίσης σημαντική κατεύθυνση αφορά στην ανάπτυξη μεθόδων αξιολόγησης που θα λαμβάνουν υπόψη την αβεβαιότητα των επισημάνσεων, μιας και τα όρια μεταξύ των ιστολογικών περιοχών δεν καθορίζονται πάντα με σαφήνεια. Συγκεκριμένα, η χρήση Probabilistic Segmentation ή Soft Labels είναι μια πρακτική που θα μπορούσε να οδηγήσει σε ρεαλιστικότερη εκτίμηση της απόδοσης [108].

Όσον αφορά την ανάλυση του κολλαγόνου μια μελλοντική προσέγγιση θα μπορούσε να είναι η αντικατάσταση της χρήσης του k – means με μεθόδους βαθιάς μάθησης, οι οποίες δίνουν την δυνατότητα εκτίμησης του δείκτη CPA απευθείας από την εικόνα, χωρίς την ύπαρξη ενδιάμεσης ομαδοποίησης. Επίσης, η χρήση πολυκαναλικής πληροφορίας θα συνέβαλλε στην βελτίωση της ακρίβειας του χρωματικού διαχωρισμού, μειώνοντας έτσι τα σφάλματα στις περιοχές με χαμηλή χρωματική αντίθεση.

Τέλος, η χρήση της προτεινόμενης μεθοδολογίας σε συστήματα υποστήριξης κλινικών αποφάσεων, θα μπορούσε να λειτουργήσει θετικά στην αυτοματοποίηση της ποσοτικής αξιολόγησης ιστολογικών δειγμάτων συμβάλλοντας έτσι στην ελάττωση της υποκειμενικότητας της ανθρώπινης εκτίμησης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] A. C. Bateman and P. Patel, “Lower gastrointestinal endoscopy: guidance on indications for biopsy,” *Frontline Gastroenterol.*, vol. 5, no. 2, pp. 96–102, Apr. 2014, doi: 10.1136/flgastro-2013-100412.
- [2] V. Villanacci *et al.*, “Histopathology of non-IBD Colitis. A practical approach from the Italian group for the study of the gastrointestinal tract (GIPAD),” Feb. 01, 2021, *Pacini Editore S.p.A./AU-CNS*. doi: 10.32074/1591-951X-234.
- [3] J. D. Wang *et al.*, “The histologic features, molecular features, detection and management of serrated polyps: a review,” 2024, *Frontiers Media SA*. doi: 10.3389/fonc.2024.1356250.
- [4] World Health Organization, “Colorectal cancer_world Health Organization, 11 July 2023”, Accessed: Jun. 19, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer>
- [5] “Statistics at a glance.” Accessed: Jun. 19, 2024. [Online]. Available: <https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf>
- [6] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.
- [7] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global patterns and trends in colorectal cancer incidence and mortality,” *Gut*, vol. 66, no. 4, pp. 683–691, Apr. 2017, doi: 10.1136/gutjnl-2015-310912.
- [8] R. Labianca *et al.*, “Colon cancer,” May 2010. doi: 10.1016/j.critrevonc.2010.01.010.
- [9] CDC, “Colorectal Cancer - <https://www.cdc.gov/cancer/colorectal/>.”
- [10] A. Hadjipetrou, D. Anyfantakis, C. G. Galanakis, M. Kastanakis, and S. Kastanakis, “Colorectal cancer, screening and primary care: A mini literature review,” Sep. 07, 2017, *Baishideng Publishing Group Co*. doi: 10.3748/wjg.v23.i33.6049.
- [11] T. Sawicki, M. Ruszkowska, A. Danielewicz, E. Niedźwiedzka, T. Artukowicz, and K. E. Przybyłowicz, “A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis,” May 01, 2021, *MDPI AG*. doi: 10.3390/cancers13092025.
- [12] K. W. Jasperson, T. M. Tuohy, D. W. Neklason, and R. W. Burt, “Hereditary and Familial Colon Cancer,” *Gastroenterology*, vol. 138, no. 6, pp. 2044–2058, 2010, doi: 10.1053/j.gastro.2010.01.054.
- [13] K. Zhang, J. Civan, S. Mukherjee, F. Patel, and H. Yang, “Genetic variations in colorectal cancer risk and clinical outcome,” *World J. Gastroenterol.*, vol. 20, no. 15, pp. 4167–4177, Apr. 2014, doi: 10.3748/wjg.v20.i15.4167.

- [14] M. L. Bisgaard, K. Fenger, S. Bülow, E. Niebuhr, and J. Mohr, "Familial adenomatous polyposis (FAP): Frequency, penetrance, and mutation rate," *Hum. Mutat.*, vol. 3, no. 2, pp. 121–125, 1994, doi: 10.1002/humu.1380030206.
- [15] O. M. Sieber *et al.*, "Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations in MYH," 2003. [Online]. Available: www.nejm.org
- [16] F. Amersi, M. Agustin, and C. Y. Ko, "Colorectal Cancer: Epidemiology, Risk Factors, and Health Services."
- [17] N. Shussman and S. D. Wexner, "Colorectal polyps and polyposis syndromes," Feb. 01, 2014, *Oxford University Press*. doi: 10.1093/gastro/got041.
- [18] H. S. Gandomani *et al.*, "Colorectal cancer in the world: incidence, mortality and risk factors," *Biomedical Research and Therapy*, vol. 4, no. 10, p. 1656, Oct. 2017, doi: 10.15419/bmrat.v4i10.372.
- [19] Y. Ma *et al.*, "Type 2 diabetes and risk of colorectal cancer in two large U.S. prospective cohorts," *Br. J. Cancer*, vol. 119, no. 11, pp. 1436–1442, Nov. 2018, doi: 10.1038/s41416-018-0314-4.
- [20] C. Yao, G. F. Nash, and T. Hickish, "Management of colorectal cancer and diabetes," 2014, *SAGE Publications Ltd*. doi: 10.1177/0141076813512121.
- [21] "463 PEOPLE LIVING WITH DIABETES million."
- [22] P. Rawla, T. Sunkara, and A. Barsouk, "Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors," 2019, *Termedia Publishing House Ltd*. doi: 10.5114/pg.2018.81072.
- [23] Aicr and WCRF, "Diet, nutrition, physical activity and colorectal cancer". Accessed: Jun. 16, 2024. [Online]. Available: <https://wcrf.org/wp-content/uploads/2021/02/Colorectal-cancer-report.pdf>."
- [24] G. U. 2017 American Cancer Society: Atlanta, "Colorectal-cancer-facts-and-figures-2017-2019".
- [25] N. Murphy *et al.*, "Lifestyle and dietary environmental factors in colorectal cancer susceptibility," Oct. 01, 2019, *Elsevier Ltd*. doi: 10.1016/j.mam.2019.06.005.
- [26] N. N. Keum and E. Giovannucci, "Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies," Dec. 01, 2019, *Nature Research*. doi: 10.1038/s41575-019-0189-8.
- [27] L. Klampfer, "Vitamin D and Colon Cancer," *World J. Gastrointest. Oncol.*, vol. 6, no. 11, pp. 430–437, 2014, doi: 10.4251/wjgo.v6.i11.430.
- [28] F. Amersi, M. Agustin, and C. Y. Ko, "Colorectal Cancer: Epidemiology, Risk Factors, and Health Services." Accessed: Jun. 15, 2024. [Online]. Available: <https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-2005-916274>
- [29] C. Thélin and S. Sikka, "Epidemiology of Colorectal Cancer — Incidence, Lifetime Risk Factors Statistics and Temporal Trends," in *Screening for Colorectal Cancer with Colonoscopy*, InTech, 2015. doi: 10.5772/61945.

- [30] P. Rawla, T. Sunkara, and A. Barsouk, “Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors[CrossRef],” 2019, *Termedia Publishing House Ltd.* doi: 10.5114/pg.2018.81072.
- [31] C. Thélin and S. Sikka, “Epidemiology of Colorectal Cancer — Incidence, Lifetime Risk Factors Statistics and Temporal Trends,” in *Screening for Colorectal Cancer with Colonoscopy*, InTech, 2015. doi: 10.5772/61945.
- [32] S. Ahmed, K. Johnson, O. Ahmed, and N. Iqbal, “Advances in the management of colorectal cancer: From biology to treatment,” 2014, *Springer Verlag*. doi: 10.1007/s00384-014-1928-5.
- [33] T. Mizoue *et al.*, “Alcohol drinking and colorectal cancer in Japanese: a pooled analysis of results from five cohort studies.,” *Am. J. Epidemiol.*, vol. 167, no. 12, pp. 1397–1406, 2008, doi: 10.1093/aje/kwn073.
- [34] American Cancer Society., “Colorectal Cancer Facts & Figures 2017–2019; American Cancer Society: Atlanta, GA, USA, 2017.”, Accessed: Jun. 15, 2024. [Online]. Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf>
- [35] Andrew R Marley and Hongmei Nan, “Epidemiology of colorectal cancer PMID: 27766137 PMCID: PMC5069274”, Accessed: Jun. 15, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27766137/>
- [36] N. N. Keum and E. Giovannucci, “Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies,” Dec. 01, 2019, *Nature Research*. doi: 10.1038/s41575-019-0189-8.
- [37] S. Koc, M. N. Esin, and A. Ardic, “Colorectal Cancer Prevention and Risk Counseling, [CrossRef],” in *Colorectal Cancer - From Pathogenesis to Treatment*, InTech, 2016. doi: 10.5772/63206.
- [38] M. Ahmed, “Elmer Press [Internet], Gastroenterol Res. 2020;13(1):1-10, Colon Cancer: A Clinician’s Perspective in 2019 [PubMed],” *Gastroenterology Res.*, vol. 13, no. 1, pp. 1–10, 2020, doi: 10.14740/gr1239.
- [39] H. J. Freeman, “World J Gastroenterol. 2013 Dec 14; 19(46): 8468–8473 Early stage colon cancer. [NCBI free article] [PubMed].,” *World J. Gastroenterol.*, vol. 19, no. 46, pp. 8468–8473, Dec. 2013, doi: 10.3748/wjg.v19.i46.8468.
- [40] R. D. Rosen, ; Amit, and S. Affiliations, “StatPearls [Internet], StatPearls Publishing; Treasure Island (FL): Feb 13, 2023. TNM Classification. [NCBI free article] [PubMed],” 2024.
- [41] G. Menon, A. Recio-Boiles, S. Lotfollahzadeh, and B. Cagir, “Colon Cancer 2024 Apr 20. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 29262132.” , 2024.
- [42] J. Pawlowski, W. E. Jones, and I. Affiliations, “StatPearls [Internet]. StatPearls Publishing; Treasure Island (FL): Feb 28, 2023. Radiation Therapy for Anal Cancer.”

Accessed: Jun. 15, 2024. [Online]. Available:
<https://pubmed.ncbi.nlm.nih.gov/30726027/>

- [43] M. P. de Leon and A. Percesepe, "Pathogenesis of colorectal cancer," 2000. doi: 10.1016/s1590-8658(00)80361-8.
- [44] E. F. Fearon and B. Vogelstein, "A Genetic Model for Colorectal Tumorigenesis," 1990.
- [45] H. Brenner, M. Kloor, and C. P. Pox, "Lancet 2014; 383: 1490-1502 [PMID: 24225001 DOI: 10.1016/S0140-6736(13)61649-9] Colorectal cancer," in *The Lancet*, Elsevier B.V., 2014, pp. 1490–1502. doi: 10.1016/S0140-6736(13)61649-9.
- [46] Winawer SJ *et al.*, "Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993; 329: 1977-1981 [PMID: 8247072 DOI: 10.1056/NEJM199312303292701]".
- [47] F. Stracci, M. Zorzi, and G. Grazzini, "Colorectal cancer screening: Tests, strategies, and perspectives," Oct. 27, 2014, *Frontiers Media SA*. doi: 10.3389/fpubh.2014.00210.
- [48] C. A. Doubeni *et al.*, "Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: A large community-based study," *Gut*, vol. 67, no. 2, pp. 291–298, Feb. 2018, doi: 10.1136/gutjnl-2016-312712.
- [49] E. H. Schreuders *et al.*, "Colorectal cancer screening: A global overview of existing programmes," *Gut*, vol. 64, no. 10, pp. 1637–1649, Oct. 2015, doi: 10.1136/gutjnl-2014-309086.
- [50] Κ. Παπαγεωργίου and Γ Ι Στυλιανίδης, "Εισαγωγή Έκδοση Β' Νοσοκομειακά χρονικά, σελ: 198 – 199."
- [51] A. B. Ballinger and C. Anggiansah, "Ballinger, A. B., & Anggiansah, C. (2007). Colorectal cancer. In *British Medical Journal* (Vol. 335, Issue 7622, pp. 715–718). <https://doi.org/10.1136/bmj.39321.527384.BE>," Oct. 06, 2007. doi: 10.1136/bmj.39321.527384.BE.
- [52] Ι Στυλιανίδης–, "Χειρουργική Θεράπεία –Έκδοση Β' Νοσοκομειακά χρονικά, σελ: 209 – 226."
- [53] E. Orsenigo, G. Gasparini, and M. Carlucci, "Clinicopathological factors influencing lymph node yield in colorectal cancer: A retrospective study," *Gastroenterol. Res. Pract.*, vol. 2019, 2019, doi: 10.1155/2019/5197914.
- [54] C. B. Aarons, S. Shanmugan, and J. I. S. Bleier, "Management of malignant colon polyps: Current status and controversies," Nov. 21, 2014, *WJG Press*. doi: 10.3748/wjg.v20.i43.16178.
- [55] S. Nivatvongs *et al.*, "Risk of lymph node metastasis in T1 carcinoma of the colon and rectum. *Dis Colon Rectum*. 2002;45(2):200-206."
- [56] "Ν. Κ. Παπαγεωργίου, Γ Ι Στυλιανίδης - Εισαγωγή Έκδοση Β' Νοσοκομειακά χρονικά, σελ: 198 – 199."
- [57] Γ Ι Στυλιανίδης, "Χειρουργική Θεράπεία – Έκδοση Β' Νοσοκομειακά χρονικά σελ. 209-226."

- [58] “Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology”.
- [59] R. Costi, F. Leonardi, D. Zanoni, V. Violi, and L. Roncoroni, “Palliative care and end-stage colorectal cancer management: The surgeon meets the oncologist,” *World J. Gastroenterol.*, vol. 20, no. 24, pp. 7602–7621, 2014, doi: 10.3748/wjg.v20.i24.7602.
- [60] A. M. Borowiec *et al.*, “Colonic Stents for Colorectal Cancer Are Seldom Used and Mainly for Palliation of Obstruction: A Population-Based Study,” *Can. J. Gastroenterol. Hepatol.*, vol. 2016, 2016, doi: 10.1155/2016/1945172.
- [61] Y. Ohhara *et al.*, “Role of targeted therapy in metastatic colorectal cancer,” *World J. Gastrointest. Oncol.*, vol. 8, no. 9, p. 642, 2016, doi: 10.4251/wjgo.v8.i9.642.
- [62] D. Cunningham *et al.*, “Cetuximab Monotherapy and Cetuximab plus Irinotecan in Irinotecan-Refractory Metastatic Colorectal Cancer,” 2004. [Online]. Available: www.nejm.org
- [63] Κ. Γεωργούλη, Σ. Κάτσικας, and Ε. Μ. Πολυτεχνείο, *Τεχνητή Νοημοσύνη Συγγραφή Κριτικός αναγνώστης Συντελεστές έκδοσης Το παρόν έργο αδειοδοτείται υπό τους όρους της άδειας Creative Commons Αναφορά Δημιουργού-Μη Εμπορική Χρήση-Όχι Παράγωγα Έργα 3.0. ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ*. [Online]. Available: www.kallipos.gr
- [64] R. Anyoha, “Can Machines Think?” [Online]. Available: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- [65] Κ. Π. Β. Ν. Βλαχάβας Ιωάννης, Κόκκορας Φώτης, and Σακελλαρίου Ηλίας, “Τεχνητή νοημοσύνη”.
- [66] “Artificial Intelligence A Modern Approach Third Edition.”
- [67] Y. Mintz and R. Brodie, “Introduction to artificial intelligence in medicine,” *Minimally Invasive Therapy and Allied Technologies*, vol. 28, no. 2, pp. 73–81, Mar. 2019, doi: 10.1080/13645706.2019.1575882.
- [68] “FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients,” 2018. [Online]. Available: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-clinical-decision-support-software-alerting-providers-potential-stroke>
- [69] R. Shadmi, V. Mazo, O. Bregman-Amitai, and E. Elnekave, *FULLY-CONVOLUTIONAL DEEP-LEARNING BASED SYSTEM FOR CORONARY CALCIUM SCORE PREDICTION FROM NON-CONTRAST CHEST CT*. 2018.
- [70] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, “Artificial intelligence in colorectal cancer screening, diagnosis and treatment. A new era,” 2021, *MDPI*. doi: 10.3390/curroncol28030149.
- [71] H. P. Hu, Z. J. Niu, Y. P. Bai, and X. H. Tan, “Cancer classification based on gene expression using neural networks,” *Genetics and Molecular Research*, vol. 14, no. 4, pp. 17605–17611, Dec. 2015, doi: 10.4238/2015.December.21.33.
- [72] K. H. Chang *et al.*, “MicroRNA signature analysis in colorectal cancer: Identification of expression profiles in stage II tumors associated with aggressive disease,” *Int. J.*

Colorectal Dis., vol. 26, no. 11, pp. 1415–1422, Nov. 2011, doi: 10.1007/s00384-011-1279-4.

- [73] I. G. Tsoulos and A. Tzallas, “Training Artificial Neural Networks Using a Global Optimization Method That Utilizes Neural Networks,” *AI*, vol. 4, no. 3, pp. 491–508, Jul. 2023, doi: 10.3390/ai4030027.
- [74] F. Sohil, M. U. Sohali, and J. Shabbir, “An introduction to statistical learning with applications in R,” *Stat. Theory Relat. Fields*, vol. 6, no. 1, pp. 87–87, Jan. 2022, doi: 10.1080/24754269.2021.1980261.
- [75] Δ. ΕΔΙΠ Πανεπιστημίου Πατρών, “ΣΠΥΡΙΔΩΝ ΛΥΚΟΘΑΝΑΣΗΣ Καθηγητής Πανεπιστημίου Πατρών ΔΗΜΗΤΡΙΟΣ ΚΟΥΤΣΟΜΗΤΡΟΠΟΥΛΟΣ Υπολογιστική νοημοσύνη και βαθιά μάθηση.”
- [76] S. S. . Haykin, *Neural networks and learning machines*. Prentice Hall/Pearson, 2009.
- [77] J. Kufel *et al.*, “What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine,” Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics13152582.
- [78] M. Z. Alom *et al.*, “A state-of-the-art survey on deep learning theory and architectures,” Mar. 01, 2019, *MDPI AG*. doi: 10.3390/electronics8030292.
- [79] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [80] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, “Deep convolution neural network for image recognition,” *Ecol. Inform.*, vol. 48, pp. 257–268, Nov. 2018, doi: 10.1016/j.ecoinf.2018.10.002.
- [81] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” Aug. 01, 2018, *Springer Verlag*. doi: 10.1007/s13244-018-0639-9.
- [82] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Fine Tuning or Full Training?,” 2016.
- [83] J. G. Lee *et al.*, “Deep learning in medical imaging: General overview,” 2017, *Korean Radiological Society*. doi: 10.3348/kjr.2017.18.4.570.
- [84] S. K. Zhou *et al.*, “A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises,” *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, doi: 10.1109/JPROC.2021.3054390.
- [85] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, “Deep Convolutional Neural Networks in Medical Image Analysis: A Review,” Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info16030195.

- [86] F. Aeffner *et al.*, “Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association,” *J. Pathol. Inform.*, vol. 10, no. 1, Jan. 2019, doi: 10.4103/jpi.jpi_82_18.
- [87] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, “Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology,” *Nat. Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019, doi: 10.1038/s41571-019-0252-y.
- [88] A. Davri *et al.*, “Deep Learning on Histopathological Images for Colorectal Cancer Diagnosis: A Systematic Review,” Apr. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics12040837.
- [89] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.
- [90] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” Nov. 2015, [Online]. Available: <http://arxiv.org/abs/1505.00853>
- [91] M. A. Wakili *et al.*, “Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/8904768.
- [92] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.”
- [93] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast Cancer Histopathological Image Classification using Convolutional Neural Networks.” [Online]. Available: <http://web.inf.ufpr.br/vri/breast-cancer-database>.
- [94] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, “Transfer learning for medical image classification: a literature review,” Dec. 01, 2022, *BioMed Central Ltd*. doi: 10.1186/s12880-022-00793-7.
- [95] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, “MuDeRN: Multi-category classification of breast histopathological image using deep residual networks,” *Artif. Intell. Med.*, vol. 88, pp. 14–24, Jun. 2018, doi: 10.1016/j.artmed.2018.04.005.
- [96] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 2020, [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [97] M. Talo, “Automated classification of histopathology images using transfer learning,” *Artif. Intell. Med.*, vol. 101, Nov. 2019, doi: 10.1016/j.artmed.2019.101743.
- [98] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” Dec. 01, 2017, *Elsevier B.V.* doi: 10.1016/j.media.2017.07.005.
- [99] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015, [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [100] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.

- [101] E. office Of journal “Morphologia,” “Histology: A Text and Atlas: With Correlated Cell and Molecular Biology. Eighth Edition, 2018 Authors: Wojciech Pawlina; Michael H. Ross,” *Morphologia*, vol. 13, no. 4, pp. 76–89, Dec. 2019, doi: 10.26641/1997-9665.2019.4.76-89.
- [102] “rearcover”.
- [103] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN.”
- [104] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” Dec. 01, 2017, *Elsevier B.V.* doi: 10.1016/j.media.2017.07.005.
- [105] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection.”
- [106] J. Macqueen, “SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS.”
- [107] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2010.11929>
- [108] A. Hatamizadeh *et al.*, “UNETR: Transformers for 3D Medical Image Segmentation,” Oct. 2021, [Online]. Available: <http://arxiv.org/abs/2103.10504>