



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Κατάτμηση αιμοφόρων αγγείων
αμφιβληστροειδούς χρησιμοποιώντας τεχνικές
Βαθιάς Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΑΘΑΝΑΣΙΑΣ ΓΙΑΤΣΟΥ

Επιβλέπων: Ανδρέας Σταφυλοπάτης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Γεώργιος Σιόλας
Ε.ΔΙ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΤΗΣ ΚΑΙ ΜΑΘΗΣΗΣ
Αθήνα, Μάρτιος 2020



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Κατάτμηση αιμοφόρων αγγείων αμφιβληστροειδούς χρησιμοποιώντας τεχνικές Βαθιάς Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΑΘΑΝΑΣΙΑΣ ΓΙΑΤΣΟΥ

Επιβλέπων: Ανδρέας Σταφυλοπάτης
Αναπληρωτής Καθηγητής Ε.Μ.Π.
Συνεπιβλέπων: Γεώργιος Σιόλας
Ε.ΔΙ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3^η Μαρτίου 2020.

(Τιπογραφή)

(Τιπογραφή)

(Τιπογραφή)

Ανδρέας Σταφυλοπάτης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Ανδρέας Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2020

(Υπογραφή)

.....
ΓΙΑΤΣΟΥ ΑΘΑΝΑΣΙΑ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

©2020 – All rights reserved ΓΙΑΤΣΟΥ ΑΘΑΝΑΣΙΑ, 2022.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Περίληψη

Το θέμα της παρούσας διπλωματικής εργασίας είναι ...

Λέξεις Κλειδιά

Μηχανική Μάθηση, Βαθιά Μηχανική Μάθηση

Abstract

Subject of this diploma thesis is

Keywords

Machine Learning, Deep Learning

Ευχαριστίες

Ευχαριστώ τον καθηγητή Ανδρέα-Γεώργιο Σταφυλοπάτη και τα μέλη του εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης για την ευκαιρία που μου δόθηκε να εργαστώ στο συγκεκριμένο θέμα της διπλωματικής εργασίας μου.

Ευχαριστώ ιδιαίτερα τον κυρίο Γεώργιο Σιόλα, που μου προσέφερε την κάθε δυνατή βοήθεια και καθοδήγηση κατά την εκπόνηση της διπλωματικής μου εργασίας. Η στήριξη και η εμπιστοσύνη του σε εμένα μου ήταν πολύτιμες.

Τέλος, θέλω να ευχαριστήσω πολύ την οικογένειά μου και τους φίλους μου που είναι πάντα δίπλα μου και με υποστήριξαν καθ' όλη τη διάρκεια των σπουδών μου και της εκπόνησης της διπλωματικής μου εργασίας.

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	5
1 Εισαγωγή	17
1.1 Αντικείμενο	18
1.2 Παρεμφερείς Εργασίες	19
1.2.1 Μέθοδοι Χωρίς Επίβλεψη - <i>Unsupervised Methods</i>	19
1.2.2 Επιβλεπόμενες Μέθοδοι - <i>Supervised Methods</i>	20
1.2.3 Αρχιτεκτονικές Βαθιών Συνελικτικών Δικτύων	20
1.2.3.1 Πρώιμες διερευνήσεις βασισμένες σε Συνελικτικά Νευρωνικά Δίκτυα	20
1.2.3.2 Προσεγγίσεις βασισμένες σε Πλήρως Συνελικτικά Δίκτυα . . .	21
1.2.3.3 Προσεγγίσεις με αρχιτεκτονικές κωδικοποιητή/αποκωδικοποιητή	21
1.3 Στόχοι	22
1.4 Οργάνωση Εγγράφου	22
2 Θεωρητικό Τπόβαθρο	23
2.1 Διαγνωστική ανάλυση και απεικόνιση αμφιβληστροειδούς	23
2.1.1 Ανατομία και Φυσιολογία του Αμφιβληστροειδούς	23
2.1.2 Απεικόνιση βυθού	25
2.1.3 DRIVE Βάση Δεδομένων	26
2.1.4 Εκδήλωση και Απεικόνιση Παθήσεων Αμφιβληστροειδούς	27
2.2 Τεχνητή Νοημοσύνη	29
2.3 Μηχανική Μάθηση	29
2.3.1 Επιβλεπόμενη Μάθηση	29
2.3.2 Μη Επιβλεπόμενη Μάθηση	30
2.3.3 Βαθιά Μηχανική Μάθηση	30
2.4 Τεχνητά Νευρωνικά Δίκτυα	32
2.4.1 Ο αλγόριθμος Perceptron	33
2.4.2 Θεώρημα Καθολικής Προσέγγισης	33

2.4.3	Πολυεπίπεδα Νευρωνικά Δίκτυα	34
2.4.4	Αλγόριθμοι Εκπαίδευσης και Βασικές Συναρτήσεις	35
2.4.4.1	Συνάρτηση Κόστους	35
2.4.4.2	Αλγόριθμος Backpropagation	36
2.4.4.3	Αλγόριθμος Κατάβασης Κλίσης	37
2.4.4.4	Αλγόριθμοι Βελτιστοποίησης	39
2.4.4.5	Συνάρτηση Ενεργοποίησης	40
2.5	Συνελικτικά Νευρωνικά Δίκτυα (CNN)	45
2.5.1	Τρόπος λειτουργίας	46
2.5.2	Επίπεδα επεξεργασίας	47
2.5.2.1	Επίπεδο Εισόδου	47
2.5.2.2	Συνελικτικό Επίπεδο	47
2.5.2.3	Συγκεντρωτικό Επίπεδο	50
2.5.2.4	Πλήρως Συνδεδεμένο Επίπεδο	51
2.6	Πλήρως Συνελικτικά Δίκτυα - Fully Convolutional Networks	52
2.6.1	Μετατιθέμενη Συνέλξη - Transposed Convolution	52
2.6.2	Συνδέσεις Παράβλεψης - Skip Connections	53
2.7	Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNN)	53
2.7.1	Τρόπος λειτουργίας	54
2.8	Μετρικές Αξιολόγησης	54
3	Μεθοδολογία	59
3.1	Επεξεργασία Δεδομένων	59
3.1.1	Επεξεργασία Εικόνων (Image Preprocessing)	60
3.1.2	Μετατροπή Χρώματος Εικόνας	60
3.1.3	Διαχωρισμός σε patches	61
3.1.4	Κανονικοποίηση (Standardization)	62
3.1.5	Μετασχηματισμοί Προεπεξεργασίας των Εικόνων Εισόδου	62
3.1.5.1	Κανονικοποίηση Ολικής Αντίθεσης - Global Contrast Normalization (GCN)	62
3.1.5.2	Λεύκανση Μηδενικής Φάσης - Zero-phase Component Analysis (ZCA Whitening)	63
3.1.5.3	Εξισορρόπηση ιστογράμματος - Contrast Limited Adaptive Histogram Equalization (CLAHE)	64
3.1.5.4	Top-Hat Μετασχηματισμός Πολλαπλής Κλίμακας Multi-Scale Top-Hat Transform (MTHT)	64
3.1.6	Επαύξηση Δεδομένων (Data Augmentation)	65
3.2	Αρχιτεκτονικές Νευρωνικών Δικτύων για την Τυμπατοποίηση των Αγγείων του Αμφιβληστροειδούς	66
3.2.1	Προσέγγιση Ταξιμόμησης - Patch-wise Classification	66

3.2.2 Προσέγγιση Τμηματοποίησης - Patch Segmentation or Pixel-wise Classification	69
4 Προτεινόμενα Συστήματα	73
4.1 Συστήματα Ταξινόμησης Τεμαχίων Εικόνας	73
4.1.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία - Plain-Net	74
4.1.2 Προτεινόμενη αρχιτεκτονική και μεθοδολογία - No POOL-Net	77
4.2 Συστήματα Κατάτμησης Τεμαχίων Εικόνας	78
4.2.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία	78
4.3 Συνδυασμός Συστημάτων - Ensemble Learning	79
4.3.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία	80
5 Πειραματική Διαδικασία	83
5.1 Σύστημα Ταξινόμησης Τεμαχίων Εικόνας	84
5.1.1 Εκπαίδευση Δικτύου	85
5.1.1.1 Callbacks	85
5.1.1.2 Κλάση ImageDataGenerator	86
5.1.2 Προκλήσεις κατά την πειραματική διαδικασία και βελτιστοποίηση του δικτύου	87
5.1.2.1 Ανισοροπία Κλάσεων (Class Imbalance)	87
5.1.3 Αξιολόγηση Συστήματος	88
5.1.3.1 Συμπληρωματικές Μετρικές	88
5.1.3.2 Βέλτιστος Συντονισμός Κατωφλίου Απόφασης στο Στάδιο Ελέγχου	89
5.1.3.3 Αποτελέσματα Αξιολόγησης στα Δεδομένα Εκπαίδευσης	90
5.1.3.4 Αποτελέσματα Αξιολόγησης στα Δεδομένα Ελέγχου	92
5.2 Σύστημα Κατάτμησης Τεμαχίων Εικόνας	96
5.2.1 Εκπαίδευση Δικτύου	96
5.2.2 Προκλήσεις κατά την πειραματική διαδικασία και βελτιστοποίηση του δικτύου	98
5.2.2.1 Ανισοροπία Κλάσεων (Class Imbalance)	98
5.2.3 Αξιολόγηση Συστήματος	99
5.2.3.1 Αποτελέσματα Αξιολόγησης στα Δεδομένα Εκπαίδευσης	99
5.2.3.2 Αποτελέσματα Αξιολόγησης στα Δεδομένα Ελέγχου	101
5.2.3.3 Βελτιστοποίηση Αποτελεσμάτων Κατάτμησης μέσω Διαδικασίας Post Processing	102
5.3 Συνδυασμός Συστημάτων	104
5.4 Σύγκριση Μεθόδων με το State-of-the-Art	105
6 Συμπεράσματα και Μελλοντικές Επεκτάσεις	107
6.1 Συμπεράσματα	107
6.2 Μελλοντικές Επεκτάσεις και Προτάσεις	110

Βιβλιογραφία

111

Κατάλογος Σχημάτων

2.1	Ανατομία Οφθαλμού	23
2.2	Ανατομία Αμφιβληστροειδούς	24
2.3	Κάμερα απεικόνισης βυθού	25
2.4	Εικόνες αμφιβληστροειδούς από το DRIVE: (α) υγιής αμφιβληστροειδής, (β) αμφιβληστροειδής που παρουσιάζει παθολογίες.	26
2.5	Πρωτότυπη DRIVE εικόνα (α), η χειροκίνητη κατάτμηση (β) και η δυαδική μάσκα (γ).	27
2.6	Μορφολογία αμφιβληστροειδούς: (α) υγιής αμφιβληστροειδής, (β) αμφιβληστροειδής με διαβητική αμφιβληστροειδοπάθεια.	28
2.7	Η σχέση μεταξύ Artificial Intelligence, Machine Learning και Deep Learning	31
2.8	Βιολογικός Νευρώνας και Τεχνητός Νευρώνας	32
2.9	Η δομή ενός νευρώνα Perceptron	33
2.10	Τα επίπεδα ενός τεχνητού νευρωνικού δικτύου	34
2.11	Γραφική απεικόνιση του αλγόριθμου κατάβασης κλίσης	38
2.12	Σιγμοειδής συνάρτηση	41
2.13	Υπερβολική εφαπτομένη	42
2.14	Συνάρτηση ReLU	43
2.15	Συνάρτηση Leaky ReLU	43
2.16	Συνάρτηση Softmax	44
2.17	Συνελικτικό Νευρωνικό Δίκτυο	45
2.18	Ένα απλό Συνελικτικό Δίκτυο: με δεδομένο ένα patch εικόνας που παρέχει ένα πλαίσιο γύρω από ένα pixel προς ταξινόμηση (εδώ μπλε), εφαρμόζεται μια σειρά από συνελίξεις και λειτουργίες συγχέντρωσης (φίλτρα που ολισθαίνουν μέσω των επιπέδων εισόδου) (εδώ, πέντε συνελίξεις 4×4 , ακόλουθούμενες από μία συγχέντρωση 2×2 , ακόλουθούμενη από δύο 2×2 συνελίξεις). Κάθε επίπεδο εξόδου 1×1 ερμηνεύεται ως μια βαθμολογία για μια δεδομένη τάξη.	46
2.19	Παράδειγμα εισόδου και πυρήνα	48
2.20	Κατασκευή πίνακα χαρακτηριστικών	49
2.21	Στοιβαγμένοι πίνακες χαρακτηριστικών	49
2.22	Padding στα δεδομένα εισόδου	50
2.23	Εφαρμογή υποδειγματοληψίας	51
2.24	Πλήρως συνδεδεμένα επίπεδα	52

2.25	Παράδειγμα λειτουργίας Transposed Convolution	53
2.26	Δομή νευρώνα RNN	54
2.27	Κλάσεις προβλέψεων	55
2.28	Ερμηνεία της καμπύλης ROC-AUC	57
3.1	Αναπαράσταση Έγχρωμης Εικόνας ως Πίνακα	60
3.2	Παράδειγμα ενός θετικού συνόλου (α) και ενός αρνητικού συνόλου από patches (β)	61
3.3	Παράδειγμα συνόλου patches εκπαίδευσης, 27×27 όπως εξήχθησαν από τη βάση DRIVE	63
3.4	Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή GCN	63
3.5	Αρχικά patches εικόνας (α) και αντίστοιχα αποτελέσματα μετασχηματισμού αύξησης δεδομένων (β)	66
3.6	Οι διαφορές ανάμεσα στο patch-wise classification και στο pixel-wise classification.[36]	67
3.7	Παράδειγμα Αρχιτεκτονικής Συνελικτικού Νευρωνικού Δικτύου	67
3.8	Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή GCN	69
3.9	Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή ZCA	69
3.10	Παράδειγμα συνόλου patches εκπαίδευσης μετά την διαδικασία επαύξησης	69
3.11	Περιγραφή της encoder-decoder αρχιτεκτονικής U-net με την αναγνώριση των διαδρομών συστολής και επέκτασης	70
3.12	Περιγραφή της encoder-decoder αρχιτεκτονικής Sine-net	71
4.1	Δομή του μοντέλου Plain Lisk-Net όπως αυτό προτάθηκε	74
4.2	Οι παράμετροι του Batch Norm layer	75
4.3	Οι υπολογισμοί που πραγματοποιούνται από το Batch Norm layer	76
4.4	Απεικόνιση της διαδικασίας dropout. Στα αριστερά είναι ένα παράδειγμα της τυπικής διαδικασίας και στα δεξιά είναι το αποτέλεσμα του Dropout.	76
4.5	Δομή επιπέδων αρχιτεκτονικής Plain-Net με προσθήκη επιπέδων Batch Normalization	77
4.6	Δομή επιπέδων αρχικής αρχιτεκτονικής No-POOL-Net	78
4.7	Δομή του μοντέλου Sine-Net	79
4.8	Αναπαράσταση της γενικής διαδικασίας του Stacking Ensemble	80
4.9	Αναπαράσταση της αρχιτεκτονικής που χρησιμοποιήθηκε στην εργασία για το Stacking Ensemble	81
5.1	Κατανομή διαθέσιμων δεδομένων DRIVE	84
5.2	Accuracy (i) και Loss (ii) του Plain_Net για 300,000 training patches και batch size = 64	90
5.3	Accuracy (i) και Loss (ii) του Plain_Net με Batch_Normalization για 300,000 training patches και batch size = 128	90

5.4 ROC του <i>Plain_Net</i> με <i>Batch_Normalization</i> για 300,000 training patches και batch size = 128	91
5.5 Accuracy (i) και Loss (ii) του <i>No_Pool_Net</i> με <i>Batch_Normalization</i> για 300,000 training patches και batch size = 64	91
5.6 ROC του <i>No_Pool_Net</i> με <i>Batch_Normalization</i> για 300,000 training patches και batch size = 64	92
5.7 Ground truth segmentation testing εικόνας και αντίστοιχο αποτέλεσμα πρόβλεψης από το μοντέλο <i>Plain_Net_BN</i>	93
5.8 Classification Reports των εικόνων από τα δεδομένα ελέγχου του σχήματος 5.7	93
5.9 Ground truth segmentation testing εικόνας και αντίστοιχο αποτέλεσμα πρόβλεψης από το μοντέλο <i>No_Pool_Net</i>	94
5.10 Classification Reports των εικόνων από τα δεδομένα ελέγχου του σχήματος 5.9	94
5.11 Accuracy (i) και Loss (ii) του <i>Sine_Net_1</i> με Dice Binary CrossEntropy Loss	99
5.12 Accuracy (i) και Loss (ii) του <i>Sine_Net_2</i> με Balanced Sampler και Dice Binary CrossEntropy Loss	100
5.13 Accuracy (i) και Loss (ii) του <i>Sine_Net_3</i> με Weighted Dice Binary CrossEntropy Loss	100
5.14 ROC-Curve (i) PR-Curve (ii) του <i>Sine_Net</i>	101
5.15	101
5.16 Binary rediction output (i) και Line Detector output για τις εικόνες 1 (επάνω) και 2 (κάτω) του test set με επισημασμένες περιοχές με αισθητή βελτιστοποίηση.	104

Κατάλογος Πινάκων

2.1	Μετρικές απόδοσης για ανάλυση εικόνας αμφιβληστροειδούς	28
3.1	Οι δύο βασικές αρχιτεκτονικές που προτάθηκαν από τους Liskowski and Krawiec [31]	68
4.1	Αρχιτεκτονική των αξιολογούμενων CNN. Τα ονόματα των επιπέδων ακολουθούνται από τους αριθμούς των χαρτών χαρακτηριστικών. Οι τετράγωνες αγκύλες καθορίζουν το μέγεθος RF, τον διασκελισμό και την επένδυση	73
5.1	Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net	92
5.2	Οι τιμές που λήφθηκαν για μία σειρά μετρικών στο στάδιο αξιολόγησης του Plain-Net	95
5.3	Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net	95
5.4	Οι τιμές που λήφθηκαν για μία σειρά μετρικών στο στάδιο αξιολόγησης του Sine-Net	102
5.5	Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net και Sine-Net	102
5.6	Τιμές επίδοσης μοντέλων και σύγκριση με Στατε-οφ-τηε-Αρτ	105

Κεφάλαιο 1

Εισαγωγή

Η ιατρική απεικόνιση είναι ένας τρόπος για την παραγωγή γραφικών αναπαραστάσεων σε μορφή εικόνας διαφορετικών τμημάτων του ανθρώπινου σώματος για τη διάγνωση και την παρακολούθηση διαφόρων ασθενειών. Υπάρχουν διάφορες τεχνολογίες που χρησιμοποιούνται στην ιατρική απεικόνιση για την απόκτηση ποικίλων πληροφοριών σχετικά με διάφορα μέρη του σώματος. Κάθε τεχνολογία χρησιμοποιείται σε διαφορετικές περιπτώσεις όπως: Η ακτινογραφία χρησιμοποιείται συνήθως για την απεικόνιση των οστικών δομών και η απεικόνιση μαγνητικού συντονισμού χρησιμοποιείται συχνά για εικόνες υψηλής ανάλυσης του εγκεφάλου και του εσωτερικού ιστού. Η υποβοηθούμενη διάγνωση από υπολογιστή (CAD) ενσωματώνει την πολυδιάστατη ανάλυση ιατρικών εικόνων, η οποία βοηθά τους γιατρούς στη λήψη αποφάσεων. Αυτά τα συστήματα εκτελούν διάφορες διαδικασίες σε ψηφιακές ιατρικές εικόνες για να εκθέσουν τη συγκεκριμένη πάσχουσα περιοχή και να βοηθήσουν έτσι τους ακτινολόγους και τους ειδικούς γιατρούς στη λήψη αποφάσεων. Το CAD είναι ένας συνδυασμός διαφόρων κλάδων στην ψηφιακή επεξεργασία εικόνας, την τεχνητή νοημοσύνη και την όραση υπολογιστών. Έχουν αναπτυχθεί συστήματα CAD για τη διάγνωση και την παρακολούθηση απειλητικών για τη ζωή ασθενειών, όπως η αυτόματη ανίχνευση διαφόρων όγκων (εγκεφάλου, μαστού, πνεύμονα, κλπ.), καρδιαγγειακών παθήσεων, κανθώς και η αυτόματη ανάλυση και ανίχνευση παθήσεων που σχετίζονται με την όραση και τον αμφιβληστροειδή.^[1]

Ο αμφιβληστροειδής είναι ένας πολυεπίπεδος ιστός που καλύπτει το εσωτερικό του ματιού υπεύθυνο για το σχηματισμό εικόνων, δηλαδή για την αίσθηση της όρασης. Επιτρέποντας τη μετατροπή του φωτός σε νευρικό σήμα που θα υποβληθεί αργότερα σε επεξεργασία στον οπτικό φλοιό του εγκεφάλου, αυτή η ανατομική δομή είναι μία από τις σημαντικότερες για την ευημερία του ανθρώπου. Ως αποτέλεσμα της λειτουργίας του, ο ιστός του αμφιβληστροειδούς ταξινομείται ως εξαιρετικά μεταβολικά ενεργός, με διπλή παροχή αίματος που επιτρέπει την άμεση μη επεμβατική παρακολούθηση του κυκλοφορικού συστήματος του ατόμου.^[2]

Τα αιμοφόρα αγγεία του αμφιβληστροειδούς είναι μια κύρια ανατομική δομή που μπορεί να ανιχνευθεί στην εικόνα του βυθού του αμφιβληστροειδούς, με τη χρήση ειδικής κάμερας βυθού. Τα αγγεία του βυθού είναι οι πιο σταθερές και σημαντικές δομές με συμμετρικά και ασύμμετρα μοτίβα που είναι ανιχνεύσιμα.^[3]

Η ανάλυση εικόνας αμφιβληστροειδούς είναι ολοένα και πιο εξέχουσα ως μέθοδος μη επεμβατικής διάγνωσης στη σύγχρονη οφθαλμολογία. Η αγγείωση του αμφιβληστροειδούς και πιο συγκεκριμένα η μορφολογία του αιμοφόρου αγγείου του αμφιβληστροειδούς είναι ένας σημαντικός δομικός δείκτης για την αξιολόγηση της παφουσίας και της σοβαρότητας των παθήσεων του αμφιβληστροειδούς όπως η διαβητική αμφιβληστροειδοπάθεια, η υπέρταση, η αρτηριοσκλήρωση, το γλαύκωμα, οι αιμορραγίες, η απόφραξη των φλεβών και η νεοαγγείωση. Ωστόσο, προκειμένου να εκτιμηθούν οι μορφολογικές αυτές ιδιότητες όπως η διάμετρος, το μήκος, το μοτίβο διακλάδωσης, οι γωνίες και η στρεβλότητα του αιμοφόρου αγγείου του αμφιβληστροειδούς, η χειροκίνητη πλανημετρία έχει κοινώς χρησιμοποιηθεί από οφθαλμίατρους, η οποία είναι γενικά χρονοβόρα και επιρρεπής σε ανθρώπινο λάθιος, ειδικά όταν οι δομές των αγγείων είναι περίπλοκες ή ένας μεγάλος αριθμός εικόνων αποκτάται ώστε να επισημαίνονται με το χέρι. Ως εκ τούτου, είναι ζωτικής σημασίας στον τομέα της διάγνωσης με τη βοήθεια υπολογιστή η ανάπτυξη μιας αξιόπιστης αυτοματοποιημένης μεθόδου με υψηλή ακρίβεια για τον κατακερματισμό αιμοφόρων αγγείων αμφιβληστροειδούς, είναι ελκυστική στη διάγνωση με τη βοήθεια υπολογιστή.^[4, 5]

1.1 Αντικείμενο

Το κρίσιμο βήμα πριν την εξαγωγή αυτών των μορφολογικών χαρακτηριστικών των αγγείων του αμφιβληστροειδούς από τις εικόνες του βυθού του αμφιβληστροειδούς είναι η κατάτμηση των αγγείων (vessel segmentation), δηλαδή η σημασιολογική (semantic) αντιστοίχιση ετικετών στα αγγεία και το φόντο αναλόγως.

Ο στόχος της κατάτμησης των αγγείων του αμφιβληστροειδούς είναι να αναθέσει μία από τις δύο ετικέτες, δηλαδή το αγγείο (1) και το φόντο (0), σε κάθε εικονοστοιχείο (pixel) της εικόνας του αμφιβληστροειδούς, επομένως αναφέρεται ως ένα πρόβλημα δυαδικής ταξινόμησης. Σε αυτήν την εργασία, χρησιμοποιούνται μέθοδοι βαθιάς εκμάθησης (deep learning), για παράδειγμα βαθιά συνελικτικά νευρωνικά δίκτυα (deep convolutional neural networks - CNN), καθώς αυτά είναι ικανά να μάθουν αυτόματα μια ολοένα και πιο σύνθετη ιεραρχία χαρακτηριστικών απευθείας από τα δεδομένα εισόδου και να εμφανίσουν βελτίωση σε σχέση με τα συστήματα που βασίζονται σε χειροποίητα χαρακτηριστικά σε θέματα οπτικής αναγνώρισης.

Στην πρώτη προτεινόμενη μέθοδο κατάτμησης των αγγείων του αμφιβληστροειδούς της εργασίας αυτής, χρησιμοποιήθηκαν τρία κύρια στάδια πριν από την ταξινόμηση μέσω βαθιών CNNs: η προεπεξεργασία της εικόνας αμφιβληστροειδούς με μεθόδους κανονικοποίησης, η εξαγωγή χιλιάδων "μπαλωμάτων", υποεικόνων (patches) από κάθε εικόνα αμφιβληστροειδούς και η αύξηση δεδομένων με παραμορφώσεις προκειμένου να δημιουργηθούν περισσότερες εικόνες για να τροφοδοτηθεί το δίκτυο, καθώς τα χρησιμοποιούμενα σύνολα δεδομένων αμφιβληστροειδούς που είναι επισημασμένα χειροκίνητα είναι περιορισμένων δεδομένων. Η πρώτη αρχιτεκτονική που χρησιμοποιήθηκε, αποτελείται από συνεχόμενα συνελικτικά επίπεδα (convolutional layers) που ακολουθούνται ανά δύο από ένα επίπεδο συγκέντρωσης (pooling layer) που τα διαδέχονται στο τέλος δύο πλήρως συνδεδεμένα επίπεδα (fully connected layers). Η απόδοση του συστήματος μας αυτού αξιολογείται στο δημόσιο ανοικτό σύνολο δεδομένων DRIVE ^[6]. Οι μετρικές αξιολόγησης

που χρησιμοποιούμε εδώ είναι: Area Under the ROC Curve, Precision, Recall, Macro-Averaged Accuracy, Jaccard Index.

1.2 Παρεμφερείς Εργασίες

Οι υπάρχουσες προσεγγίσεις στο πρόβλημα της τυμηματοποίησης των αιμοφόρων αγγείων του αμφιβληστροειδούς μπορούν να χωριστούν χονδρικά σε δύο ομάδες: Μη εποπτευόμενες μέθοδοι (unsupervised) και εποπτευόμενες μέθοδοι (supervised), ανάλογα με το αν χρησιμοποιούνται ή όχι δεδομένα ελέγχου. Δεδομένα Ελέγχου - Ground Truth ονομάζεται το σύνολο των γνωστών και ορθά ταξινομημένων δειγμάτων, που έχουν επισημανθεί χειροκίνητα με ετικέτα από ειδικό και χρησιμοποιούνται ως δεδομένα εκπαίδευσης. Οι ακόλουθες υποενότητες παρουσιάζουν εν συντομία ορισμένα αντιπροσωπευτικές εργασίες από τις προαναφερθείσες δύο ομάδες.

1.2.1 Μέθοδοι Χωρίς Επίβλεψη - *Unsupervised Methods*

Στις μεθόδους χωρίς επίβλεψη δεν υπάρχει φάση εκπαίδευσης ενώ μπορούν να κατηγοριοποιηθούν περεταίρω σε πέντε βασικές κατηγορίες: αντιστοιχισμένο φιλτράρισμα (matched filtering), μορφολογική επεξεργασία, παρακολούθηση (tracking) αγγείων, ανάλυση πολλαπλών κλιμάκων και αλγόριθμοι που βασίζονται σε μοντέλα.

Η μεθοδολογία matched filtering πραγματοποιεί τη συνέλιξη ενός πυρήνα βασιζόμενου στη Γκαουσιανή ή στις παραγώγους της με την εικόνα αμφιβληστροειδή ώστε να βελτιώσει τα χαρακτηριστικά των αγγείων. Προκειμένου να σχεδιαστεί ο αντιστοιχισμένος πυρήνας φίλτρου, η μεθοδολογία αυτή εκμεταλλεύεται τη γραμμική προσέγγιση κατά τμήματα, τη μείωση της διαμέτρου των αγγείων και το προφίλ της έντασης που μοιάζει με Γκαουσιανή. Άλλη μία μέθοδος για την κατάτμηση των αγγείων είναι αυτή της μαθηματικής μορφολογίας που συνδιάζει την αξιολόγηση της καμπυλότητας και την μέθοδο matched-filtering για την ανίχνευση της κεντρικής φωτεινής γραμμής των αγγείων. Η μεθοδολογία, που βασίζεται στην παρακολούθηση των αγγείων τυμηματοποιεί ένα αγγείο χρησιμοποιώντας τοπικές πληροφορίες στο επίπεδο του ενός αγγείου παρά ολόκληρου του αγγειακού συστήματος. Οι προσεγγίσεις πολλαπλών κλιμάκων βασίζονται στην ανάλυση κλίμακας-χώρου. Εξετάζεται η πολυχλιμακωτή δεύτερης-τάξης τοπική δομή της εικόνας (Hessian) και λαμβάνεται μια μετρική αγγείων στη βάση της ανάλυσης ιδιοτιμής Hessian.[7]

Οι μέθοδοι χωρίς επίβλεψη σχεδιάζουν φίλτρα που είναι ευαίσθητα ως προς τα αγγεία και τα μοτίβα πο.υ μοιάζουν με αγγεία, και έτσι οδηγούν στο να μην ταυτοποιούνται πλήρως τα αιμοφόρα αγγεία ή στο να εντοπίζονται λανθασμένα ψευδο-μοτίβα που μοιάζουν με αγγεία.[8]. Γενικά, οι μη εποπτευόμενες μέθοδοι παρουσιάζουν ελαττώματα στην επίδοσή τους, επειδή δεν μπορούν να επωφεληθούν από τα δεδομένα ελέγχου επισημασμένα από άνθρωπο.

1.2.2 Επιβλεπόμενες Μέθοδοι - Supervised Methods

Οι επιβλεπόμενες μέθοδοι μάθησης που χρησιμοποιούν δεδομένα ground truth για να εκπαιδεύσουν τον ταξινομητή, αποτελούνται από δύο στάδια: (1) εξαγωγή χαρακτηριστικών, (2) ταξινόμηση. Κατά τη διάρκεια του σταδίου εξαγωγής χαρακτηριστικών, τα χαρακτηριστικά μπορούν να αποκτηθούν με δύο τρόπους: χειροποίητα χαρακτηριστικά και χαρακτηριστικά που βασίζονται στη μάθηση. Η πρώτη προσέγγιση βασίζεται στη χειροποίητα σχεδιασμένη εξαγωγή χαρακτηριστικών, όπου κάθε εικονοστοιχείο (pixel) προβάλλεται σε ένα σύνολο χαρακτηριστικών προκαθορισμένων από προηγούμενη γνώση, πριν από την ταξινόμηση. Τα χαρακτηριστικά των αγγείων μπορούν να εξαχθούν με το μετασχηματισμό Gabor, το διακριτό μετασχηματισμό κυματομορφής, το φιλτράρισμα αγγείων, το Γκαουσιανό φιλτράρισμα και άλλα.

Η δεύτερη προσέγγιση αναπτύσσει ένα τεχνητό νευρικό δίκτυο (Artificial Neural Network - ANN) ή τις παραλλαγές του για την κατάτμηση των αιμοφόρων αγγείων του αμφιβληστροειδούς, όπου η εξαγωγή και η ταξινόμηση των χαρακτηριστικών ενσωματώνονται σε ένα pipeline: η εξαγωγή χαρακτηριστικών μαθαίνεται άμεσα από τη μη-επεξεργασμένη αρχική εικόνα αντί να επιβάλλεται από τους εξωτερικούς σχεδιαστές και στη συνέχεια ακολουθεί η παραγωγή των προβλέψεων από τη διεργασία του τεχνητού νευρωνικού δικτύου.

1.2.3 Αρχιτεκτονικές Βαθιών Συνελικτικών Δικτύων

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) έχουν προσληφθεί αρκετό ενδιαφέρον, αφού μπορούν να μάθουν αυτόματα πολύπλοκες ιεραρχίες χαρακτηριστικών από τα ακατέργαστα δεδομένα εισόδου. Τα CNNs έχουν εφαρμοστεί ευρέως στην ταξινόμηση, αναγνώριση και κατάτμηση εικόνων. Η χρησιμότητά του αυτή ως προς τα δεδομένα εικόνων σχετίζεται με την ικανότητά τους να μαθαίνουν πολύπλοκες αναλλοίωτες όπως η κλίμακα και η αναλογίωτη της περιστροφής.^[7, 8]

Με τη χρήση ισχυρών τεχνικών βαθιάς εκμάθησης όπως τα Convolutional Neural Networks (CNNs), τα Fully Convolutional Network (FCNs) και πρόσφατα το U-Net, έχει επιτευχθεί εξαιρετική απόδοση στην επίλυση του προβλήματος της κατάτμησης των αιμοφόρων αγγείων αμφιβληστροειδούς. Ωστόσο, εξακολουθούν να υπάρχουν ορισμένοι παράγοντες που το καθιστούν ένα δύσκολο έργο. Αυτοί οι παράγοντες περιλαμβάνουν τη μεταβαλλόμενη αντίθεση και ένταση μεταξύ των διαφορετικών συνόλων εικόνων, ενδο-αγγειακές διαφορές μεταξύ παχιών και λεπτών αγγείων, η παρουσία οπτικού δίσκου και αλλοιώσεων, περιορισμένα σχολιασμένα δεδομένα και ούτω καθεξής. Θα συζητήσουμε πώς αντιμετωπίστηκαν αυτά τα προβλήματα στις ακόλουθες υποενότητες:

1.2.3.1 Πρώιμες διερευνήσεις βασισμένες σε Συνελικτικά Νευρωνικά Δίκτυα

Προτού χρησιμοποιηθούν ευρέως τα πλήρως συνελικτικά δίκτυα, η κατάτμηση αγγείων αντιμετωπίζοταν ως ένα πρόβλημα ταξινόμησης pixel-προς-pixel και η δομημένη πρόβλεψη (structured prediction) ήταν ακόμα ένα πρόβλημα που έπρεπε να επιλυθεί. Πρόκειται για ένα πρόβλημα συμπερασμάτων με πολλαπλές ετικέτες που λύνεται βρίσκοντας τις εξαρτήσεις των κλάσεων των

ετικετών από γειτονικά pixels: δεδομένου μιας έγχρωμης φωτογραφίας, ορίζεται pixel στην τοποθεσία (x,y) και γειτονιά αυτού ως $N(x,y)$. Σκοπός αποτελεί να ταξινομηθούν όλα τα γειτονικά pixels σε μια από τις δύο κλάσεις: (αγγείο, φόντο). Η συνήθης προσέγγιση ήταν η περικοπή των εικόνων σε patches ως είσοδος και η χρησιμοποίηση CNNs των οποίων τα τελευταία στρώματα είναι πλήρως συνδεδεμένα για να προβλέψουν την ετικέτα του κεντρικού εικονοστοιχείου κάθε patch. Η εργασία από τους Khalaf et al.(2016) είναι μία τυπική τέτοια προσέγγιση. Χρησιμοποίησαν ένα CNN που περιείχε τρία επίπεδα συνέλιξης για να πραγματοποιήσουν την προσέγγιση αγγείων. Το τελευταίο Fully Connected στρώμα του προτεινόμενου CNN περιέχει τρεις νευρώνες, που αντιπροσωπεύουν την πιθανότητα τα κεντρικά εικονοστοιχεία να είναι μεγάλα αγγεία, μικρά αγγεία ή φόντο, αντίστοιχα. Οι Yu et al.(2020) χρησιμοποίησαν επίσης ένα CNN του οποίου τα τελευταία στρώματα είναι στρώματα FC και στη συνέχεια αφού εξήγαγαν τα αγγειακά δέντρα από τα προηγουμένως την προσέγγιση αγγεία, χρησιμοποιώντας μια μέθοδο που βασίζεται σε γραφήματα, πρότειναν δύο αλγόριθμους για την ιεραρχική διαίρεση των αγγειακών δικτύων του αμφιβληστροειδούς (hierarchical division).

1.2.3.2 Προσεγγίσεις βασισμένες σε Πλήρως Συνελικτικά Δίκτυα

Τα πλήρως συνελικτικά δίκτυα παρέχουν μια από άκρη σε άκρη λύση, αντιμετωπίζοντας το ζήτημα της δομημένης πρόβλεψης. Ως εκ τούτου, εφαρμόστηκαν γρήγορα στην την προσέγγιση αγγείων. Οι Fu et al.(2016) πρότειναν ένα πλήρως συνελικτικό δίκτυο που ονομάζεται Deep-Vessel. Χρησιμοποίησαν ένα επίπεδο πλευρικής εξόδου για να βοηθήσουν το δίκτυο να μάθει χαρακτηριστικά πολλαπλής κλίμακας. Στο τέλος του δικτύου, χρησιμοποιήθηκε ένα στρώμα από Υπό Συνθήκη Τυχαία Πεδία Conditional Random Fields - CRF για την περαιτέρω μοντελοποίηση μη τοπικών συσχετίσεων εικονοστοιχείων. Οι Dasgupta et Singh (2017) πρότειναν επίσης ένα πλήρως συνελικτικό δίκτυο που περιέχει έξι επίπεδα συνέλιξης (convolutional), ένα επίπεδο υποδειγματοληψίας (downsampling) και ένα επίπεδο υπερδειγματοληψίας upsampling. Οι Feng et al. (2017) πρότειναν ένα πλήρως συνελικτικό δίκτυο το οποίο μπορεί να θεωρηθεί ως μια απλουστευμένη έκδοση του U-Net. Το δίκτυό τους πραγματοποιεί upsampling και downsampling μόνο δύο φορές. Προκειμένου να επιλύσουν το πρόβλημα της ανισορροπίας των τάξεων του φόντου και των αιμοφόρων αγγείων, όρισαν μια εντροπία που μετρά την αναλογία των εικονοστοιχείων αγγείων στο patch της εικόνας. Κατά τη διάρκεια της εκπαίδευσης, τα μισά patches επιλέγονται από τα patches με την υψηλότερη εντροπία και τα άλλα μισά επιλέγονται τυχαία. Οι Oliveira et al. (2018) πρότειναν μια αρχιτεκτονική FCN η οποία είναι παρόμοια με αυτή των Feng et al (2017). Στο στάδιο της προεπεξεργασίας χρησιμοποίησαν ένα μετασχηματισμό στάσιμων κυμάτων Stationary Wavelet Transform (SWT) ώστε να αποκτήσουν επιπλέον κανάλια για τις εικόνες εισόδου και να κρατήσουν την πληροφορία πολλαπλής ανάλυσης.

1.2.3.3 Προσεγγίσεις με αρχιτεκτονικές κωδικοποιητή/αποκωδικοποιητή

Χάρη στην εξαιρετική τους ικανότητα να εξάγουν χαρακτηριστικά και στην εξαιρετική τους απόδοση στην πράξη, οι αρχιτεκτονικές κωδικοποιητή - αποκωδικοποιητή (encoder-decoder), και

ειδικά η αρχιτεκτονική U-Net, παραμένουν τα πιο διάσημα frameworks που εφαρμόζονται στις εικόνες βυθού έως και σήμερα. Μία πολύ ενδιαφέρουσα κατεύθυνση που διερευνήθηκε πρόσφατα ήταν αυτή των Wu et al. (2019) που πρότειναν μια αρχιτεκτονική πολλαπλών κλιμάκων, το Vessel-Net, για να βελτιώσουν την μέχρι τότε φτωχή απόδοση στα τριχοειδή αγγεία των εικόνων. Πρώτα κατασκεύασαν ένα Inception-Residual (IR) block, εμπνευσμένο από την τεχνική Inception και ResNet που μπορούν να ενσωματωθούν στο U-Net. Τέσσερις ροές επίβλεψης παρουσιάστηκαν στο δίκτυο, συμπεριλαμβανομένων: μία παραδοσιακή ροή επίβλεψης, μία πλουσιότερων χαρακτηριστικών η οποία αλλάζει το μέγεθος όλων των σταδίων εξόδου του κωδικοποιητή ώστε να έχουν το ίδιο μέγεθος με τα patches εισόδου (48×48) και έπειτα τα συνδέει, και δύο ροές επίβλεψης πολλαπλών κλιμάκων όπου οι χάρτες χαρακτηριστικών που δημιουργούνται από τον κωδικοποιητή μεγέθους 12×12 και 24×24 προωθούνται διαμέσω ενός συνελικτικού επιπέδου με ReLu και Softmax συναρτήσεις ενεργοποίησης. [9]

1.3 Στόχοι

Σε αυτή την εργασία, θα παρουσιάσουμε μια προσπάθεια λύσης για την αυτόματη κατάτμηση των αιμοφόρων αγγείων του αμφιβληστροειδούς, χρησιμοποιώντας τεχνικές επεξεργασίας εικόνας και βαθιάς μηχανικής μάθησης. Τα αποτελέσματα της λύσης αυτής θα αναλυθούν, θα ερμηνευθούν, θα συγχριθούν και θα αξιολογηθούν και παρόμοιες διαυθέσιμες λύσεις.

1.4 Οργάνωση Εγγράφου

Αυτό το έγγραφο χωρίζεται σε έξι κύριες ενότητες. Η ενότητα που παρουσιάστηκε αποτελεί την εισαγωγή της παρούσας διπλωματικής εργασίας. Το υπόλοιπο του παρόντος εγγράφου είναι οργανωμένο ως εξής:

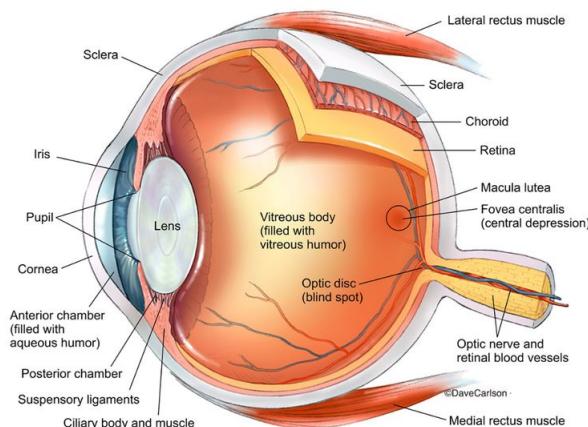
Το Κεφάλαιο 2 είναι μια περιγραφή ορισμένων σχετικών πληροφοριών θεωρητικού υποβάθρου σχετικά με την ανατομική δομή του αμφιβληστροειδούς, τη δομή των αγγείων, την τεχνική φωτογραφίας βυθού και την περιγραφή του συνόλου δεδομένων που χρησιμοποιούνται στην εργασία. Επιπλέον, αναλύεται το υπόβαθρο των μεθόδων και τεχνολογιών που χρησιμοποιήθηκαν στη συνέχεια. Στο κεφάλαιο 3 παρουσιάζεται η μεθοδολογία της εργασίας, οι προτεινόμενες μέθοδοι και τεχνικές και η αναλυτικές εφαρμογές τους. Στο κεφάλαιο 4 παρουσιάζεται η μεθοδολογία της εργασίας, η προτεινόμενη λύση και η αναλυτική εφαρμογή της. Τα αποτελέσματα που προέκυψαν κατά τη διάρκεια της εργασίας, η ανάλυση και η συζήτησή της παρουσιάζονται στο Κεφάλαιο 5. Τέλος, στο τελευταίο κεφάλαιο 6 συζητούνται τα συμπεράσματα και οι πιθανές μελλοντικές εργασίες σχετικά με τις προτεινόμενες λύσεις.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Διαγνωστική ανάλυση και απεικόνιση αμφιβληστροειδούς

2.1.1 Ανατομία και Φυσιολογία του Αμφιβληστροειδούς



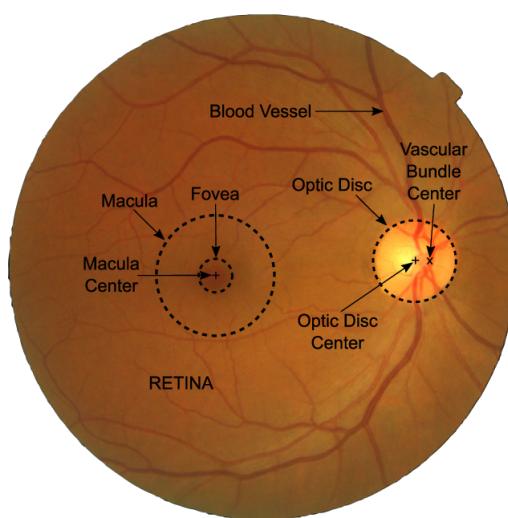
Σχήμα 2.1: Ανατομία Οφθαλμού

Η ανατομία του ματιού απεικονίζεται στην εικόνα 2.1. Η συνήθως λευκή μπάλα του ματιού ονομάζεται σκληρός χιτώνας, ο οποίος έχει ένα διαφανές μετωπιαίο τμήμα που ονομάζεται κερατοειδής. Κάτω από τον κερατοειδή, υπάρχει η ίριδα, η οποία προσαρμόζει την ποσότητα του φωτός που εισέρχεται στο μάτι, και ο φακός, ο οποίος εστιάζει την φως στο πίσω μέρος του ματιού. Το πίσω μέρος του ματιού είναι το σημείο όπου βρίσκονται τα φωτοευαίσθητα κύτταρα σε έναν στρωματοποιημένο ιστό που ονομάζεται αμφιβληστροειδής. Ο αμφιβληστροειδής είναι προσκολλημένος στο εσωτερικό στρώμα του οφθαλμού (τον χοριοειδή χιτώνα) με το επιθήλιο της χρωστικής του αμφιβληστροειδούς στη μέση. Ο χοριοειδής χιτώνας, μια λεπτή, πλούσια σε αίμα μεμβράνη βρίσκεται ανάμεσα στον αμφιβληστροειδή και τον σκληρό χιτώνα και είναι υπεύθυνη την παροχή αίματος στον αμφιβληστροειδή. Ο εσωτερικός χώρος του ματιού ονομάζεται

υαλοειδές σώμα και είναι γεμάτος με διαυγές τζελ που ονομάζεται υαλοειδές υγρό. [10]

Ο αμφιβληστροειδής είναι ο φωτοευαίσθητος χιτώνας του ματιού που συμπεριφέρεται σαν φιλμ σε κάμερα και είναι το πιο σημαντικό ανατομικό τμήμα του ματιού στο πλαίσιο αυτής της εργασίας. Μια σειρά από χημικά και ηλεκτρικά γεγονότα λαμβάνουν χώρα μέσα στον αμφιβληστροειδή πηροδοτούμενα από τα οπτικά στοιχεία του ματιού καθώς εστιάζουν σε ένα φωτεινό ερέθισμα, μια εικόνα. Τα ηλεκτρικά σήματα που προκύπτουν από αυτή την αλυσίδα γεγονότων αποστέλλονται στη συνέχεια στον εγκέφαλο μέσω νευρικών ινών, όπου και ερμηνεύονται ως οπτικές εικόνες. Ο ίδιος ο αμφιβληστροειδής είναι ένας πολυστρωματικός ιστός που αποτελείται από διαφορετικά χύτταρα για τη μετατροπή της φωτεινής ενέργειας, την προεπεξεργασία οπτικών πληροφοριών και τη μετάδοση του νευρικού σήματος. Αποτελείται από δέκα διαφορετικές στοιβάδες χυττάρων με πολύπλοκες συνδέσεις μεταξύ τους ενώ διαθέτει δύο μεγάλους πληρυματικούς φωτουποδοχέων, τα ραβδία (75-150 εκατομμύρια) και τα κωνία (7 εκατομμύρια). Τα πρώτα είναι υπεύθυνα για την νυχτερινή όραση ενώ τα δεύτερα για την ευχρινή όραση. [10]

Η δομή του αμφιβληστροειδούς, στους ανθρώπους, απεικονίζεται στην εικόνα 2.2 και εντοπίζεται στην εσωτερική επιφάνεια του οπίσθιου τμήματος των δύο τρίτων έως τα τρία τέταρτα του ματιού. Ένα λεπτό, διαφανές φύλλο ιστού που προέρχεται από το νευροεκτόνερμα, περιλαμβάνει τους αισθητήριους νευρώνες, που είναι η αρχή της οπτικής διαδρομής. Πολλαπλοί νευρώνες συνθέτουν τον νευρικό αμφιβληστροειδή (neuroretina) που χωρίζεται σε εννέα στρώματα και αποτελεί βασικό συστατικό στην παραγωγή και τη μετάδοση των ηλεκτρικών παλμών. Εποι, μεταδίδονται ριπές σημάτων μέσω γαγγλιακών χυττάρων που συνιστούν το οπτικό νεύρο στον ινιακό λοβό του εγκεφάλου όπου βρίσκονται τα ανώτερα οπτικά κέντρα μέσω των οποίων γίνονται αντιληπτές οι εικόνες. Η αντίληψη των εικόνων ως ανώτερη επεξεργασία από τον εγκέφαλο γίνεται με τέτοιον τρόπο ώστε οι εικόνες αντιπροσωπεύονται στον αμφιβληστροειδή ανεστραμμένες (κάτω ημισφαίριο αντιστοιχεί στο άνω οπτικό πεδίο, και ρινικό ημισφαίριο στο χροταφικό οπτικό πεδίο). [10]



Σχήμα 2.2: Ανατομία Αμφιβληστροειδούς

Στην ανατομία του αμφιβληστροειδούς το τμήμα που είναι υπεύθυνο για την έγχρωμη προβολή υψηλής ανάλυσης με εστίαση καρφίτσας είναι το βιοθρίο, όπου η πυκνότητα των κωνίων είναι η υψηλότερη. Στην υπόλοιπη επιφάνεια του αμφιβληστροειδούς, τα ραβδία υπερτερούν αριθμητικά των κωνίων. Οι οπικοί δίσκοι (optical disc - ΟΔ) είναι το μέρος του αμφιβληστροειδούς όπου οι νευρωνικές ίνες και τα αιμοφόρα αγγεία εισέρχονται στον αμφιβληστροειδή - δεν υπάρχουν φωτοδεκτικά κύτταρα στον ΟΔ και γι' αυτό είναι επίσης γνωστό ως το τυφλό σημείο. Όταν τα αιμοφόρα αγγεία εισέρχονται στον αμφιβληστροειδή μέσα στον ΟΔ, μια αρτηρία και μια φλέβα εισέρχονται, και στη συνέχεια, με διακλάδωση, γεμίζουν τον ιστό του αμφιβληστροειδούς. Από τεχνική άποψη, στον πραγματικό τρισδιάστατο χώρο κάθε αγγείο σχηματίζει μια δομή σαν δέντρο με μια ρίζα στον ΟΔ. Στις φωτογραφίες του αμφιβληστροειδούς, οι δισδιάστατες προβολές των δέντρων επικαλύπτονται, δημιουργώντας διασταυρώσεις αγγείων και κύκλους. Ωστόσο, μια σημαντική ιδιότητα είναι ότι ακόμη και στις δισδιάστατες προβολές, οι αρτηρίες δεν διασχίζουν τις φλέβες και οι φλέβες δεν διασταυρώνονται με τις φλέβες. [10]

Όσον αφορά στις μετρικές ποσοτικοποίησης του αμφιβληστροειδούς, επιτρέπουν και έχουν χρησιμοποιηθεί για να περιγράψουν τη σχέση μεταξύ των συστηματικών καρδιαγγειακών παθήσεων και των αλλαγών στον αμφιβληστροειδή. Οι συσχετίσεις μεταξύ των μετρήσεων και διαφόρων κλινικών παραμέτρων όπως η ηλικία, η αρτηριακή πίεση και ο δείκτης μάζας σώματος (ΔΜΣ) διερευνώνται σε μεγάλες πληθυσμιακές μελέτες [11], [12] που επιτρέπουν την εφαρμογή των μετρήσεων ως προγνωστικών παραγόντων σε διαγνωστικά συστήματα. Οι ποσοτικές μετρήσεις του αμφιβληστροειδούς βασίζονται όλες σε μετρήσεις της δομής των αιμοφόρων αγγείων στον αμφιβληστροειδή και περιλαμβάνουν συμβολικούς εκθέτες, γωνίες στις διακλαδώσεις, μέτρα αγγειακής στρέψης, μήκος: αναλογίες διαμέτρου, διαστάσεις φράκταλ και AVR. [11]

2.1.2 Απεικόνιση βυθού

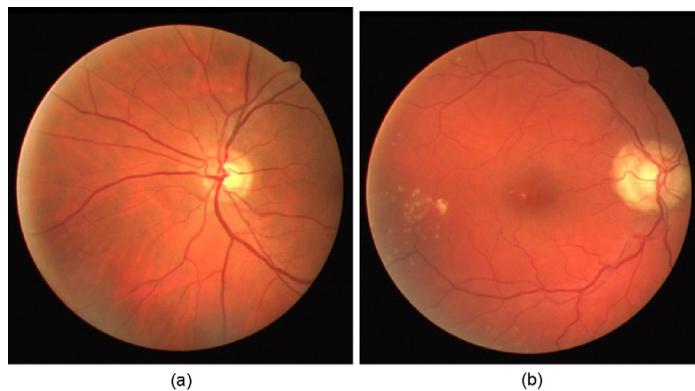


Σχήμα 2.3: Κάμερα απεικόνισης βυθού

Η φωτογραφία βυθού χρησιμοποιεί εξειδικευμένη κάμερα βυθού για τη λήψη της φωτογραφίας του αμφιβληστροειδούς. Η κάμερα του βυθού (Εικόνα 2.3) χρησιμοποιεί ειδικά φίλτρα και βαρές για να συλλάβει την κύρια δομή του ματιού, δηλαδή τον οπικό δίσκο, την ωχρά κηλίδα και τον κεντρικό και περιφερειακό αμφιβληστροειδή. Η φωτογραφία βυθού έχει διαχωριστεί σε διαφο-

ρετικούς τρόπους, π.χ. έγχρωμη, χωρίς κόκκινο, αγγειογραφία και ταυτόχρονη στερεοφωνική φωτογραφία βυθού. Ο αμφιβληστροειδής φωτίζεται χρησιμοποιώντας λευκό φως και εξετάζεται έγχρωμος στην έγχρωμη φωτογραφία. Σε φωτογραφίες βυθού χωρίς κόκκινο, χρησιμοποιείται ειδικό φίλτρο, δηλαδή πράσινο φίλτρο με μήκος κύματος 540-570 nm για την παρατήρηση της επιφανειακής βλάβης και την αφαίρεση του κόκκινου φωτός. Στη διαδικασία της αγγειογραφίας, μια φθορίζουσα χρωστική εγχέεται στην κυκλοφορία του αίματος για τη λήψη των φωτογραφών. Η αγγειογραφία φλουορεσκείνης νατρίου χρησιμοποιεί το μπλε φως 490 και φθορίζει 530 του κίτρινου φωτός. Αυτή είναι εξαιρετικά αποδεκτή φωτογραφία για τη διαβητική αμφιβληστροειδοπάθεια. Η αγγειογραφία πράσινης ινδοκυανίνης χρησιμοποιείται για την παρατήρηση του χοριοειδούς αγγείου και χρησιμοποίησε λέιζερ υπέρυυθρης διόδου 805 nm. [13]

2.1.3 DRIVE Βάση Δεδομένων

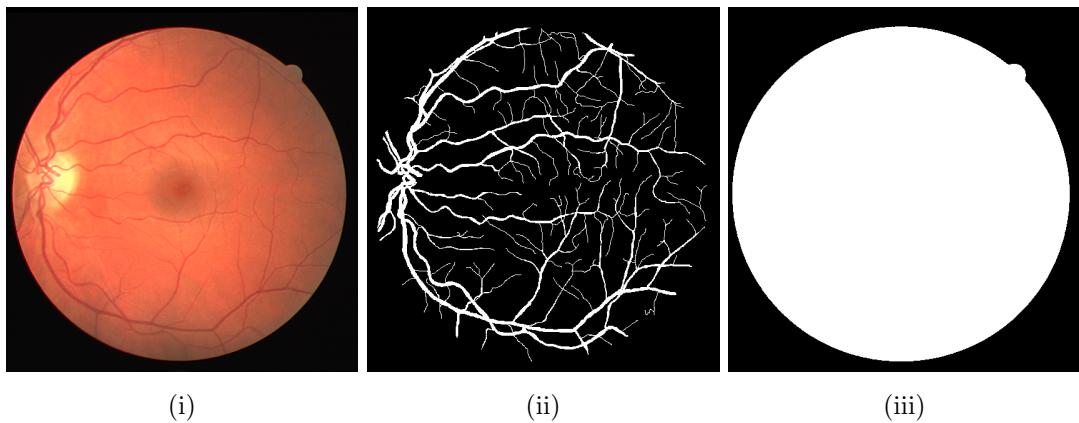


Σχήμα 2.4: Εικόνες αμφιβληστροειδούς από το DRIVE: (α) υγιής αμφιβληστροειδής, (β) αμφιβληστροειδής που παρουσιάζει παθολογίες.

To DRIVE (Digital Retinal Images for Vessel Extraction) [14] είναι μια δημόσια διαθέσιμη βάση δεδομένων, που αποτελείται από συνολικά 40 έγχρωμες φωτογραφίες βυθού. Οι φωτογραφίες ελήφθησαν από πρόγραμμα προσυμπτωματικού ελέγχου διαβητικής αμφιβληστροειδοπάθειας στην Ολλανδία. Ο πληθυσμός προσυμπτωματικού ελέγχου αποτελούνταν από 453 άτομα ηλικίας μεταξύ 31 και 86 ετών. Κάθε εικόνα έχει συμπιεστεί JPEG, κάτι που είναι κοινή πρακτική στα προγράμματα ελέγχου. Από τις 40 εικόνες της βάσης δεδομένων, οι 7 περιέχουν παθολογία, συγκεκριμένα εξιδρώματα, αιμορραγίες και αλλαγές στο χρωστικό επιθήλιο. Στην εικόνα 2.4 παρουσιάζεται ένα παράδειγμα τόσο μιας φυσιολογικής όσο και μιας παθολογικής εικόνας.

Οι εικόνες λήφθηκαν χρησιμοποιώντας μια άνον P5 μη μυδριατική κάμερα 3-CCD με οπτικό πεδίο 45 ° (FOV). Κάθε εικόνα λαμβάνεται χρησιμοποιώντας 8 βίτ ανά επίπεδο χρώματος στα 768×584 πιξελ. Το FOV κάθε εικόνας είναι κυκλικό με διάμετρο περίπου 540 πιξελ. Για αυτήν τη βάση δεδομένων, οι εικόνες έχουν περικοπεί γύρω από το FOV. Για κάθε εικόνα, παρέχεται μια εικόνα μάσκας που οριοθετεί το FOV (Εικόνα 2.5(iii)).

Το σύνολο των 40 εικόνων χωρίστηκε σε ένα σετ δοκιμής και σε σετ εκπαίδευσης που περιείχαν και τα δύο 20 εικόνες. Η εικόνα 2.5(i) παρουσιάζει ένα παράδειγμα εικόνας εκπαίδευσης.



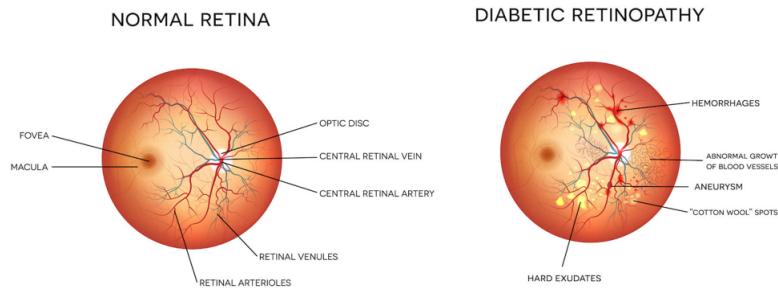
Σχήμα 2.5: Πρωτότυπη DRIVE εικόνα (α), η χειροκίνητη κατάτμηση (β) και η δυαδική μάσκα (γ).

Μια χειροκίνητη τμηματοποίηση του αγγειακού συστήματος είναι διαθέσιμη για καινεμία από τις εικόνες εκπαίδευσης και για το σύνολο δεδομένων δοκιμής είναι διαθέσιμες δύο τμηματοποιήσεις. Το ένα χρησιμοποιείται ως χρυσός κανόνας (Εικόνα 2.5(ii)), το άλλο μπορεί να χρησιμοποιηθεί για τη σύγχριση τμηματοποιήσεων που παράγονται από υπολογιστή με εκείνες από έναν ανεξάρτητο ανθρώπινο παρατηρητή. Τρεις παρατηρητές τμηματοποίησαν χειροκίνητα έναν αριθμό εικόνων και όλοι εκπαίδευτηκαν από έμπειρο οφθαλμίατρο. Ο πρώτος παρατηρητής τμηματοποίησε 14 εικόνες του σετ εκπαίδευσης ενώ ο δεύτερος παρατηρητής τις άλλες 6 εικόνες. Το σύνολο δοκιμής τμηματοποιήθηκε δύο φορές με αποτέλεσμα ένα σύνολο X και ένα Y. Το σύνολο X τμηματοποιήθηκε τόσο από τον πρώτο όσο και από τον δεύτερο παρατηρητή (13 και 7 εικόνες, αντίστοιχα) ενώ το σύνολο Y κατατμήθηκε πλήρως από τον τρίτο παρατηρητή. Η απόδοση των αλγορίθμων τμηματοποίησης αγγείων μετράται στο δοκιμαστικό σύνολο. Στο σετ X οι παρατηρητές σημειώσαν 577.649 pixels ως αγγεία και 3.960.494 ως φόντο (12,7% αγγεία). Στο σύνολο Y 556.532 pixels σημειώνονται ως αγγεία και 3.981.611 ως φόντο (12,3% αγγεία).

2.1.4 Εκδήλωση και Απεικόνιση Παυθήσεων Αμφιβληστροειδούς

Η ανάλυση της μορφολογίας του αμφιβληστροειδούς παρέχει μια εικόνα για τη γενική κατάσταση της υγείας ολόκληρου του ανθρώπινου σώματος, καθώς πολλές συστηματικές ασθένειες και ασθένειες του αμφιβληστροειδούς εκδηλώνονται ως βλάβες των οφθαλμικών δομών.

Η διαβητική αμφιβληστροειδοπάθεια (ΔA) και ο σαχχαρώδης διαβήτης ($\Sigma \Delta$) είναι μια ασθένεια που προκαλείται από κυμαινόμενα επίπεδα γλυκαυγίας (υπογλυκαιμία, υπεργλυκαιμία). Η πιο διαδεδομένη συνέπεια του σαχχαρώδους διαβήτη είναι η διαβητική αμφιβληστροειδοπάθεια. Παρά τη χρήση εντατικής θεραπείας γλυκαυγικού ελέγχου, η ΔA παραμένει μια επιπλοκή του διαβήτη που απειλεί την όραση. Τα σοβαρά συμπτώματα της ΔA περιλαμβάνουν ανάπτυξη μικρο-ανευρυσμάτων, νεοαγγείωση, αιμορραγίες, κηλίδες από βαμβάκι και εξιδρώματα στην περιοχή του αμφιβληστροειδούς. Παυθογενέσεις όπως το κλείσιμο των τριχοειδών και η δυσλειτουργία του φραγμού αίματος-αμφιβληστροειδούς είναι οι κύριες ανατομικές αλλαγές στον αμφιβληστροειδή



Σχήμα 2.6: Μορφολογία αμφιβληστροειδούς: (α) υγιής αμφιβληστροειδής, (β) αμφιβληστροειδής με διαβητική αμφιβληστροειδοπάθεια.

Performance metrics for retinal image analysis.

Performance metric	Description
Sensitivity (SN)	Ratio of classified true positives to the actual true positives in the ground truth. Also known as true positive rate (TPR)
Specificity (SP)	Ratio of classified true negatives to the actual true negatives in the ground truth. (1-SP) is also known as false positive rate (FPR)
Receiver Operating Characteristic curve (ROC)	Mapping of TPR to FPR at different threshold levels or mapping SN to SP in a 2 dimensional plane
Area Under the Curve (AUC)	Area covered by ROC curve when optimality is achieved

Πίνακας 2.1: Μετρικές απόδοσης για ανάλυση εικόνας αμφιβληστροειδούς

που οδηγούν σε ΔΑ.

Πιο αναλυτικά, πολλές διαταραχές του αμφιβληστροειδούς προέρχονται από την εξάλειψη των τριχοειδών αγγείων του αμφιβληστροειδούς, όπως η νεοαγγείωση αναφέρεται ότι προηγείται η μη αιμάτωση και το κλείσιμο των μικρών αιμοφόρων αγγείων του αμφιβληστροειδούς. Το κλείσιμο των αγγείων του αμφιβληστροειδούς υποκινεί την ανάπτυξη νεοαγγείωσης, μικροανευρυσμάτων, κηλίδων από βαμβάκι και αιμορραγιών. Τέτοιες αλλοιώσεις του αμφιβληστροειδούς φαίνονται και στην εικόνα 2.6. [15]

Ένας αλγόριθμος θεωρείται αποτελεσματικός εάν οι προβλέψεις του και τα δεδομένα ελέγχου ground truth βρίσκονται σε άμεση εγγύτητα μεταξύ τους. Αυτή η εγγύτητα μπορεί να ελεγχθεί μόνο με τη χρήση ορισμένων ποσοτικών μετρήσεων. Αυτές οι ποσοτικές μετρήσεις βοηθούν επίσης στην αξιολόγηση και σύγκριση των δυνατοτήτων διαφορετικών αλγορίθμων και προέρχονται από βασικά μέτρα απόδοσης (αληθώς θετικά, ψευδώς θετικά, αληθώς αρνητικά και ψευδώς αρνητικά). Οι μετρήσεις απόδοσης που χρησιμοποιούνται για ανάλυση και σύγκριση αλγορίθμων σε αυτή τη μελέτη καταγράφονται στον Πίνακα 2.1. Για παράδειγμα, η αναλογία των ταξινομημένων pixels εξιδρώματος προς τα συνολικά pixels συν την αναλογία των pixels που ταξινομούνται ως μη εξιδρώματα προς τα συνολικά pixels. [15]

2.2 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη (Artificial Intelligence-AI) είναι ένας τομέας της πληροφορικής που ασχολείται με την σχεδίαση και την ανάπτυξη συστημάτων που μιμούνται την ανθρώπινη συμπεριφορά. Όπως υποδηλώνει και το όνομα της, τα συστήματα τεχνητής νοημοσύνης προσπαθούν να μοντελοποιήσουν και να προσομοιώσουν νοήμονες συμπεριφορές που χαρακτηρίζουν τον άνθρωπο. Επίσης, η ανάπτυξη τέτοιων συστημάτων προϋποθέτει την μελέτη και άλλων επιστημών, όπως για παράδειγμα της ψυχολογίας και της ιατρικής, συνθέτοντας έτσι ένα σημείο τομής μεταξύ πληροφορικής και των επιστημών αυτών. Η τεχνητή νοημοσύνη επεκτείνεται σε πολλές κατηγορίες όπως η επεξεργασία φυσικής γλώσσας (NTLK), η ρομποτική (Robotics), η οραση υπολογιστών (Computer Vision), η μηχανική μάθηση (Machine Learning) κ.α.

2.3 Μηχανική Μάθηση

Η μηχανική μάθηση (Machine Learning-ML) αποτελεί παρακλάδι της τεχνητής νοημοσύνης με σκοπό την ανάπτυξη συστημάτων που μπορούν να μαθαίνουν. Το όνομα μηχανική μάθηση αποδόθηκε από τον Arthur Samuel το 1959 [16], πολύ αργότερα το 1997 ο Tom M. Mitchell θα δώσει έναν πιο επίσημο και συχνά αναφερόμενο ορισμό της μηχανικής μάθησης. Σύμφωνα με τον προηγούμενο, ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από μια εμπειρία Ε ως προς μια κλάση εργασιών Τ και ένα μέτρο επίδοσης Π, αν η επίδοσή του στις εργασίες της κλάσης Τ, όπως αποτιμάται από το μέτρο Ρ, βελτιώνεται με την εμπειρία Ε [17, p. 2]. Ουσιαστικά, οι αλγόριθμοι μηχανικής μάθησης χτίζουν ένα μαθηματικό μοντέλο με την χρήση κάποιων δεδομένων, γνωστά και ως δεδομένα εκπαίδευσης (training data), ώστε να μπορέσουν να κάνουν προβλέψεις και να πάρουν αποφάσεις χωρίς να έχουν προγραμματιστεί ρητά για το συγκεκριμένο έργο. Οι αλγόριθμοι μηχανικής μάθησης ταξινομούνται σε τρεις κύριες κατηγορίες, βάσει του τρόπου εκμάθησης, την επιβλεπόμενη μάθηση (Supervised learning), τη μη επιβλεπόμενη μάθηση (Unsupervised learning) και την ενισχυμένη μάθηση (Reinforcement learning).

2.3.1 Επιβλεπόμενη Μάθηση

Στην επιβλεπόμενη μάθηση, ή αλλιώς επιτηρούμενη μάθηση, το σύνολο των δεδομένων (dataset) που χρησιμοποιεί ο εκάστοτε αλγόριθμος αποτελείται από ζευγάρια εισόδου (input) και της επιθυμητής εξόδου (output). Συχνά τα δεδομένα εξόδου αναφέρονται και ως ετικέτες (labels). Είναι σύνηθες, στη διαδικασία της επιβλεπόμενης μάθησης το σύνολο των δεδομένων να χωρίζεται σε δύο υποσύνολα, τα δεδομένα εκπαίδευσης (training data) και τα δεδομένα δοκιμής (test data). Με τα δεδομένα εκπαίδευσης επιτυγχάνεται η εκπαίδευση του αλγορίθμου ενώ με τα δεδομένα δοκιμής αξιολογείται η επιτυχία του. Σχοπός ενός αλγορίθμου επιβλεπόμενης μάθησης είναι η προσέγγιση μιας συνάρτησης, με τη βοήθεια της εκπαίδευσης, που θα αντιστοιχεί μία ετικέτα σε κάθε δεδομένο εισόδου. Οι ετικέτες των δεδομένων εισόδου είναι αυτές που θα καθοδηγήσουν τον αλγόριθμο να γενικεύσει την αντιστοίχηση εισόδου-ετικέτας, να μάθει να αποδίδει δηλαδή τις σωστές ετικέτες σε οποιαδήποτε είσοδο. Οι ετικέτες μπορούν να πάρουν

είτε ποιοτικές ή ποσοτικές τιμές. Για τον λόγο αυτό, τα προβλήματα που επιλύει η επιβλεπόμενη μάθηση χωρίζονται σε δύο κατηγόριες, τα προβλήματα κατηγοριοποίησης (Classification) και τα προβλήματα παλινδρόμησης (Regression).

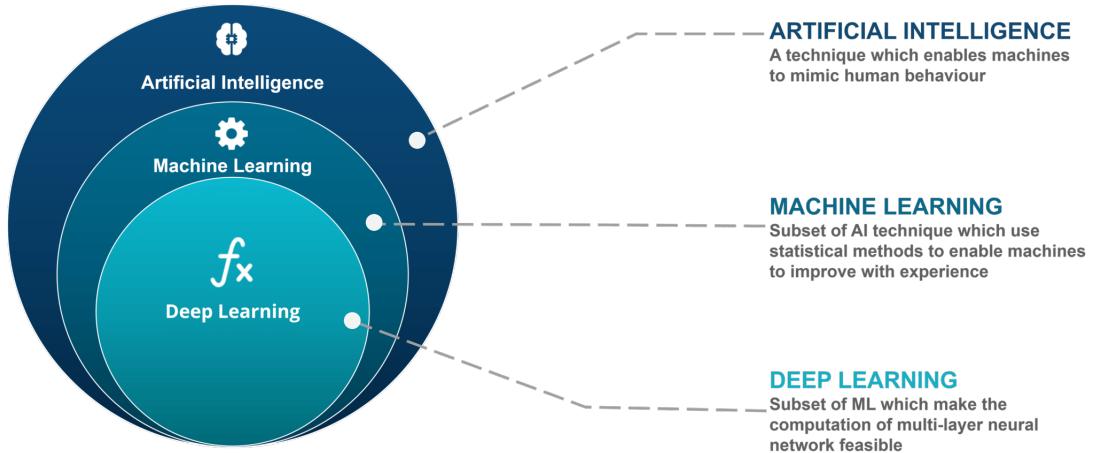
- **Κατηγοριοποίηση (Classification)** Στα προβλήματα κατηγοριοποίησης, οι τιμές των ετικετών ανήκουν σε ένα περιορισμένο σύνολο, οι ετικέτες δηλαδή εκφράζουν ένα ποιοτικό χαρακτηριστικό. Ο αλγόριθμος έχει στόχο να ταξινομήσει τα δεδομένα της εισόδου στην κατηγορία που ανήκουν. Για παράδειγμα, προβλήματα ταξινόμησης μπορεί να είναι η απόφαση αν σε μια εικόνα υπάρχει μια γάτα ή ο διαχωρισμός τραγουδιών σε μουσικά είδη.
- **Παλινδρόμηση (Regression)** Στα προβλήματα παλινδρόμησης, οι τιμές των ετικετών παίρνουν συνεχείς τιμές, οι ετικέτες δηλαδή εκφράζουν ένα ποσοτικό χαρακτηριστικό. Ο αλγόριθμος σε αυτή την περίπτωση έχει ως στόχο να προβλέψει μια τιμή για τα δεδομένα της εισόδου. Προβλήματα παλινδρόμησης θα μπορούσαν να είναι η πρόβλεψη της τιμής μιας μετοχής ή η πρόβλεψη της θερμοκρασίας της ακόλουθης μέρας.

2.3.2 Μη Επιβλεπόμενη Μάθηση

Στη μη επιβλεπόμενη μάθηση, το σύνολο των δεδομένων δεν είναι επισημασμένα, δεν έχουν δηλαδή ετικέτες, σε αντίθεση με την επιβλεπόμενη μάθηση. Σκοπός της μη επιβλεπόμενης μάθησης είναι να αναγνωρίσει κοινά χαρακτηριστικά των δεδομένων εισόδου και να τα ομαδοποιήσει σε κατηγορίες. Αυτή η τεχνική ομαδοποίησης ονομάζεται συσταδοποίηση (Clustering) καθώς επικεντρώνεται στην ταξινόμηση των δεδομένων σε συστάδες. Η ανάλυση της συσταδοποίησης τοποθετεί στην ίδια ομάδα παρατηρήσεις με κοινά χαρακτηριστικά, σύμφωνα με κάποια προκαθορισμένα χριτήρια, ενώ παρατηρήσεις με ανόμοια χαρακτηριστικά τις απομακρύνει. Ανάλογα με τη φύση του προβλήματος, ο αριθμός των συστάδων, που προσπαθούν οι αλγόριθμοι μη επιβλεπόμενης μάθησης να ταξινομήσουν τα δεδομένα, δεν είναι πάντα γνωστός εκ των προτέρων. Έστω ότι έχουμε εικόνες από ζώα και θέλουμε να τα κατηγοριοποιήσουμε σε ομάδες ώστε κάθε ομάδα να περιέχει εικόνες ίδιων ζώων. Είναι φανερό ότι ο αριθμός των ομάδων σε αυτό το πρόβλημα είναι αρχικά άγνωστος. Αντίθετα, στο πρόβλημα διαχωρισμού αλληλογραφίας σε κακόβουλη ή όχι ο αριθμός των συστάδων είναι γνωστός εκ των προτέρων.

2.3.3 Βαθιά Μηχανική Μάθηση

Η Βαθιά Μηχανική Μάθηση είναι ένα παρακλάδι της Μηχανικής Μάθησης που τα τελευταία χρόνια παρουσιάζει ραγδαία ανάπτυξη (2.7). Η βαθιά μάθηση χρησιμοποιεί δίκτυα πολλαπλών επιπέδων, δηλαδή αυξημένου βάθους, για την εξαγωγή χαρακτηριστικών υψηλού επιπέδου από τα δεδομένα. Η χρήση πολλαπλών επιπέδων επέτρεψε την αποτελεσματική εκπαίδευση δικτύων σε μη-δομημένα δεδομένα όπως οι εικόνες, η φωνή, κ.α. Οι μέθοδοι βαθιάς μηχανικής μάθησης έδωσαν μεγάλη ώθηση στην όραση υπολογιστών (computer vision), την επεξεργασία φυσικής γλώσσας (natural language processing), αυτόματη αναγνώριση ομιλίας και ήχου και σε πολλές άλλες εφαρμογές.



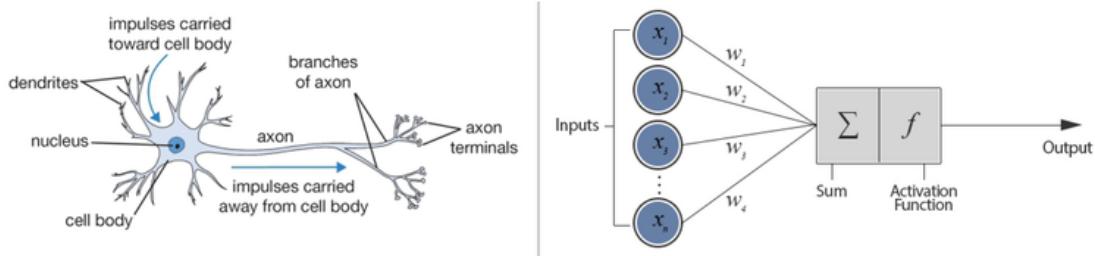
Σχήμα 2.7: Η σχέση μεταξύ Artificial Intelligence, Machine Learning και Deep Learning

Η μετάβαση στη βαθιά μάθηση ήρθε ως εξέλιξη στην αναπαραστατική μάθηση (Representation Learning). Αρχικά, η χειροκίνητη εξαγωγή χαρακτηριστικών αποτέλεσε μια πολύ απαιτητική και χρονοβόρα διαδικασία, αλλά αυτό αντιμετωπίστηκε με την ανάπτυξη της αναπαραστατικής μάθησης, που είναι η εκμάθηση της χαρτογράφησης (αντιστοίχισης) της αρχικής υψηλών-διαστάσεων αναπαράστασης δεδομένων σε μια ενδιάμεση χαμηλών-διαστάσεων αναπαράσταση και μετά πίσω στην αρχική αναπαράσταση. Ένα απλό παράδειγμα Representation Learning είναι ο αυτόματος κωδικοποιητής (auto-encoder).

Το πρόβλημα που εμφανίζει το Representation Learning είναι οι πολλοί παράγοντες μεταβλητότητας που σχετίζονται με τα χαρακτηριστικά που μαθαίνονται, για παράδειγμα, οι διαφορετικές γωνίες θέασης από τις οποίες λαμβάνεται μια εικόνα ενός αυτοκινήτου. Εδώ έρχεται η βαθιά Μηχανική Μάθηση για την επίλυση αυτού του κρίσιμου προβλήματος. Η ιεραρχική φύση των τεχνικών βαθιάς μάθησης επιτρέπει τη δημιουργία πολύπλοκων εννοιών από απλούστερες. Το βασικό παράδειγμα μιας τεχνικής βαθιάς μάθησης είναι το δίκτυο προώθησης που ονομάζεται πολυστρωματικό perceptron (MLP). Τα οφέλη της βαθιάς μάθησης επιτυγχάνονται μέσω της χρήσης βαθιών Νευρωνικών Δικτύων (Deep Neural Networks - DNN). Είναι μια μορφή τεχνητού νευρωνικού δικτύου στο οποίο η διάταξη των νευρώνων εμπνέεται από τη διάταξη των νευρώνων του οπτικού φλοιού των ζώων. Το DNN παρέχει τα ακόλουθα σημαντικά πλεονεκτήματα:

1. Ιεραρχική εξαγωγή χαρακτηριστικών, δηλαδή δεν χρειάζεται χειροποίητα χαρακτηριστικά.
2. Περιορισμένη προεπεξεργασία εικόνων εισόδου. Το DNN έχει τρία είδη επιπέδων: στρώμα εισόδου, χρυφά επίπεδα και στρώμα εξόδου. Κάθε επίπεδο έχει μη γραμμικές μονάδες γνωστές ως κόμβοι που βοηθούν στη μοντελοποίηση σύνθετων χαρακτηριστικών. Το DNN μπορεί να μάθει λειτουργίες ακολουθώντας είτε την επιβλεπόμενη μέθοδο εκμάθησης είτε τη μέθοδο χωρίς επίβλεψη.

2.4 Τεχνητά Νευρωνικά Δίκτυα



Σχήμα 2.8: Βιολογικός Νευρώνας και Τεχνητός Νευρώνας

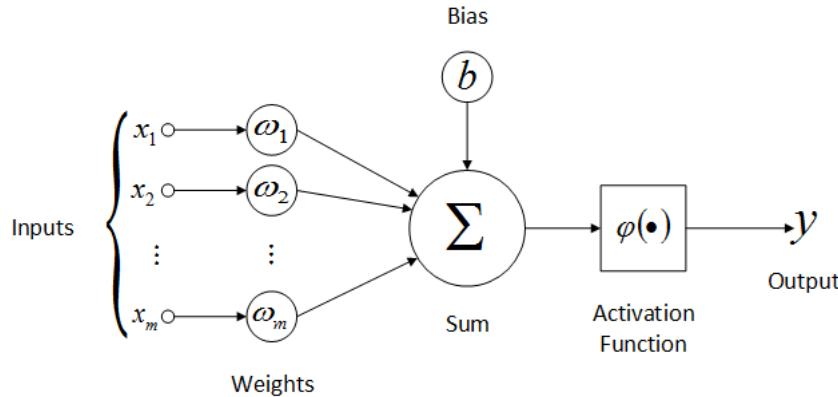
Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks-ANN) είναι υπολογιστικά συστήματα εμπνευσμένα από τα βιολογικά νευρωνικά δίκτυα του εγκεφάλου (2.8). Τα βιολογικά νευρωνικά δίκτυα συνθέτονται από συνάψεις μονάδων (νευρώνων) με σκοπό να εκτελέσουν συγκεκριμένες λειτουργίες όταν ενεργοποιηθούν. Οι συνάψεις των νευρώνων σε ένα βιολογικό δίκτυο είναι συνδυασμός χημικών και ηλεκτρικών λειτουργιών που μαθαίνουν να εκτελούν διεργασίες χωρίς να έχουν προγραμματιστεί ρητά για το σκοπό αυτό. Τα τεχνητά νευρωνικά δίκτυα, επομένως, προσπαθούν να συνθέσουν ένα απλοποιημένο μοντέλο των βιολογικών και όχι να τα αντιγράψουν.

Ιστορικά, ο πρώτος τεχνητός νευρώνας δημιουργείται το 1943 [18] όπου οι Warren McCulloch και Walter Pitts παρουσίασαν ένα μοντέλο νευρώνα με δυαδική συνάρτηση ενεργοποίησης, ικανό να αναπαριστά boolean συναρτήσεις. Η απλότητα όμως του μοντέλου των McCulloch και Pitts δημιουργούσε ορισμένα προβλήματα καθώς επέτρεπε μόνο δυαδικές εισόδους, χρησιμοποιούσε μόνο τη βηματική σαν συνάρτηση ενεργοποίησης, δεν ενσωμάτωνε βάρη στην είσοδο και δεν μπορούσε να περιγράψει μη γραμμικές συναρτήσεις.

Μερικά από τα προβλήματα αυτά μπόρεσαν να αντιμετωπιστούν το 1958 όπου ο Frank Rosenblatt παρουσίασε μια πιο ολοκληρωμένη μοντέλοποίηση του τεχνητού νευρώνα, το perceptron (αντίληπτρο) [19]. Τα επόμενα χρόνια ήταν γεμάτα ανακαλύψεις, αξιοσημείωτη είναι ωστόσο η δημοσίευση των Minsky και Papert το 1969 [20], όπου αποδεικνύουν ότι οι νευρώνες perceptrons μπορούν να εκπαιδευτούν για την επίλυση γραμμικών διαχωριζόμενων προβλημάτων και μόνο.

Ένα από τα πιο χαρακτηριστικά παραδείγματα μη γραμμικών διαχωριζόμενων προβλημάτων είναι το αποκλειστικό OR (XOR)¹. Προκειμένου να επιλυθεί επιτυχώς το πρόβλημα XOR πρέπει η έξοδος του δικτύου να είναι αληθής αν και μόνο αν μία είσοδος του δικτύου είναι αληθής. Οι Minsky και Papert γνώριζαν ότι αυτό μπορούσε να αντιμετωπιστεί μόνο με τη χρήση πολυεπίπεδων δικτύων, η μέχρι τότε γνώση όμως δεν επέτρεπε την εκπαίδευση ενός τέτοιου δικτύου. Λύση στο πρόβλημα αυτό βρέθηκε να δώσει ο αλγόριθμος backpropagation του Paul Jhon Werbos το 1975.

¹https://en.wikipedia.org/wiki/XOR_gate



Σχήμα 2.9: Η δομή ενός νευρώνα Perceptron [21]

2.4.1 Ο αλγόριθμος Perceptron

Ο νευρώνας perceptron, ή αλγόριθμος perceptron, αποτελεί το βασικό δομικό στοιχείο των σύγχρονων τεχνητών νευρωνικών δικτύων. Στην ουσία, ο αλγόριθμος perceptron είναι ένας δυαδικός ταξινομητής, δηλαδή είναι μια συνάρτηση που μπορεί να αποφανθεί αν η είσοδος ανήκει σε μία κλάση ή όχι. Στο σχήμα 2.9 παρουσιάζεται γραφικά η εσωτερική δομή του νευρώνα, στη συνέχεια περιγράφεται αναλυτικότερα ο αλγόριθμος για ένα perceptron ενός επιπέδου.

Όπως αναφέρθηκε ο αλγόριθμος perceptron αποτελεί επί της ουσίας μία συνάρτηση που αντιστοιχεί ένα διάνυσμα εισόδου $\mathbf{x} = [x_1, x_2, x_3, \dots, x_n] \in \mathbb{R}^n$ σε μία τιμή εξόδου $\varphi(\mathbf{x})$. Η τιμή της εξόδου υπολογίζεται από τη σχέση:

$$y = \varphi(\mathbf{w} \cdot \mathbf{x} + b) \quad (2.1)$$

όπου το $\mathbf{w} \cdot \mathbf{x} = \sum_{i=1}^n x_i w_i$ είναι το εσωτερικό γινόμενο του διανύσματος εισόδου \mathbf{x} με το διάνυσμα βαρών $\mathbf{w} = [w_1, w_2, \dots, w_n] \in \mathbb{R}^n$ και b είναι η πόλωση του νευρώνα. Η τιμή της εξόδου y εξαρτάται από τη συνάρτηση ενεργοποίησης φ , συνήθως παίρνει τιμές στο \mathbb{R} , ή στο $[0, 1]$ ή ακόμα και στο $\{0, 1\}$. Η φύση του αλγόριθμου, ως γραμμικός ταξινομητής, του επιτρέπει να ταξινομήσει σωστά όλα τα διανύσματα εισόδου σε δύο κλάσεις, μόνο εάν αυτά είναι γραμμικά διαχωρίσιμα. Σε αντίθετη περίπτωση, αν τα διανύσματα εισόδου δεν είναι γραμμικά διαχωρίσιμα όπως στο πρόβλημα XOR που αναφέρθηκε στην προηγούμενη ενότητα, χρειάζονται περισσότερα επίπεδα νευρώνων για να μπορέσουν να ταξινομηθούν σε δύο κλάσεις. Τα πολυεπίπεδα νευρωνικά δίκτυα θα περιγραφούν με λεπτομέρεια στη συνέχεια (2.4.3).

2.4.2 Θεώρημα Καθολικής Προσέγγισης

Η σημαντικότητα του θεωρήματος καθολικής προσέγγισης (Universal Approximation Theorem) συνδέεται άμεσα με την ισχύ των νευρωνικών δικτύων, συγκεκριμένα των εμπρόσθια τροφοδοτούμενων δικτύων (feedforward networks). Ένα εμπρόσθια τροφοδοτούμενο δίκτυο είναι ένα νευρωνικό δίκτυο στο οποίο οι συνδέσεις μεταξύ των νευρώνων δεν σχηματίζουν κύκλο. Το

Θεώρημα καθολικής προσέγγισης αυτό που αποδεικνύει είναι ότι ένα εμπρόσθια τροφοδοτούμενο δίκτυο με μόνο ένα χρυφό επίπεδο, πέρα από τα επίπεδα της εισόδου και της εξόδου, και πεπερασμένο αριθμό νευρώνων μπορεί να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση ορισμένη σε ένα κλειστό σύνολο των πραγματικών αριθμών, με οσοδήποτε μικρό σφάλμα.

Μία πιο επίσημη έκφραση του θεωρήματος καθολικής προσέγγισης διατυπωμένη με μαθηματικούς όρους είναι η παρακάτω. Έστω μια συνάρτηση $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ η οποία είναι μη σταθερή, συνεχής και φραγμένη, γνωστή ως συνάρτηση ενεργοποίησης. Συμβολίζουμε με I_m τον m -διάστατο υπερκύβο στο χώρο $[0, 1]^m$ και με $C(I_m)$ την κλάση όλων των πραγματικών συνεχών συναρτήσεων στο I_m . Τότε, δεδομένου οποιοδήποτε αριθμού $\varepsilon > 0$ και οποιαδήποτε συνάρτηση $f \in I_m$, υπάρχει ένας ακέραιος N , σταθερές $r_i, b_i \in \mathbb{R}$ και διανύσματα $w_i \in \mathbb{R}^m$ για $i = 1, 2, \dots, N$ τέτοια ώστε να οριστεί η συνάρτηση $F(\mathbf{x})$:

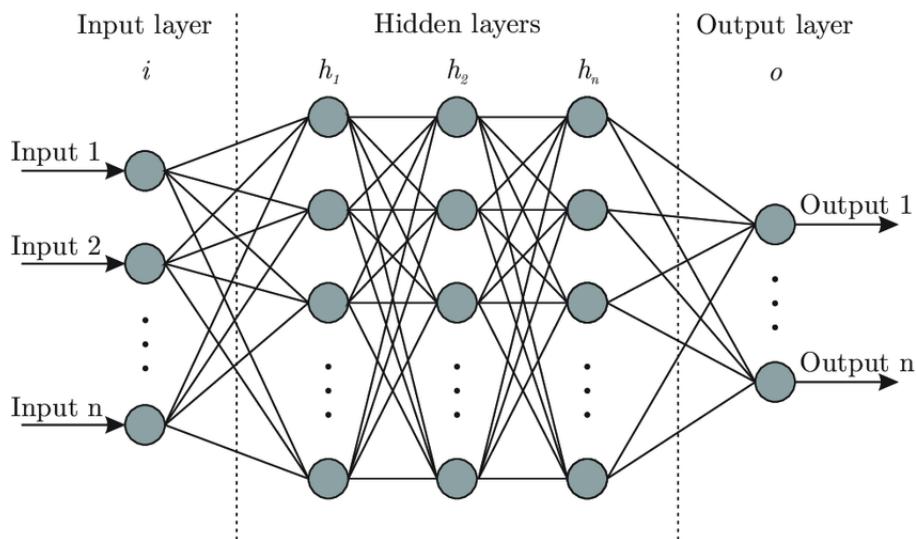
$$F(\mathbf{x}) = \sum_{i=1}^N r_i \varphi(\mathbf{w}_i^T \mathbf{x} + b_i) \quad (2.2)$$

ως μια προσέγγιση της συνάρτησης f , τέτοια ώστε να ισχύει:

$$|F(\mathbf{x}) - f(\mathbf{x})| < \varepsilon \quad (2.3)$$

για όλα τα $\mathbf{x} \in I_m$.

2.4.3 Πολυεπίπεδα Νευρωνικά Δίκτυα



Σχήμα 2.10: Τα επίπεδα ενός τεχνητού νευρωνικού δικτύου [22]

Στις προηγούμενες υποενότητες έγινε αναφορά στα πολυεπίπεδα νευρωνικά δίκτυα (Multi-layer Perceptrons-MLP) και στη σημαντικότητα της χρήσης τους. Όπως ειπώθηκε, ένας νευρώνας Perceptron μπορεί να προσεγγίσει μόνο γραμμικές συναρτήσεις, μη γραμμικά προβλήματα

απαιτούν την χρήση μη γραμμικών συναρτήσεων, οι οποίες μπορούν να προσεγγιστούν από πολλαπλά επίπεδα νευρώνων Perceptron. Ωστόσο πολλές φορές, ο όρος πολυεπίπεδα νευρωνικά δίκτυα χρησιμοποιείται γενικότερα για να περιγράψει όλα τα εμπρόσθια τροφοδοτούμενα νευρωνικά δίκτυα και όχι μόνο δίκτυα που αποτελούνται από νευρώνες Perceptrons. Τα πολυεπίπεδα νευρωνικά δίκτυα συνθέτονται από τουλάχιστον τρία επίπεδα νευρώνων όπου το πρώτο επίπεδο ονομάζεται επίπεδο εισόδου και περιγράφει την είσοδο του δικτύου, το τελευταίο ονομάζεται επίπεδο εξόδου και περιγραφεί την έξοδο του δικτύου ενώ όλα τα ενδιάμεσα ονομάζονται ‘χρυφά’ επίπεδα (hidden layers) όπως φαίνεται και στην εικόνα ;;. Επί της ουσίας, κάθε επίπεδο έχει ως είσοδο την έξοδο του προηγούμενου και τροφοδοτεί με την έξοδό του το επόμενο. Συγκεκριμένα, στις οριακές περιπτώσεις, είσοδο στο επίπεδο εισόδου αποτελούν τα δεδομένα εισόδου ενώ η έξοδος στο επίπεδο εξόδου συνθέτει την έξοδο του δικτύου.

Στην περίπτωση που τα ενδιάμεσα επίπεδα είναι παραπάνω από ένα τότε αναφερόμαστε στην κατηγορία των βαθιά νευρωνικών δικτύων (Deep Neural Networks-DNN). Τα βαθιά νευρωνικά δίκτυα χρησιμοποιούνται κατά κόρον σε εφαρμογές βαθιάς μηχανικής μάθησης (Deep Learning), όπου η αρχιτεκτονική τους βοηθά στην εξαγωγή αφηρημένων χαρακτηριστικών. Για παράδειγμα, σε ένα νευρωνικό δίκτυο με τρία χρυφά επίπεδα, όπου ως είσοδο θα είχε εικόνες προσώπων και θα προσπαθούσε να τα αναγνωρίσει, η διαδικασία εξαγωγής χαρακτηριστικών θα γινόταν διαδικτικά. Δηλαδή, το πρώτο χρυφό επίπεδο θα μπορούσε να αναγνωρίζει απλά χαρακτηριστικά όπως γραμμές και απλές καμπύλες, το δεύτερο χρυφό επίπεδο θα μπορούσε να αναγνωρίζει πιο σύνθετα χαρακτηριστικά του προσώπου όπως τα μάτια ή η μύτη και τέλος το τρίτο χρυφό επίπεδο να αναγνωρίζει ολόκληρη την σύνθεση ενός προσώπου.

Τα πολυεπίπεδα νευρωνικά δίκτυα εκπαιδεύονται, δηλαδή ανανεώνουν τα βάρη τους, κάνοντας χρήση μια τεχνική επιβλεπόμενης μάθησης που ονομάζεται οπισθοδιάδοση (Backpropagation). Στην περίπτωση ενός προβλήματος ταξινόμησης της εισόδου σε τρεις κατηγορίες, το δίκτυο θα ανανέωνε τα βάρη του με τέτοιο τρόπο ώστε το επίπεδο εξόδου να αποδώσει τις κατάλληλες τιμές και η είσοδος να ταξινομηθεί στη σωστή κατηγορία. Οι τιμές που θα δώσει η έξοδος αλλά και κάθε επίπεδο του δικτύου εξαρτάται από την συνάρτηση ενεργοποίησης που χρησιμοποιεί κάθε επίπεδο. Στο παράδειγμα ταξινόμησης της εισόδου σε τρεις κατηγορίες το επίπεδο εξόδου θα αποτελούνται από τρεις νευρώνες, όπου κάθε ένας θα αντιπροσώπευε μία κατηγορία, και η συνάρτηση ενεργοποίησης των νευρώνων 2.16 θα ήταν υπεύθυνη για την ταξινόμηση της εισόδου στις τρεις κατηγορίες.

2.4.4 Αλγόριθμοι Εκπαίδευσης και Βασικές Συναρτήσεις

2.4.4.1 Συνάρτηση Κόστους

Η συνάρτηση κόστους (Cost function) είναι μία μετρική της επίδοσης των συστημάτων μηχανικής μάθησης. Η επίδοση στα συστήματα μηχανικής μάθησης συνήθως μετράται ως τη διαφορά που έχουν οι προβλεπόμενες τιμές του συστήματος σε σχέση με τις αναμενόμενες τιμές. Σχοπός λοιπόν της συνάρτησης κόστους είναι να υπολογίσει το σφάλμα μεταξύ προβλεπόμενων και πραγματικών τιμών. Οι τιμές που αποκτά το σφάλμα συνηθίζεται να είναι πραγματικοί αριθμοί. Ο

σκοπός της συνάρτησης κόστους είναι τις περισσότερες φορές να ελαχιστοποιήσει την τιμή της, άλλα ενδέχεται να πρέπει και να την μεγιστοποιήσει, ανάλογα με τη φύση του προβλήματος. Για παράδειγμα, αν η συνάρτηση κόστους υπολογίζει μία ανταμοιβή για το σύστημα τότε επιθυμητό θα ήταν να μεγιστοποιήσει την τιμή της. Ενώ σε αντίθετη περίπτωση, αν η συνάρτηση κόστους υπολογίζει ένα κόστος του συστήματος, όπως δηλώνει και το όνομά της, τότε η τιμή της θα πρέπει να ελαχιστοποιηθεί.

Ένα παράδειγμα συνάρτησης που προσπαθεί να ελαχιστοποιήσει την τιμή της είναι η συνάρτηση μέσου τετραγωνικού σφάλματος που περιγράφεται με την σχέση 2.4. Στη συνάρτηση μέσου τετραγωνικού σφάλματος (Mean Squared Error) μετράται το τετραγωνικό σφάλμα μεταξύ της προβλεπόμενης ($f(x_i|\vartheta)$) και της αναμενόμενης (y_i) τιμής της εξόδου για N δεδομένα εισόδου (x_i) και παραμέτρους δικτύου (ϑ).

$$MSE(\vartheta) = \frac{1}{N} \sum_{i=1}^N (f(x_i|\vartheta) - y_i)^2 \quad (2.4)$$

Πρέπει να γίνει κατανοητό ότι η επιτυχία ενός νευρωνικού δικτύου συνδέεται με την ικανότητα του να γενικεύσει. Δηλαδή, για ένα σύνολο δεδομένων εκπαίδευσης το δίκτυο θα πρέπει να είναι σε θέση να προβλέψει με επιτυχία και για το σύνολο των δεδομένων δοκιμής. Η τιμή της συνάρτησης κόστους μειώνεται όσο το δίκτυο εκπαίδευται με τα δεδομένα εκπαίδευσης. Η εκτεταμένη εκπαίδευση του δικτύου συχνά μπορεί να οδηγήσει και στο πρόβλημα της υπερπροσαρμογής (overfitting), αυτό ουσιαστικά σημαίνει ότι το δίκτυο μαθαίνει να προβλέπει πολύ καλά για τα 'γνωστά' δεδομένα εκπαίδευσης ενώ αδυναμεί στις προβλέψεις των 'άγνωστων' δεδομένων δοκιμής. Σκοπός του δικτύου, όπως αναφέρθηκε, είναι να μπορέσει να γενικεύσει να μπορεί δηλαδή να προβλέψει με μεγάλη ακρίβεια για αγνωστά, στο δίκτυο, δεδομένα και όχι να προσεγγίσει την συνάρτηση που διέπει τα δεδομένα εκπαίδευσης.

2.4.4.2 Αλγόριθμος Backpropagation

Ο αλγόριθμος Backpropagation είναι ένα από τα σημαντικότερα δομικά στοιχεία ενός νευρωνικού δικτύου. Ο όρος backpropagation και η γενική χρήση του στα νευρωνικά δίκτυα αρχικά εδραιώθηκε το 1986 [23] από τους Rumelhart, Hinton και Williams, ενώ η ιδέα για την διαδικασία που περιγράφει ο αλγόριθμος υπήρχε ήδη από τη δεκαετία του 60. Ο αλγόριθμος ουσιαστικά χρησιμοποιείται για να εκπαίδευσει ένα δίκτυο κάνοντας χρήση του κανόνα της αλυσίδας. Μια απλή περιγραφή είναι ότι μετά από κάθε εμπρόσθιο πέρασμα (forward pass) στο δίκτυο ο αλγόριθμος πραγματοποιεί ένα πέρασμα με αντίθετη φορά (backward pass) υπολογίζοντας τις μερικές παραγώγους των παραμέτρων του δικτύου, δηλαδή τα βάρη και τις πολώσεις. Παρακάτω γίνεται μια μαθηματική περιγραφή της διαδικασίας αυτής.

Βασικός στόχος του αλγορίθμου είναι να υπολογίσει τη μερική παράγωγο μιας συνάρτησης κόστους, έστω J αναλυτική αναφορά γίνεται στην υποενότητα 2.4.4.1, ως προς τις παραμέτρους του δικτύου. Έστω ένας κόμβος με βάρος $w_{j,k}^l$ και πόλωση b_j^l , όπου τα j, k δηλώνουν την θέση του στοιχείου στο αντίστοιχο διάνυσμα ενώ το l τον αριθμό του επιπέδου στο οποίο αναφερόμαστε.

Η μερική παραγωγος του βάρους $w_{j,k}^l$ υπολογίζεται με τον κανόνα της αλυσίδας ως:

$$\frac{\partial J}{\partial w_{j,k}^l} = \frac{\partial J}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{j,k}^l} \quad (2.5)$$

Εξ ορισμού, έστω m ο αριθμός των κόμβων στο επίπεδο $l - 1$:

$$z_j^l = \sum_{k=1}^m w_{j,k}^l a_k^{l-1} + b_j^l \quad (2.6)$$

λόγω παραγώγισης προκύπτει:

$$\frac{\partial J}{\partial w_{j,k}^l} = a_k^{l-1} \quad (2.7)$$

Άρα η τελική τιμή της μερικής παραγώγου είναι:

$$\frac{\partial J}{\partial w_{j,k}^l} = \frac{\partial J}{\partial z_j^l} a_k^{l-1} \quad (2.8)$$

Η ίδια διαδικασία ακολουθείται και για τον υπολογισμό της μερικής παραγώγου της πόλωσης b_j^l :

$$\frac{\partial J}{\partial b_j^l} = \frac{\partial J}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} \quad (2.9)$$

Λόγω της εξίσωσης 2.6 προκύπτει:

$$\frac{\partial J}{\partial b_j^l} = 1 \quad (2.10)$$

και τελικά σε αυτή την περίπτωση η τιμή της μερικής παραγώγου είναι:

$$\frac{\partial J}{\partial b_j^l} = \frac{\partial J}{\partial z_j^l} \quad (2.11)$$

Στη συνέχεια, οι τιμές των μερικών παραγώγων που υπολογίστηκαν από τον αλγόριθμο backpropagation θα χρησιμοποιηθούν από τον αλγόριθμο κατάβασης κλήσης για την ανανέωση των παραμέτρων του δικτύου. Η διαδικασία αυτή αναλύεται στην επόμενη ενότητα 2.4.4.3.

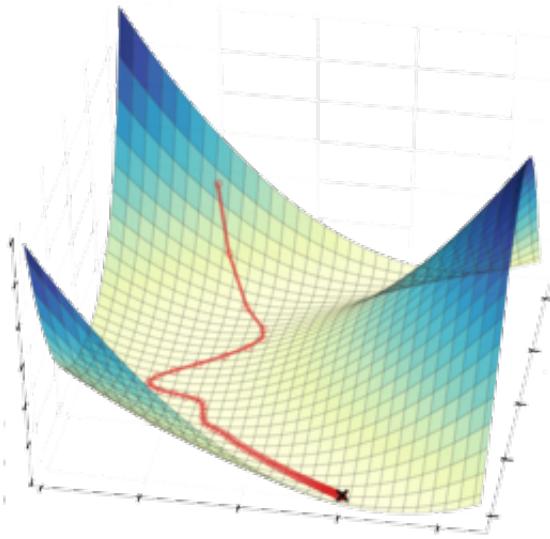
2.4.4.3 Αλγόριθμος Κατάβασης Κλίσης

Ο αλγόριθμος κατάβασης κλίσης (Gradient Descent Algorithm) είναι ένας αλγόριθμος βελτιστοποίησης και στον τομέα της μηχανικής μάθησης η χρήση του είναι πολύ συχνή. Πιο συγκεκριμένα, ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει μια συνάρτηση κόστους, έστω $J(\vartheta)$ όπου ϑ είναι το διάνυσμα που αντιπροσωπεύει τις παραμέτρους του δικτύου, εν προκειμένω τα βάρη και τις πολώσεις. Ο αλγόριθμος κατάβασης κλίσης είναι επαναληπτικός και σε κάθε επανάληψη αφαιρεί μια μικρή ποσότητα από τις παραμέτρους του συστήματος. Την ποσότητα αυτή συνθέτει το γινόμενο της μερικής παραγώγου της συνάρτησης κόστους ως προς την παράμετρο που ανανεώνεται, όπως υπολογίζεται από τον αλγόριθμο Backpropagation, επί την παράμετρο r που ονομάζεται ρυθμός μάθησης (learning rate) και καθορίζει το μέγεθος του βήματος εκμάθησης

που θα κάνει κάθε φόρα ο αλγόριθμος. Με την παραχώτω σχέση 2.12 περιγράφεται πως γίνεται η ανανέωση της παραμέτρου θ_j τη χρονική στιγμή $t + 1$

$$\vartheta_{j,t+1} = \vartheta_{j,t} - r \frac{\partial}{\partial \vartheta_{j,t}} J(\vartheta) \quad (2.12)$$

Στην παραχώτω εικόνα 2.11 γίνεται παρουσίαση ενός παραδείγματος της διαδρομής που ακολουθεί ο αλγόριθμος κατάβασης κλίσης προσεγγίζοντας την ελάχιστη τιμή της συνάρτησης κόστους.



Πηγή: datasciencecentral.com/profiles/blogs/alternatives-to-the-gradient-descent-algorithm
Σχήμα 2.11: Γραφική απεικόνιση του αλγόριθμου κατάβασης κλίσης

Πολλές φορές ο αλγόριθμος κατάβασης κλίσης συναντάται και σε διάφορες εκδοχές οι οποίες, σε αντίθεση με την γενική περίπτωση (Batch Gradient Descent - BGD), χρησιμοποιούν ένα μέρος των δεδομένων εισόδου κάθε φορά και όχι το σύνολό τους. Συγκεκριμένα, αυτές οι εκδοχές του αλγόριθμου, χωρίζουν τα δεδομένα εισόδου σε τεμάχια (batches), όπου κάθε τέτοιο τεμάχιο χρησιμοποιείται με την σειρά για την ανανέωση των παραμέτρων του δικτύου.

Ο Batch Gradient Descent θα υπολογίσει την παράγωγο σε όλα τα δεδομένα και θα πραγματοποιήσει μια ανανέωση. Για το λόγο αυτό μπορεί να είναι πολύ αργός και για μεγάλα δεδομένα εισόδου που δεν χωρούν στη μνήμη ο έλεγχος να προκύψει πολύ δύσκολος.

Μια εκδοχή του αλγόριθμου κατάβασης κλήσης είναι ο Stochastic Gradient Descent (SGD), ο οποίος σε αντίθεση με τον προηγούμενο πραγματοποιεί ανανέωση παραμέτρων για κάθε δεδομένο εκπαίδευσης. Λόγω των συχνών ανανέωσεων, η ενημέρωση των παραμέτρων παρουσιάζει υψηλή διακύμανση και με την σειρά της η συνάρτηση κόστους λαμβάνει κυματινόμενες τιμές. Αυτό είναι εξαιρετικά χρήσιμο, επειδή επιτρέπει στην συνάρτηση κόστους να εντοπίσει νέα τοπικά ελάχιστα. Όμως, οι συνεχείς διακυμάνσεις μπορούν να προκαλέσουν την συνεχή προσπέραση του ολικού ελαχίστου, έτσι τελικά η σύγκλιση του αλγορίθμου να προκύψει αρκετά αργή.

Μια άλλα εκδοχή του αλγορίθμου που αντιμετωπίζει το πρόβλημα αυτό είναι ο Mini Batch

Gradient Descent (MBGD), που χρησιμοποιεί τα προτερήματα των δύο προηγούμενων μεθόδων. Σύμφωνα με αυτή την τεχνική τα δεδομένα εκπαίδευσης χωρίζονται σε μικρά τεμάχια που τροφοδοτούνται στο δίκτυο με τη σειρά και η ανανέωση των παραμέτρων γίνεται για κάθε τεμάχιο. Αυτή η τεχνική προκαλεί μικρότερες διακυμάνσεις στην ανανέωση των παραμέτρων με αποτέλεσμα η σύγκλιση του αλγόριθμου να είναι αρκετά πιο γρήγορη και αποτελεσματική.

2.4.4.4 Αλγόριθμοι Βελτιστοποίησης

Αντικείμενο της βελτιστοποίησης (Optimization) στα συστήματα μηχανικής μάθησης είναι η ελαχιστοποίηση, πιο σπάνια η μεγιστοποίηση, μιας συνάρτησης στόχου. Παραδείγματος χάρη ο αλγόριθμος κατάβασης κλίσης που παρουσιάστηκε στην προηγούμενη υποενότητα 2.4.4.3 είναι και αυτός ένας αλγόριθμος βελτιστοποίησης που ελαχιστοποιεί μια συνάρτηση κόστους. Για την αποδοτική και αποτελεσματική εκπαίδευση του συστήματος ο σωστός υπολογισμός και η ανανέωση των εσωτερικών παραμέτρων παιίζει πολύ σημαντικό ρόλο. Για τον λόγο αυτό στρατηγικές και αλγόριθμοι βελτιστοποίησης εφαρμόζονται για τον υπολογισμό των παραμέτρων του συστήματος που θα επηρεάσουν την διαδικασία εκπαίδευσης και την έξοδό του.

Οι αλγόριθμοι βελτιστοποίησης χωρίζονται σε δύο κατηγορίες, με βάση την τάξη της παραγώγου που χρησιμοποιούν. Οι αλγόριθμοι βελτιστοποίησης *Πρώτης Τάξης* (*First Order Optimization Algorithms*) υπολογίζουν και χρησιμοποιούν την πρώτη παράγωγο της συνάρτησης στόχου με σκοπό την βελτιστοποίηση της. Η παράγωγος πρώτης τάξης μιας συνάρτησης σε ένα σημείο ουσιαστικά υπολογίζει την τιμή της κλίσης της στο σημείο αυτό, δηλαδή αν η συνάρτηση έχει την τάση να αυξήσει ή να μειώσει την τιμή της στο επόμενο σημείο. Ένας αλγόριθμός βελτιστοποίησης πρώτης τάξης είναι για παράδειγμα ο αλγόριθμος κατάβασης κλίσης.

Η δεύτερη κατηγορία αλγόριθμων βελτιστοποίησης είναι οι *Αλγόριθμοι Βελτιστοποίησης Δεύτερης Τάξης* (*Second Order Optimization Algorithms*). Όπως είναι αναμενόμενο, οι αλγόριθμοι αυτοί υπολογίζουν και χρησιμοποιούν τις παραγώγους δεύτερης τάξης της συνάρτησης στόχου. Η δεύτερη παράγωγος μιας συνάρτησης σε ένα σημείο εκφράζει την κλίση της πρώτης παραγώγου στο σημείο αυτό, δηλαδή την κυρτότητα της συνάρτησης. Όμως ο υπολογισμός της δεύτερης παραγώγου είναι υπολογιστικά πολύ πιο δαπανηρός, έτσι οι αλγόριθμοι αυτοί δεν χρησιμοποιούνται συχνά στην πράξη. Το πλεονέκτημα των αλγόριθμων δεύτερης τάξης είναι ότι είναι αποτελεσματικότεροι καθώς δεν αγνοούν την καμπυλότητα της επιφάνειας.

Στη συνέχεια, με σκοπό την αντιμετώπιση των έντονων ταλαντώσεων που εμφανίζει ο αλγόριθμος κατάβασης κλίσης και κατά συνέπεια την δυσκολία σύγκλισης, παρουσιάζεται μια τεχνική βελτιστοποίησης που ονομάζεται *Ορμή* (*Momentum*). Η τεχνική αυτή εισάγει έναν όρο γ στην 2.12 με αποτέλεσμα να επιταχύνει την διαδικασία σύγκλισης οδηγώντας τις παραμέτρους προς την σχετική κατεύθυνση και εξομαλύνοντας τις ταλαντώσεις στις άσχετες κατευθύνσεις. Όταν χρησιμοποιείται και η ορμή για την ανανέωση παραμέτρων ϑ η σχέση που τις ανανεώνει δίνεται από τις εξισώσεις:

$$V(t) = \gamma V(t-1) + r \nabla J(\vartheta) \quad (2.13)$$

$$\vartheta = \vartheta - V(t) \quad (2.14)$$

, μία τυπική τιμή για την ορμή είναι $\gamma = 0.9$.

Τρεις από τους πιο συνηθισμένους αλγόριθμους βελτιστοποίησης που χρησιμοποιούν την τεχνική της ορμής είναι οι παρακάτω:

- *Adagrad*

Ο Adagrad [24] επιτρέπει στον ρυθμό μάθησης r να προσαρμόζεται σε κάθε παράμετρο βασιζόμενος στις παρελθοντικές κλίσεις (gradients) του. Έτσι για παραμέτρους που ανανεώνονται τις τιμές τους συχνά, ο αλγόριθμος κάνει μικρά βήματα στην ενημέρωση αυτών των παραμέτρων. Ενώ σε αντίθετη περίπτωση κάνει μεγάλα βήματα για παραμέτρους που οι τιμές τους ανανέωνονται λιγότερο συχνά. Ένα πλεονέκτημα του αλγόριθμου είναι ότι δεν απαιτείται να γίνει χειροκίνητη ρύθμιση του ρυθμού εκμάθησης, ενώ μειονέκτημα είναι ότι ο ρυθμός εκμάθησης παίρνει συνεχώς μικρότερες τιμές περιορίζοντας σημαντικά τη διαδικασία εκμάθησης.

- *AdaDelta*

Ο αλγόριθμος AdaDelta [25] είναι μία βελτίωση του Adagrad που αντιμετωπίζει το πρόβλημα του φθίνοντος ρυθμού εκμάθησης. Χρησιμοποιεί μόνο ένα τμήμα των προηγούμενων τιμών της κλίσης της συνάρτησης για την ανανέωση των παραμέτρων σε αντίθεση με τον Adagrad που χρειαζόταν όλες τις προηγούμενες τιμές της κλίσης. Ένα ακόμη προτέρημα του αλγόριθμου είναι ότι δεν χρειάζεται προκαθορισμένη τιμή για το ρυθμό εκμάθησης.

- *Adam*

Ο αλγόριθμός Adam (Adaptive Moment Estimation) [26] είναι και αυτός μία μέθοδος προσαρμοστική ως προς το ρυθμό εκμάθησης. Ο αλγόριθμός χρησιμοποιεί τις τιμές των ροπών πρώτης (μέσος όρος) και δεύτερης (διακύμανση) τάξης για την ανανέωση των παραμέτρων. Σε πολλά σύγχρονα συστήματα μηχανικής μάθησης γίνεται χρήση του αλγόριθμου Adam λόγω της ικανότητας του να συγκλίνει πολύ γρήγορα και να αντιμετωπίζει τα προβλήματα που εμφανίζουν οι προηγούμενες τεχνικές.

2.4.4.5 Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης (Activation function) είναι η συνάρτηση που σε ένα τεχνητό νευρωνικό δίκτυο θα εφαρμοστεί στην έξοδο κάθε κόμβου και η τιμή της θα τροφοδοτηθεί στο επόμενο επίπεδο του δικτύου. Η χρήση της συνάρτησης ενεργοποίησης στα νευρωνικά δίκτυα είναι απαραίτητη καθώς εισάγει μη γραμμικότητα στο σύστημα αλλά και κανονικοποιεί την έξοδο των νευρώνων σε ένα κλειστό σύνολο.

Για να γίνει πιο κατανοητή η σημασία της κανονικοποίησης της εξόδου, ας υποθέσουμε ότι έχουμε ένα νευρωνικό δίκτυο και εξετάζουμε έναν κόμβο του δικτύου με είσοδο το διάνυσμα x , βάρη w και πόλωση b . Η τιμή της εξόδου y , όπως φαίνεται στη σχέση 2.15, μπορεί να παίρνει τιμές στο $(-\infty, +\infty)$. Με την εφαρμογή της συνάρτησης ενεργοποίησης 2.16, έστω f καταφέρνουμε

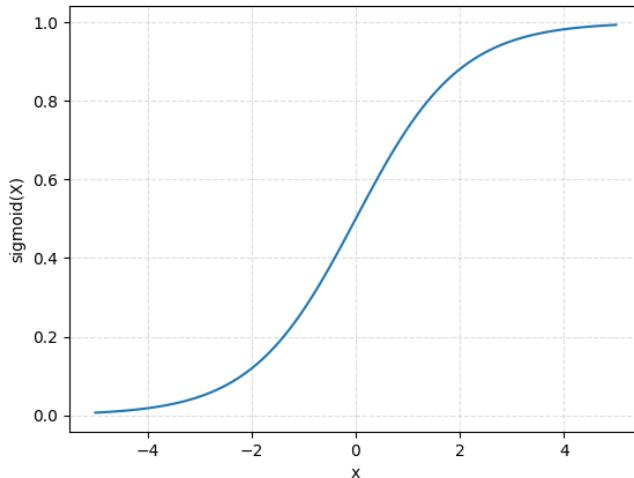
να αντιστοιχίσουμε την τιμή της εξόδου του νευρώνα z στο επιθυμητό σύνολο, συνήθως στο $[0, 1]$. Αυτή η διαδικασία είναι απαραίτητη ώστε ο νευρώνας να μπορέσει να διαχωρίσει τις περιπτώσεις που θα πρέπει να μεταδώσει την τιμή στα επόμενα επίπεδα ή να την αποκόψει. Τιμές κοντά στο ένα έχουν αποτέλεσμα την πυροδότηση (ενεργοποίηση) του νευρώνα ενώ τιμές κοντά στο μηδέν την αποκοπή του.

$$y = \mathbf{w} \cdot \mathbf{x} + b, y \in (-\infty, +\infty) \quad (2.15)$$

$$z = f(y) = f(\mathbf{w} \cdot \mathbf{x} + b), z \in [0, 1] \quad (2.16)$$

Μια άλλη σκοπιμότητα της συνάρτησης ενεργοποίησης είναι η εισαγωγή μη γραμμικότητας στο σύστημα, όπως αναφέρθηκε. Η μη γραμμικότητα σε ένα σύστημα μηχανικής μάθησης είναι αναγκαία ώστε να μπορέσει να προσεγγίσει μη γραμμικές συναρτήσεις. Η φύση πολλών συναρτήσεων ενεργοποίησης ως μη γραμμικές εξυπηρετούν τη σκοπιμότητα αυτή. Στη συνέχεια γίνεται μια σύντομη παρουσίαση των συνηθέστερων συναρτήσεων ενεργοποίησης:

- Σιγμοειδής Συνάρτηση

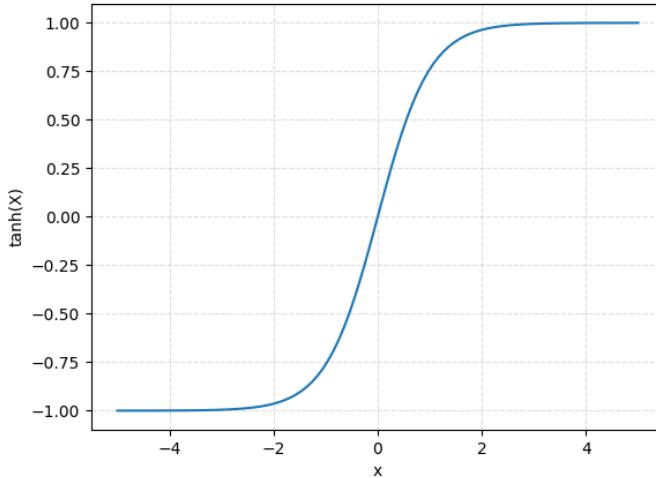


Σχήμα 2.12: Σιγμοειδής συνάρτηση

Η γραφική αναπαράσταση της σιγμοειδούς συνάρτησης Sigmoid function φαίνεται στο σχήμα 2.16, γίνεται εύκολα κατανοητή αλλά στην πράξη δεν χρησιμοποιείται συχνά. Η σιγμοειδής έχει πεδίο τιμών το $(0, 1)$ αλλά αυτό δημιουργεί πρόβλημα γιατί δεν είναι κεντραρισμένο στο μηδέν. Δύο ακόμα προβλήματα που εμφανίζει είναι η εξαφάνιση της κλίσης (Vanishing gradient problem) όπως επίσης και η αργή σύγκλιση που την χαρακτηρίζει στο στάδιο της εκπαίδευσης.

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.17)$$

- Υπερβολική εφαπτομένη

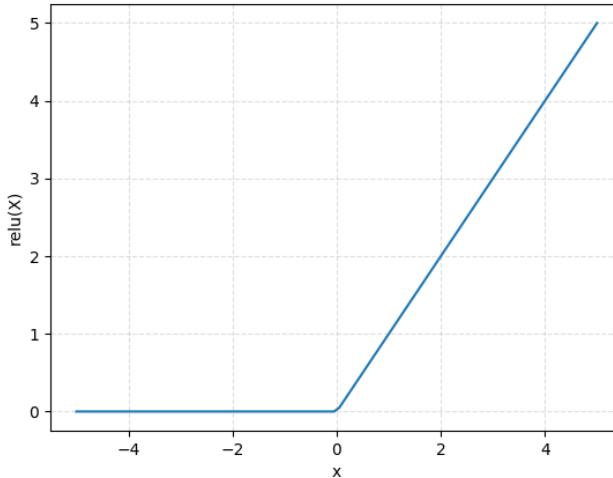


Σχήμα 2.13: Υπερβολική εφαπτομένη

Η υπερβολική εφαπτομένη (Hyperbolic tangent) 2.13 μοιάζει αρχετά με την σιγμοειδή συνάρτηση. Υπερτερεί σε σχέση με την σιγμοειδή καθώς το πεδίο τιμών της παίρνει τιμές στο $(-1, 1)$ και είναι κεντραρισμένο στο μηδέν. Ένα ακόμα προτέρημα είναι η μεγαλύτερη κλίση της που βοήθα στην ταχύτερη σύγκλιση. Όπως και η σιγμοειδής δεν καταφέρνει να αντιμετωπίσει το πρόβλημα της εξαφάνισης κλίσης.

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 = 2\sigma(2x) - 1 \quad (2.18)$$

- Rectified Linear Unit (ReLU)

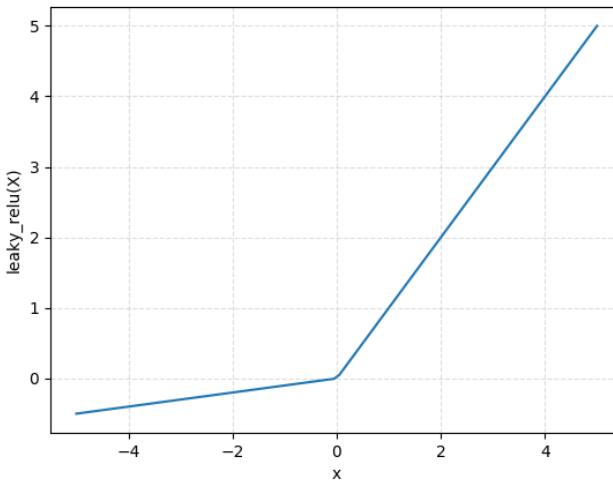


Σχήμα 2.14: Συνάρτηση ReLU

Η συνάρτηση ReLU παρουσιάζει αρκετά προτερήματα και η χρήση της στα σύγχρονα συστήματα μηχανικής μάθησης είναι συχνή. Αντιμετωπίζει με επιτυχία το πρόβλημα της εξαφάνισης κλίσης και η σύγκλισή της είναι ταχύτερη από τις δύο προηγούμενες μεθόδους. Η μαθηματική εξίσωση της ReLU 2.19 είναι πολύ απλή και αποδοτική, ο μόνος περιορισμός που εμφανίζει είναι η χρήση της μόνο στα ενδιάμεσα επίπεδα ενός νευρωνικού δικτύου.

$$f(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.19)$$

- *Leaky Rectified Linear Unit (Leaky ReLU)*

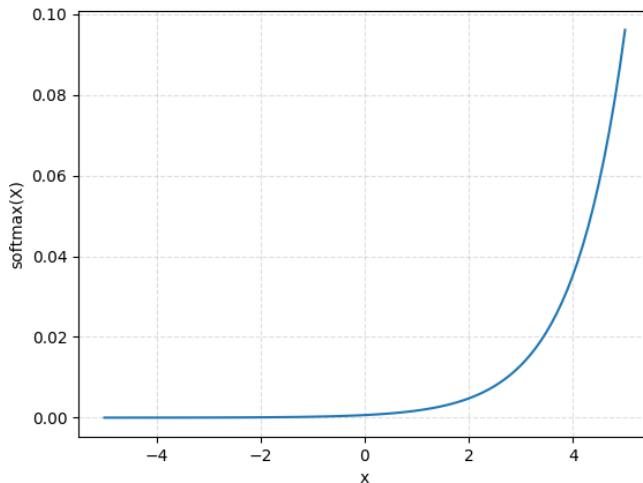


Σχήμα 2.15: Συνάρτηση Leaky ReLU

Η συνάρτηση Leaky ReLU είναι μια βελτιωμένη εκδοχή της ReLU που καταφέρνει να δώσει λύση στο πρόβλημα των Νεκρών Νευρώνων (Dead Neurons) που μπορεί να προκαλέσει η ReLU. Σύμφωνα με το πρόβλημα αυτό η ReLU μπορεί κατά την διάρκεια εκπαίδευσης να οδηγήσει στην ανανέωση κάποιου βάρους και να προκαλέσει απενεργοποίηση της εξόδου για όλα τα δεδομένα εισόδου. Όπως φαίνεται στο σχήμα 2.15, σε αντίθεση με την ReLU, διατηρεί μια πολύ μικρή κλίση για της αρνητικές τιμές αποφεύγοντας έτσι το πρόβλημα των νεκρών νευρώνων.

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (2.20)$$

- Συνάρτηση Softmax



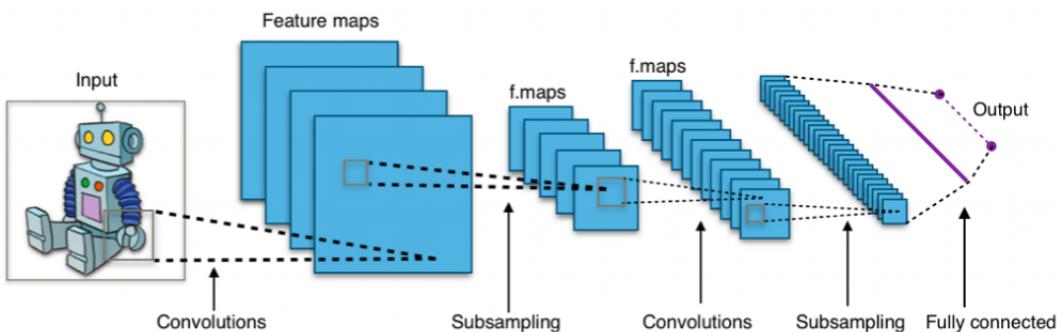
Σχήμα 2.16: Συνάρτηση Softmax

Η συνάρτηση softmax ξεχωρίζει από τις προηγούμενες κατηγορίες συναρτήσεων καθώς εκφράζει πιθανότητες στην έξοδο της. Χρησιμοποιείται συχνά για την κατηγοριοποίηση των δεδομένων εισόδου σε κλάσεις, για αυτό τον λόγο συναντάται συνήθως ως συνάρτηση ενεργοποίησης του τελευταίου επιπέδου σε ένα νευρωνικό δίκτυο. Πιο συγκεκριμένα, η softmax κανονικοποιεί τις εξόδους για κάθε κλάση και στη συνέχεια διαιρεί με το άθροισμα όλων το εξόδων, έτσι εκφράζει την πιθανότητα κάθε είσοδος να ανήκει στην αντίστοιχη κλάση. Η σχέση 2.21 δίνει την δυνατότητα οι έξοδοι να εκφράζονται ως πιθανότητες.

$$\sum_{k=1}^K y_k = \sum_{k=1}^K \text{Softmax}(\mathbf{z})_k = 1 \quad (2.21)$$

2.5 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs) αποτελούν μία κλάση των τεχνητών νευρωνικών δικτύων. Η χρήση τους είναι πολύ συχνή στην ανάπτυξη συστημάτων βαθιάς μηχανικής μάθησης και πιο συγκεκριμένα στην ανάπτυξη εφαρμογών που απαιτούν την ανάλυση εικόνας και βίντεο. Το όνομά 'συνελικτικά νευρωνικά δίκτυα' προέρχεται από την μαθηματική πράξη της συνέλιξης. Η συνέλιξη είναι μια ειδική γραμμική πράξη και τα συνελικτικά νευρωνικά δίκτυα την χρησιμοποιούν ως μια γενική μέθοδο πολλαπλασιασμού πινάκων.

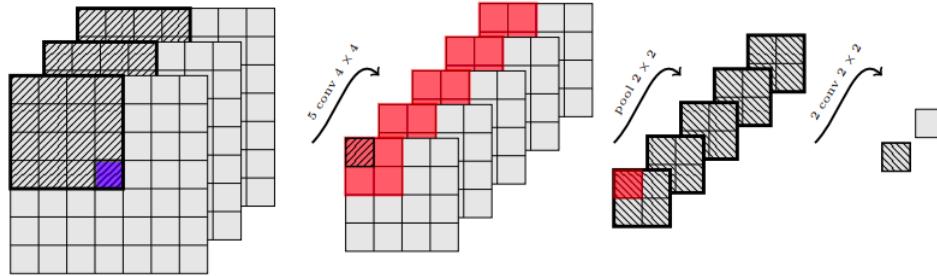


Πηγή: unite.ai/what-are-convolutional-neural-networks/

Σχήμα 2.17: Συνελικτικό Νευρωνικό Δίκτυο

Ιστορικά, το 1968 η εργασία των D. Hubel και T. Wiesel [27] για τον διαχωρισμό των οπτικών κυττάρων του εγκεφάλου σε απλά και σύνθετα άνοιξε το δρόμο για την δημιουργία των συνελικτικών νευρωνικών δικτύων. Αργότερα, το 1980 ο Kunihiko Fukushima εμπνευσμένος από τους προηγούμενους θα είναι ο πρώτος που θα παρουσιάσει το τεχνητό νευρωνικό δίκτυο neocognitron [28] που θα περιλαμβάνει συνελικτικά και υποδειγματοληπτικά επίπεδα. Για να γίνουν δημοφιλή τα CNNs στα συστήματα βαθιάς μηχανικής μάθησης χρειάστηκαν να περάσουν αρκετά χρόνια. Το 2012 ο Alex Krizhevsky με την αρχιτεκτονική που παρουσίασε ως AlexNet [29] κατάφερε να κερδίσει τον διαγωνισμό αναγνώρισης εικόνων ImageNet. Συστήματα που βασίζονται στην αρχιτεκτονική AlexNet αναπτύσσονται μέχρι και σήμερα.

2.5.1 Τρόπος λειτουργίας



Σχήμα 2.18: Ένα απλό Συνελικτικό Δίκτυο: με δεδομένο ένα patch εικόνας που παρέχει ένα πλαίσιο γύρω από ένα pixel προς ταξινόμηση (εδώ μπλε), εφαρμόζεται μια σειρά από συνελίξεις και λειτουργίες υπο-δειγματοληψίας (φίλτρα που ολισθαίνουν μέσω των επιπέδων εισόδου) (εδώ, πέντε συνελίξεις 4×4 , ακολουθούμενες από μία υπο-δειγματοληψία 2×2 , ακολουθούμενη από δύο 2×2 συνελίξεις). Κάθε επίπεδο εξόδου 1×1 ερμηνεύεται ως μια βαθμολογία για μια δεδομένη τάξη.

Τα συνελικτικά νευρωνικά δίκτυα, όπως προαναφέρθηκε, έχουν σχεδιαστεί για την ανάλυση εικόνων. Η αρχιτεκτονική τους είναι ανάλογη με τα συνδετικά πρότυπα ενός ανθρώπινου εγκέφαλου και η οργάνωση τους είναι εμπνευσμένη από τον οπτικό φλοιό. Σε αντιστοίχιση με τον ανθρώπινο εγκέφαλο τα CNNs λαμβάνοντας μια εικόνα καταφέρνουν να ξεχωρίσουν και να αναγνωρίσουν τα σημεία ενδιαφέροντος. Τα σημεία αυτά σε μία εικόνα είναι δισδιάστατα σχήματα και τα CNNs καταφέρνουν να τα αναγνωρίσουν ακόμη και όταν οι παραμορφώσεις της εικόνας είναι υψηλές. Για να το καταφέρουν αυτό τα συνελικτικά νευρωνικά δίκτυα εκπαιδεύονται με έναν επιβλεπόμενο τρόπο ακολουθώντας συγκεκριμένα βήματα εκπαίδευσης. Τα βήματα που ακολουθούν είναι:

- *Εξαγωγή Χαρακτηριστικών (Feature Extraction)*

Σαν πρώτο βήμα κάθε νευρώνας εξάγει τοπικά χαρακτηριστικά αξιοποιώντας τις εισόδους που παίρνει από το προηγούμενο επίπεδο. Για κάθε χαρακτηριστικό που εξάγεται διατηρείται η πληροφορία για τη σχετική θέση του και μειώνεται η σημαντικότητα της, έτσι ώστε να αποκτούν προτεραιότητα τα χαρακτηριστικά που δεν έχουν αναγνωριστεί ακόμα.

- *Αντιστοίχιση Χαρακτηριστικών (Feature Mapping)*

Κάθε χαρακτηριστικό που υπολογίστηκε από την προηγούμενη διαδικασία αποθηκεύεται στη συνέχεια σε έναν χάρτη χαρακτηριστικών. Πολλοί τέτοιοι χάρτες εξάγονται από κάθε συνελικτικό επίπεδο σε ένα CNN. Κάθε χάρτης προέρχεται από την συνέλιξη ενός φίλτρου και του διανύσματος που δέχεται ως είσοδο το επίπεδο. Η διαδικασία αυτή, του φιλτραρίσματος, συνήθως μειώνει τις ελεύθερες παραμέτρους και ακολουθείται από μια διαδικασία ενεργοποίησης. Ως συνάρτηση ενεργοποίησης στο βήμα αυτό χρησιμοποιείται συνήθως η ReLU.

- *Τποδεηγματοληψία (Downsampling)*

Μετά από κάθε συνελικτικό επίπεδο υπάρχει και ένα συγκεντρωτικό επίπεδο. Σκοπός του επιπέδου αυτού είναι να μειώσει τις διαστάσεις των χαρτών χαρακτηριστικών ώστε ελαττωθεί η υπολογιστική ισχύς που απαιτείται για την επεξεργασία των δεδομένων.

- *Αντιστοίχιση Προβλέψεων (Prediction Mapping)*

Στο τέλος ενός συνελικτικού νευρωνικού δικτύου βρίσκεται μια σειρά από πλήρως συνδεδεμένα επίπεδα (Fully Connected Layers). Τα πλήρως συνδεδεμένα επίπεδα ακολουθούν τα επίπεδα που περιγράφηκαν προηγούμενως με σκοπό τη μετατροπή της πληροφορίας που ρέει στο δίκτυο στην επιθυμητή έξοδο και την εξαγωγή προβλέψεων.

Ο τρόπος λειτουργίας των συνελικτικών νευρωνικών δικτύων ακολουθεί τα πρότυπα των πολυεπίπεδων νευρωνικών δικτύων. Ένα προωθητικό πέρασμα (forward pass), που τροφοδοτεί τα δεδομένα εισόδου στο δίκτυο, ακολουθείται από ένα οπισθοδρομικό (backward pass) πέρασμα και η διαδικασία αυτή επαναλαμβάνεται αρκετές φορές. Στο προωθητικό πέρασμα η ροή της πληροφορίας, από την είσοδο στην έξοδο, περνάει από τα διαδοχικά επίπεδα επεξεργασίας. Κάθε επίπεδο δέχεται ως είσοδο την έξοδο του προηγούμενου επιπέδου, τη μετασχηματίζει και την προωθεί στο επόμενο επίπεδο. Όταν τελικά γίνει μια πρόβλεψη για την έξοδο του δικτύου, η τιμή αυτή θα διαδοθεί στο δίκτυο με το οπισθοδρομικό πέρασμα και με σκοπό την ανανέωση των παραμέτρων του δικτύου για την επίτευξη καλύτερης πρόβλεψης.

2.5.2 Επίπεδα επεξεργασίας

Από αρχιτεκτονικής σκοπιάς τα συνελικτικά νευρωνικά δίκτυα δομούνται από ένα σύνολο σειριακών επιπέδων επεξεργασίας. Ένα σύνολο από ομαδοποιημένους νευρώνες συνθέτει κάθε επίπεδο και καθορίζει την λειτουργία του. Στη συνέχεια παρουσιάζεται ο τρόπος κατασκευής και η λειτουργία των διάφορων επιπέδων επεξεργασίας σε ένα συνελικτικό νευρωνικό δίκτυο.

2.5.2.1 Επίπεδο Εισόδου

Το επίπεδο εισόδου (Input layer) είναι το πρώτο επίπεδο επεξεργασίας σε ένα τεχνητό νευρωνικό δίκτυο υπεύθυνο για την τροφοδότηση των δεδομένων εισόδου στο δίκτυο. Οι διαστάσεις των δεδομένων εισόδου καθορίζουν και τις διαστάσεις του επιπέδου εισόδου. Για παράδειγμα, σε ένα συνελικτικό νευρωνικό δίκτυο όπου τα δεδομένα εισόδου είναι ψηφιακές εικόνες με αναπαράσταση RGB διαστάσεων 1024x768 pixels, το επίπεδο εισόδου θα έχει μήκος 1024 νευρώνες, πλάτος 768 και ύψος 3, όσα είναι δηλαδή τα κανάλια σε μια RGB εικόνα (Red, Green, Blue).

2.5.2.2 Συνελικτικό Επίπεδο

Το συνελικτικό επίπεδο (Convolutional layer) είναι το κύριο δομικό στοιχείο ενός συνελικτικού νευρωνικού δικτύου αλλά και το πιο σύνθετο. Δέχεται σαν είσοδο την έξοδο του προηγούμενου επιπέδου, την μετασχηματίζει, εξάγει χαρακτηριστικά και δημιουργεί πίνακες (χάρτες)

χαρακτηριστικών. Για να γίνει πιο εύκολα κατανοητός ο τρόπος λειτουργίας του επιπέδου αυτού στη συνέχεια γίνεται παρουσίαση των διαδικασιών και των εννοιών που το συνθέτουν:

- *Συνέλιξη (Convolution)*

Η συνέλιξη είναι μια μαθηματική πράξη που συνδέει δύο διαφορετικά σύνολα πληροφορίας. Στην περίπτωση των συνελικτικών επιπέδων σε ένα CNN, η πράξη της συνέλιξης γίνεται μεταξύ των δεδομένων εισόδου και ενός φίλτρου (filter) ή αλλιώς πυρήνα (kernel). Με απλά λόγια, η πράξη της συνέλιξης είναι η ολίσθηση ενός παραθύρου, εν προκειμένω του πυρήνα, πάνω στα δεδομένα εισόδου. Το παράθυρο αυτό ολισθαίνοντας θα περάσει πάνω από όλα τα στοιχεία της εισόδου. Σε κάθε βήμα της ολίσθησης πραγματοποιείται ένα-προς-ένα πολλαπλασιασμός πινάκων μεταξύ του πυρήνα και των στοιχείων που καλύπτει το παράθυρο στον πίνακα των δεδομένων εισόδου. Στο σχήμα 2.19 δίνεται ένα παράδειγμα ενός πίνακα εισόδου και ενός πίνακα πυρήνα.

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

Input

1	0	1
0	1	0
1	0	1

Filter / Kernel

Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.19: Παράδειγμα εισόδου και πυρήνα

- *Πίνακες Χαρακτηριστικών (Feature Maps)*

Το αποτέλεσμα του πολλαπλασιασμού των πινάκων εισόδου και πυρήνα αποθηκεύεται σε έναν πίνακα χαρακτηριστικών, όπως φαίνεται στο σχήμα 2.20.

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

Input x Filter

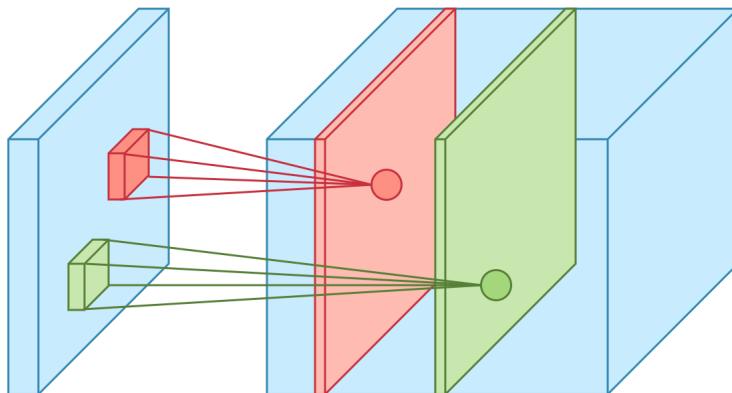
4		

Feature Map

Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.20: Κατασκευή πίνακα χαρακτηριστικών

Ο αριθμός και οι διαστάσεις των φίλτρων ορίζεται στις παραμέτρους του επιπέδου. Για την ολίσθηση κάθε φίλτρου δημιουργείται ένας ξεχωριστός δισδιάστατος πίνακας πίνακας χαρακτηριστικών. Κάθε τέτοιος πίνακας περιγράφει ξεχωριστά χαρακτηριστικά για την είσοδο. Αυτοί οι πίνακες στοιβάζονται (βλέπε σχήμα 2.21), έτσι στην έξοδο του επιπέδου θα σχηματιστεί ένας τρισδιάστατος πίνακας με ύψος όσος είναι ο αριθμός των πυρήνων που έχει καθοριστεί.



Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.21: Στοιβαγμένοι πίνακες χαρακτηριστικών

- *Υπερπαράμετροι (Hyperparameters)*

Οι υπερπαράμετροι του επιπέδου συνέλιξης θα καθορίσουν το μέγεθος των πινάκων χαρακτηριστικών και την φύση της πληροφορίας που περιέχουν. Οι τέσσερις βασικές υπερπαράμετροι που πρέπει να καθοριστούν είναι το μέγεθος του πυρήνα (kernel size), ο αριθμός των πυρήνων (kernel count), το βήμα ολίσθησης (stride) και η επέκταση του περιθωρίου (padding). Όπως αναφέρθηκε ο αριθμός των πυρήνων θα καθορίσει την τρίτη διάσταση

στη έξοδο του συνελικτικού επιπέδου, όπως επίσης και το πλήθιος των διαφορετικών χαρακτηριστικών. Το μέγεθος του πυρήνα καθορίζει τον όγκο της πληροφορίας που αναπαριστά κάθε στοιχείο στον πίνακα εξόδου. Τέλος, το βήμα ολίσθησης καθορίζει το μήκος και το πλάτος της εξόδου, αν το βήμα είναι πάνω από 1 αυτές οι διαστάσεις είναι μικρότερες από ότι είναι στην είσοδο του επιπέδου. Για να αποφύγουμε την μείωση των διαστάσεων μπορούμε να χρησιμοποιήσουμε την τελευταία υπερπαράμετρο για να επεκτείνουμε το περιθώριο της εισόδου όπως στο σχήμα 2.22.

0	0	0	0	0	0
0	35	19	25	6	0
0	13	22	16	53	0
0	4	3	7	10	0
0	9	8	1	3	0
0	0	0	0	0	0

Πηγή: towardsdatascience.com/introduction-to-convolutional-neural-networks-cnn-with-tensorflow-57e2f4837e18

Σχήμα 2.22: Padding στα δεδομένα εισόδου

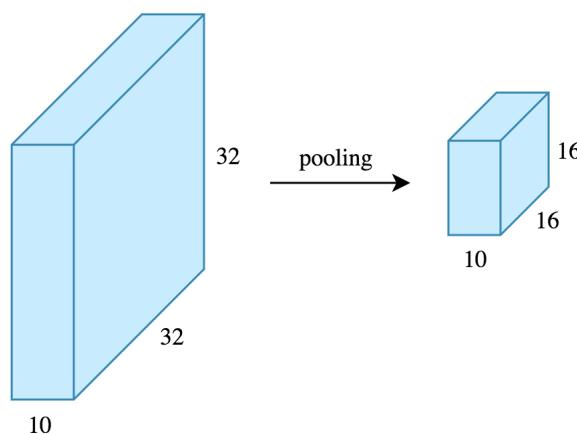
- *Μη-γραμμικότητα (Non-linearity)*

Η εισαγωγή της μη-γραμμικότητας είναι αναγκαία ώστε ένα τεχνητό νευρωνικό δίκτυο να είναι ισχυρό. Σε ένα CNN η έννοια της μη-γραμμικότητας εφαρμόζεται στο συνελικτικό επίπεδο. Συγκεκριμένα, το αποτέλεσμα της συνέλιξης περνάει πρώτα από μία μη-γραμμική συνάρτηση ενεργοποίησης, πριν αποθηκευτεί. Στην πράξη, η συνάρτηση ενεργοποίησης που χρησιμοποιείται είναι η ReLU. Τελικά, αυτό που καταφέρνει η ReLU, σύμφωνα με τη σχέση 2.19, είναι να μηδενίσει όλες τις αρνητικές τιμές σε έναν πίνακα χαρακτηριστικών.

2.5.2.3 Συγκεντρωτικό Επίπεδο

Ένα συγκεντρωτικό επίπεδο (Pooling layer) συνηθίζεται να ακολουθεί κάθε συνελικτικό επίπεδο με σκοπό την μείωση των διαστάσεων του χώρου που αναπαριστά τα δεδομένα. Ουσιαστικά, όπως φαίνεται και από το σχήμα 2.23, το συγκεντρωτικό επίπεδο δειγματοληπτεί τους πίνακες χαρακτηριστικών, μειώνοντας το μήκος και το πλάτος τους, ενώ το ύψος τους παραμένει ανέπαφο. Αυτό, επιτρέπει να ελαττωθεί ο αριθμός των παραμέτρων του συστήματος, το οποίο επιταχύνει την διαδικασία εκπαίδευσης και αντιμετωπίζει το πρόβλημα της υπερπροσαρμογής.

Ο τρόπος που γίνεται η υποδειγματοληψία στα επίπεδα αυτά θυμίζει πολύ την διαδικασία της



Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.23: Εφαρμογή υποδειγματοληψίας

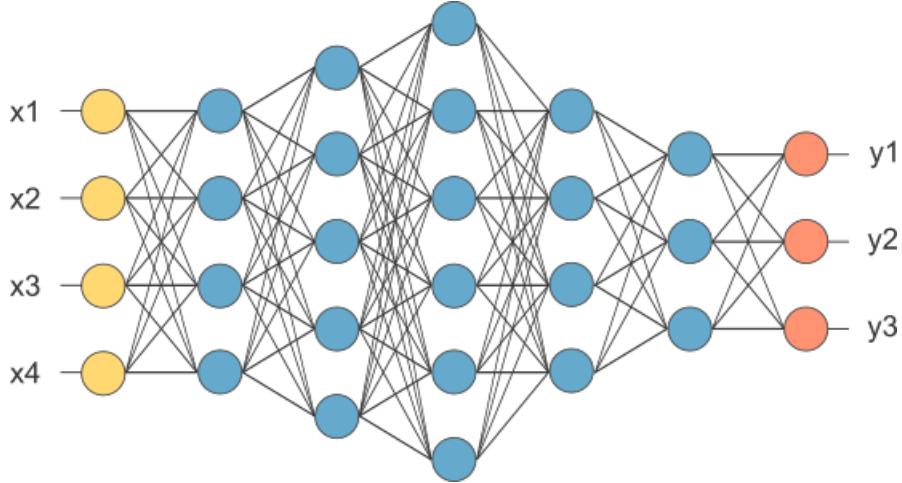
συνέλιξης. Συγκεκριμένα, πάλι ένα παράθυρο ολισθαίνει πάνω από όλα τα δεδομένα εισόδου και υπολογίζει σε κάθε βήμα μία τιμή για τα στοιχεία που καλύπτει. Το μέγεθος του παραθύρου και το βήμα ολίσθησης καθορίζονται στις υπερπαραμέτρους του συγκεντρωτικού επιπέδου. Η τιμή που θα υπολογιστεί σε κάθε βήμα ολίσθησης καθορίζεται από την μέθοδο pooling που θα εφαρμοστεί. Μερικές από τις πιο συνήθεις μεθόδους είναι το max pooling που επιλέγει την μέγιστη τιμή στην περιοχή του παραθύρου, το average pooling που υπολογίζει την μέση τιμή των στοιχείων της περιοχής και το sum pooling που υπολογίζει το άθροισμα των στοιχείων της περιοχής.

2.5.2.4 Πλήρως Συνδεδεμένο Επίπεδο

Τα πλήρως συνδεδεμένα επίπεδα (Fully connected layers) εμφανίζονται σε ομάδες. Σε ένα CNN μια ομάδα από πλήρως συνδεδεμένα επίπεδα συναντάται στο τέλος του δικτύου μετά από μια αλληλουχία συνελικτικών και συγκεντρωτικών επιπέδων. Σε ένα πλήρως συνδεδεμένο επίπεδο κάθε νευρώνας του επιπέδου συνδέεται με όλους του νευρώνες του προηγούμενου επιπέδου, όπως στο σχήμα 2.24. Ένα τέτοιο επίπεδο αναμένει να πάρει ως είσοδο ένα διάνυσμα δεδομένων μίας διάστασης, τα συνελικτικά και συγκεντρωτικά επίπεδα όμως δίνουν διανύσματα τριών διαστάσεων. Για τον λόγο αυτό σε ένα CNN πριν από το πλήρως συνδεδεμένο επίπεδο εφαρμόζεται ένας μετασχηματισμός των δεδομένων που ονομάζεται ισοπέδωση (flattening) και μετατρέπει τα τρισδιάστατα διανύσματα σε μονοδιάστατα, χωρίς να χάνεται καμία πληροφορία.

Σκοπός μιας ομάδας πλήρως συνδεδεμένων επιπέδων είναι να αξιοποιήσουν τα χαρακτηριστικά που έχουν προέλθει από τα συνελικτικά και συγκεντρωτικά επίπεδα, ώστε να παράξουν την επιθυμητή έξοδο. Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης εικόνων τα πλήρως συνδεδεμένα επίπεδα του δικτύου αναλαμβάνουν να επιλέξουν σε ποια κλάση ανήκει κάθε εικόνα της εισόδου. Τέλος, η χρησιμότητα των επιπέδων αυτών δεν περιορίζεται μόνο στην ταξινόμηση της εισόδου αλλά και στην ικανότητα εκμάθησης μη γραμμικών συνδυασμών των χαρακτηριστικών

της εισόδου.



Πηγή: [quora.com/What-is-a-fully-connected-layer](https://www.quora.com/What-is-a-fully-connected-layer)

Σχήμα 2.24: Πλήρως συνδεδεμένα επίπεδα

2.6 Πλήρως Συνελικτικά Δίκτυα - Fully Convolutional Networks

Ένα δίκτυο πλήρους συνέλιξης (FCN) είναι ένα νευρωνικό δίκτυο που εκτελεί μόνο λειτουργίες συνέλιξης (και υποδειγματοληψίας ή υδειγματοληψίας). Ισοδύναμα, ένα FCN είναι ένα CNN χωρίς πλήρως συνδεδεμένα επίπεδα. Το Πλήρως Συνελικτικό Δίκτυο (FCN) χρησιμοποιείται όλο και περισσότερο σε διάφορα προβλήματα τμηματοποίησης ιατρικών εικόνων. Το FCN εισήχθη στον τομέα τμηματοποίησης εικόνας, ως εναλλακτική λύση στη χρήση patches εικόνας. Χρησιμοποιώντας το FCN, η εικόνα θα μπορούσε να αναλυθεί καθολικά αντί να χρησιμοποιηθούν τοπικά patches. Επιπλέον, μπορεί να υπάρξει σύντηξη διαφορετικών κλιμάκων προσθέτοντας συνδέσμους που συνδυάζουν το τελικό επίπεδο πρόβλεψης με χαμηλότερα στρώματα με λεπτότερους διασκελισμούς. Δεδομένου ότι όλα τα πλήρως συνδεδεμένα επίπεδα αντικαθίστανται από στρώματα συνέλιξης, το FCN μπορεί να λάβει εισόδους αυθαίρετων μεγεθών και να παράγει εξόδους αντίστοιχου μεγέθους με αποτελεσματική εξαγωγή συμπερασμάτων (inference) και εκμάθηση (training). Σε αντίθεση με τις μεθόδους που βασίζονται σε patch, η συνάρτηση loss που χρησιμοποιεί αυτήν την αρχιτεκτονική υπολογίζεται σε ολόκληρο το αποτέλεσμα τμηματοποίησης εικόνας.

2.6.1 Μετατιθέμενη Συνέλιξη - Transposed Convolution

To Transposed Convolution μπορεί να θεωρηθεί ως αντίθετη ενέργεια από το Convolution. Σε αντίθεση με τη συνέλιξη, ένα μετατιθέμενο επίπεδο συνέλιξης χρησιμοποιείται για τη δειγματοληψία του χαρακτηριστικού μειωμένης ανάλυσης πίσω στην αρχική του ανάλυση. Ένα σύνολο τιμών διασκελισμού (stride) και γεμίσματος (padding) μαθαίνεται για να ληφθεί η τελική έξοδος

από τα χαρακτηριστικά χαμηλότερης ανάλυσης. Η παρακάτω εικόνα εξηγεί τη διαδικασία με πολύ κατανοητό τρόπο.

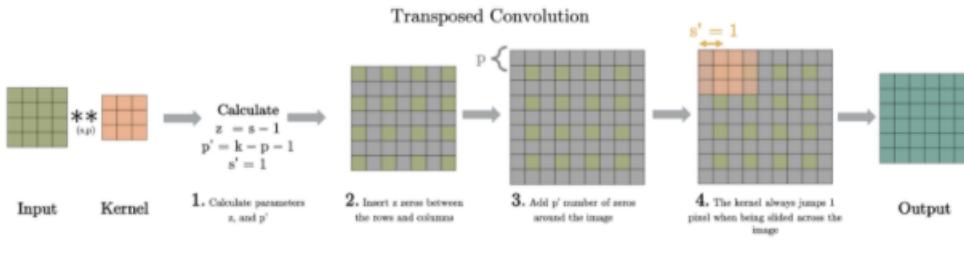


Illustration of a typical Convolution Layer (Source – Ref No 11)

Πηγή: <https://www.mygreatlearning.com/blog/fcn-fully-convolutional-network-segmentation/>

Σχήμα 2.25: Παράδειγμα λειτουργίας Transposed Convolution

2.6.2 Συνδέσεις Παράβλεψης - Skip Connections

Οι συνδέσεις παράβλεψης παρέχουν αρκετές πληροφορίες σε μεταγενέστερα επίπεδα για τη δημιουργία ακριβών ορίων τμηματοποίησης. Αυτός ο συνδυασμός λεπτών και χονδροειδών στρωμάτων οδηγεί σε τοπικές προβλέψεις με σχεδόν ακριβή συνολική (χωρική) δομή.

Η προσθήκη συνδέσεων skip μπορεί να θεωρηθεί ως μέθοδος Boosting για ένα FCN, το οποίο προσπαθεί να βελτιώσει την απόδοση των επιπέδων χρησιμοποιώντας προβλέψεις (χάρτες χαρακτηριστικών) από προηγούμενα επίπεδα.

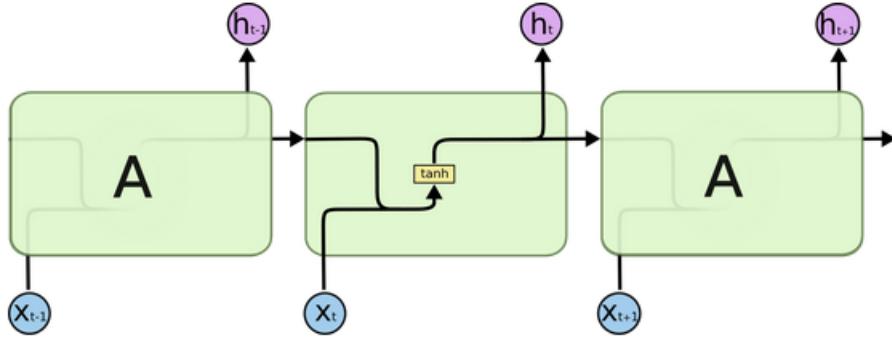
2.7 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNN)

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks) αποτελούν και αυτά μία κλάση των τεχνητών νευρωνικών δικτύων. Η ανάγκη για τη δημιουργία των (RNNs) προέκυψε από την αδυναμία των απλών νευρωνικών δικτύων να αξιοποιήσουν σωστά ακολουθιακά δεδομένα. Για παράδειγμα δεδομένα όπως τα καρέ σε ένα βίντεο ή οι λέξεις σε ένα κείμενο που έχουν ακολουθιακή φύση. Ο άνθρωπος για να επεξεργαστεί τέτοιου είδους δεδομένα χρησιμοποιεί την μνήμη του ώστε να διατηρήσει την χρονική πληροφορία που τα χαρακτηρίζει. Τα RNNs για να μπορέσουν να αξιοποιήσουν την ακολουθιακή πληροφορία των δεδομένων μιμούνται τον άνθρωπο και εισάγουν την έννοια την μνήμης.

Τα βασικά feed forward νευρωνικά δίκτυα έχουν και αυτά μνήμη, υπό την έννοια ότι ψυμούνται όσα έχουν μάθει κατά τη διαδικασία εκπαίδευσης. Τα RNNs σε επέκταση αυτών, καταφέρνουν να ψυμούνται εισόδους που χρησιμοποιήθηκαν ήδη για την παραγωγή εξόδων προηγουμένως. Έτσι, κάθε έξοδος του δικτύου δεν εξαρτάται μόνο από δεδομένα της εισόδου εκείνη την χρονική στιγμή και τους παραμέτρους του δικτύου αλλά και από ένα διάνυσμα 'χρυφής' κατάστασης (hidden

state) που αναπαριστά όλες τις παρελθοντικές εισόδους στο δίκτυο. Επομένως, ένα RNN για δύο περιπτώσεις με την ίδια είσοδο μπορεί να αποδώσει διαφορετικές εξόδους όταν τα προηγούμενα δεδομένα εισόδου είναι διαφορετικά σε κάθε περίπτωση. Η πρώτη παρουσίαση των επαναλαμβανόμενα νευρωνικών δικτύων και της έννοιας της εσωτερικής κατάστασης έγινε το 1982 από τον J. Hopfield με τα Hopfield Networks [30].

2.7.1 Τρόπος λειτουργίας



Πηγή: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
Σχήμα 2.26: Δομή νευρώνα RNN

Ο υπολογισμός της εξόδου σε ένα επαναλαμβανόμενα νευρωνικό δίκτυο δεν εξαρτάται μόνο από το διάνυσμα εισόδου αλλά και από το διάνυσμα εσωτερικής κατάστασης (hidden state), όπως αναφέρθηκε. Στην εικόνα 2.26 ο υπολογισμός της επόμενης κρυψής κατάστασης εξαρτάται από την είσοδο και την προηγούμενη κρυψή κατάσταση. Με τις σχέσεις 2.22 και 2.23 φαίνεται αναλυτικά ο τρόπος υπολογισμού της κρυψής κατάστασης h_t και της εξόδου y_t για είσοδο x_t . Στη βιβλιογραφία, οι διαστάσεις του διανύσματος h αναφέρονται ως hidden size και ουσιαστικά εκφράζει το μέγεθος της εσωτερικής μνήμη του κυττάρου (νευρώνα). Οι μεταβλητές W_h, U_h, W_y, b_h, b_y είναι διανύσματα παραμέτρων και οι διαστάσεις τους εξαρτώνται από το input size και το hidden size. Τέλος, τα σ_h, σ_y είναι συναρτήσεις ενεργοποίησης, στο παράδειγμα της εικόνας $\sigma_h = \tanh$.

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \quad (2.22)$$

$$y_t = \sigma_y(W_y h_t + b_y) \quad (2.23)$$

2.8 Μετρικές Αξιολόγησης

Η προετοιμασία των δεδομένων και εκπαίδευση είναι αντικείμενα υψίστης σημασίας για την ανάπτυξη ενός μοντέλου μηχανικής μάθησης, εξίσου σημαντική είναι όμως και η παρακολούθηση της επίδοσης του μοντέλου. Δηλαδή, πόσο καλά μπορεί να γενικεύει το μοντέλο για άγνωστα

δεδομένα. Οι μετρικές αξιολόγησης εξυπηρετούν αυτόν το σκοπό, συγκεκριμένα μετρούν κάποιο μέγεθος του εκπαιδευμένου μοντέλου ως προς κάποιο χαρακτηριστικό. Χωρίς την χρήση μετρικών αξιολόγησης η βελτίωση της προβλεπτικής ικανότητας του μοντέλου ή η σύγκρισή του με άλλα μοντέλα δεν θα ήταν εφικτή. Ακόμα, αξίζει να αναφερθεί ότι η χρήση των μετρικών αξιολόγησης δεν είναι καθολική αλλά η φύση του κάθισ προβλήματος ενθαρρύνει την χρήση διαφορετικών μετρικών αξιολόγησης. Στη συνέχεια παρουσιάζονται μερικές από τις σημαντικότερες μετρικές αξιολόγησης. Πρώτα, όμως, γίνεται αναφορά στις κλάσεις που μπορούν να ανήκουν οι προβλέψεις του συστήματος:

- *True Positive (TP):*

Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση είναι θετική.

- *True Negative (TN):*

Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση είναι αρνητική.

- *False Positive (FP):*

Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση είναι θετική.

- *False Negative (FN):*

Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση είναι αρνητική.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Σχήμα 2.27: Κλάσεις προβλέψεων

1. Accuracy

Η μετρική αξιολόγησης Accuracy ή ακρίβεια όπως φαίνεται και από τη σχέση 2.24, περιγράφει τον λόγο των σωστά ταξινομημένων δειγμάτων προς το σύνολο όλων των δειγμάτων.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.24)$$

Η μετρική Accuracy είναι η πιο σημαντική μετρική και δίνει μια άμεση και απλή αξιολόγηση του μοντέλου. Η χρήση της συνίσταται για δεδομένα που είναι καλά ισορροπημένα.

2. Precision

Η μετρική αξιολόγησης Precision (βλέπε σχέση 2.25) που εκφράζει τον λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των προβλέψεων που έχουν ταξινομηθεί ως θετικές.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.25)$$

Η μετρική Precision χρησιμοποιείται σε προβλήματα που η εγκυρότητα της πρόβλεψης είναι μεγάλης σημασίας.

3. Recall

Η μετρική αξιολόγησης Recall (βλέπε σχέση 2.26) που εκφράζει το λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των θετικών προβλέψεων.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.26)$$

Η μετρική Recall χρησιμοποιείται σε προβλήματα που έχουν ως σκοπό την μεγιστοποίηση των θετικών προβλέψεων.

4. F1 Score

Η μετρική αξιολόγησης F1 Score (βλέπε σχέση 2.27) που εκφράζει τον αρμονικό μέσο όρο των μετρικών Precision και Recall.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.27)$$

Η χρήση της μετρικής F1 Score γίνεται όταν το πρόβλημα απαιτεί καλό Precision και Recall.

5. Crossentropy or Binary Crossentropy

Η μετρική Crossentropy ή Log Loss λαμβάνει υπόψιν την αβεβαιότητα της πρόβλεψης βασιζόμενη στο πόσο διαφέρει από την πραγματική τιμή. Χρησιμοποιείται σε προβλήματα δυαδικής ταξινόμησης και υπολογίζεται από τον τύπο 2.28.

$$\text{Crossentropy} = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2.28)$$

Όπου p είναι η πιθανότητα η πρόβλεψη να είναι 1 και y είναι η πρόβλεψη του μοντέλου. Η μετρική αυτή χρησιμοποιείται όταν η έξοδος του μοντέλου είναι πιθανοτικές προβλέψεις.

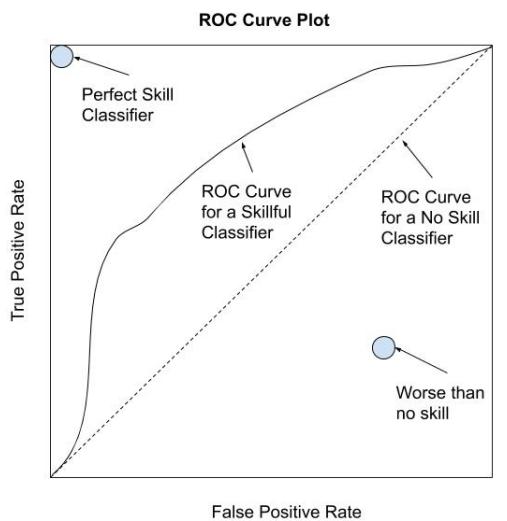
6. Categorical Crossentropy

Αυτή η μετρική αξιολόγησης είναι ίδια με την μετρική Crossentropy η μόνη διαφορά είναι ότι χρησιμοποιείται σε προβλήματα που οι κλάσεις ταξινόμησης είναι παραπάνω από δύο.

$$\text{CategoricalCrossentropy} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} * \log(p_{ij}) \quad (2.29)$$

Όπου το y_{ij} είναι 1 αν το δείγμα i ανήκει στην κλάση j , αλλιώς είναι 0 και p_{ij} είναι η πιθανότητα το μοντέλο να προβλέψει ότι το δείγμα i ανήκει στην κλάση j .

7. Καμπύλη ROC - AUC



Σχήμα 2.28: Ερμηνεία της καμπύλης ROC-AUC

Αν για λόγους απλότητας θεωρήσουμε ότι ένας δυαδικός ταξινομητής παράγει ως έξοδο την πιθανότητα ένα δείγμα να είναι θετικό p και συμπληρωματικά την πιθανότητα να είναι αρνητικό $1 - p$, τότε η κλάση στην οποία θα ταξινομηθεί το δείγμα καθορίζεται από την τιμή ενός κατωφλίου t :

- Αν $p > t$ το δείγμα θα ταξινομηθεί ως θετικό.
- Αν $p < t$ το δείγμα θα ταξινομηθεί ως αρνητικό.

Η ROC(Receiver Operating Characteristic) είναι μια καμπύλη πιθανότητας και η AUC αντιπροσωπεύει το βαθμό ή το μέτρο της διαχωρισιμότητας. Δείχνει πόσο το μοντέλο είναι ικανό να διαχρίνει μεταξύ των δύο κλάσεων. Η καμπύλη ROC απεικονίζει την απόδοση του δυαδικού ταξινομητή καθώς η τιμή του κατωφλίου μεταβάλλεται όπως φαίνεται στην εικόνα 2.28. Στον οριζόντιο άξονα χέχουμε το συμπλήρωμα του Specificity (False Positive Rate) ενώ στον κάθετο άξονα χέχουμε το Sensitivity (True Positive Rate). Κάθε σημείο της καμπύλης αναπαριστά το λόγο Sensitivity προς Specificity για μια συγκεκριμένη τιμή του κατωφλίου διαχωρισμού. Ένας ταξινομητής αποδίδει καλά όταν για μικρές τιμές του $1 -$

Specificity, το Sensitivity προσεγγίζει την μέγιστη τιμή 1. Για την ποσοτικοποίηση αυτής της ιδιότητας υπολογίζουμε το εμβαδόν "κάτω από την καμπύλη" (area under the curve, AUC). Προφανώς όσο αυξάνεται η τιμή του AUC τόσο καλύτερα προβλέπει το μοντέλο.

Κεφάλαιο 3

Μεθοδολογία

Έχουν εφαρμοστεί πολλαπλές προσεγγίσεις που βασίζονται στη βασιά μάθηση για την κατάτμηση των αιμοφόρων αγγείων του αμφιβληστροειδούς. Υπάρχουν δύο διαφορετικές μεθοδολογίες που χρησιμοποιούν τα CNN με διακριτούς τρόπους προκειμένου να τμηματοποιήσουν τα αγγεία. Αυτοί είναι οι ακρογωνιαίοι λίθοι της εργασίας που παρουσιάζεται σε αυτό το κεφάλαιο.

Στην επόμενη ενότητα, περιγράφονται οι διαφορετικές λύσεις, ξεκινώντας από τον μηχανισμό προεπεξεργασίας δεδομένων που χρησιμοποιήθηκε. Στη συνέχεια, παρέχεται αναλυτική περιγραφή των δικτύων που υλοποιήθηκαν.

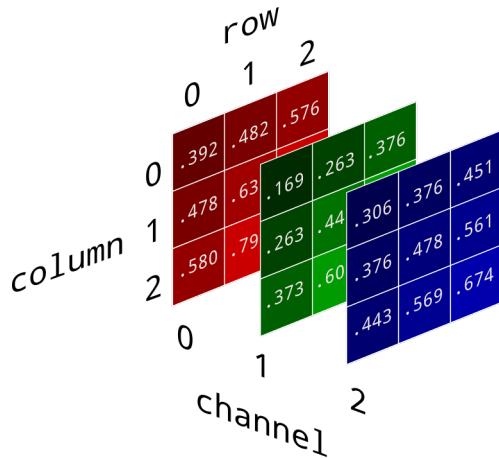
3.1 Επεξεργασία Δεδομένων

Η προεπεξεργασία δεδομένων είναι μια σημαντική εργασία. Είναι μια τεχνική εξόρυξης δεδομένων που μετατρέπει τα ακατέργαστα δεδομένα σε μια πιο κατανοητή, χρήσιμη και αποτελεσματική μορφή, που μπορεί να γίνει κατανοητή και να αναλυθεί από τους υπολογιστές και τη μηχανική μάθηση. Πολλές φορές, τα δεδομένα που θα χρησιμοποιηθούν για την ανάπτυξη ενός συστήματος μηχανικής μάθησης είναι από την φύση τους αριθμητικά και είναι εύκολο να χρησιμοποιηθούν, όπως για παράδειγμα οι τιμές των μετοχών στο χρηματιστήριο ή οι τιμές καρδιολογικών εξετάσεων ενός συνόλου ανθρώπων. Σε αυτή την περίπτωση τα δεδομένα δεν χρειάζονται ιδιαίτερη επεξεργασία για την τροφοδότησή τους στο σύστημα. Άλλες φορές όμως, υπάρχουν περιπτώσεις που τα δεδομένα δεν μπορούν να χρησιμοποιηθούν άμεσα στη μορφή που συναντώνται στον πραγματικό κόσμο, παραδείγματος χάρη το σύνολο των λέξεων ενός κειμένου ή οι κυματομορφές ήχου ή στην περίπτωσή μας οι εικόνες. Η αναπαράσταση αυτών των ακατέργαστων δεδομένων στον υπολογιστή μπορεί πάλι να γίνεται με 0 και 1, άλλα η πληροφορία που περιέχουν δεν μπορεί να γίνει άμεσα αντιληπτή από τον υπολογιστή ενώ μπορεί να περιέχει όχι μόνο σφάσματα και ασυνέπειες, αλλά να είναι συχνά ημιτελής και μη ομοιόμορφη. Σε τέτοιου είδους δεδομένα η επεξεργασία για την μετατροπή τους και την ενσωμάτωση στα συστήματα μηχανικής μάθησης είναι απαραίτητη για την επίτευξη των επιθυμητών αποτελεσμάτων.

Τα δεδομένα εισόδου παίζουν χρίσιμο ρόλο σε κάθε αλγόριθμο μηχανικής μάθησης, ακόμη

και στις προσεγγίσεις που βασίζονται σε βαθιά μάθηση. Θεωρητικά, η υψηλή χωρητικότητα αυτών των συστημάτων θα πρέπει να επιτρέπει εξίσου τη μάθηση από ακατέργαστα δεδομένα, ωστόσο, η εκπαιδευτική διαδικασία φαίνεται να επωφελείται από μια επαρκή προεπεξεργασία. Σε αυτή την ενότητα, παρουσιάζεται η προσέγγιση που έχει χρησιμοποιηθεί για τη δημιουργία προεπεξεργασμένων επιθεμάτων-patches εικόνων αμφιβληστροειδούς.

3.1.1 Επεξεργασία Εικόνων (Image Preprocessing)



Σχήμα 3.1: Αναπαράσταση Έγχρωμης Εικόνας ως Πίνακα

Προτού προχωρήσουμε ότι εξετάσουμε πως αναπαρίσταται μία εικόνα ως πίνακας με την χρήση του RGB μοντέλου. Μία έγχρωμη εικόνα με ανάλυση $H \times W$ (Height, Width) μπορεί να αναπαρασταθεί ως ένας τρισδιάστατος πίνακας διαστάσεων $H \times W \times 3$ τον οποίο μπορούμε να φανταστούμε ως τρείς δισδιάστατους $H \times W$ πίνακες, έναν για το κόκκινο κανάλι (R), έναν για το πράσινο (G), και έναν για το μπλε (B). Κάθε στοιχείο αυτών των δισδιάστατων πινάκων λαμβάνει μία τιμή στο εύρος $[0, 255]$ που εκφράζει την ένταση του συγκεκριμένου χρώματος για το συγκεκριμένο εικονοστοιχείο (εικόνα 3.1).

Παρακάτω ότι δούμε μια σειρά από τεχνικές - βήματα στην προεπεξεργασία εικόνων που χρησιμοποιήθηκαν στη μελέτη μας.

3.1.2 Μετατροπή Χρώματος Εικόνας

Μπορούν να χρησιμοποιηθούν διάφορες προσεγγίσεις για την προεπεξεργασία ιατρικών εικόνων. Για παράδειγμα, ορισμένα έργα χρησιμοποιούν τα τρία έγχρωμα κανάλια και άλλα βασίζονται μόνο στην πληροφορία έντασης του πράσινου καναλιού, καθώς είναι το κανάλι που εμφανίζει υψηλότερη αντίθεση μεταξύ των αγγείων και του φόντου. Τα κόκκινα και μπλε κανάλια παρουσιάζουν χαμηλή αντίθεση και τυχαία διακύμανση της φωτεινότητας.

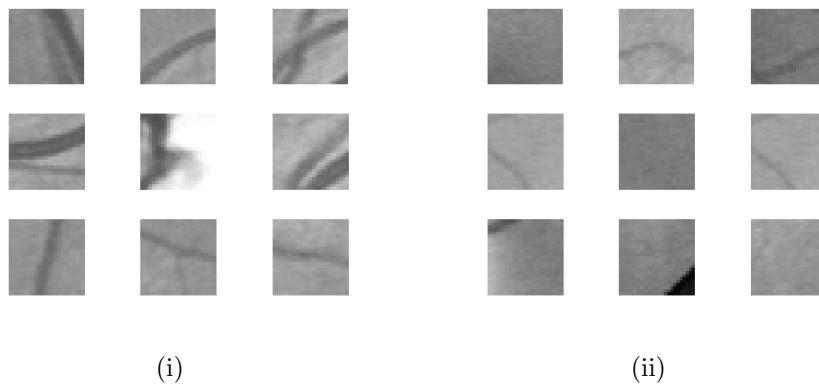
Σε αυτή τη μελέτη, χρησιμοποιήθηκαν στις δύο προσεγγίσεις μας, και οι δύο παραπάνω τεχνικές προεπεξεργασίας της εικόνας. Η δεύτερη τεχνική προεπεξεργασίας εικόνας που χρησιμοποιήθηκε σε αυτή τη μελέτη ξεκινά με τις πληροφορίες που εξάγονται από το πράσινο κανάλι

της αρχικής εικόνας RGB και στη συνέχεια εφαρμόστηκε μια κανονικοποίηση σε κλίμακα του γκρι λόγω του γεγονότος ότι ως αποτέλεσμα της διαδικασίας λήψης, πολύ συχνά οι εικόνες βυθού είναι ανομοιόμορφα φωτισμένες και παρουσιάζουν τοπική φωτεινότητα και μεταβλητότητα αντίθεσης.

3.1.3 Διαχωρισμός σε patches

Η εξαγωγή μικρών μπαλωμάτων (patches) από τις εικόνες αμφιβληστροειδούς επιτρέπει τη χρήση στρατηγικών βαθιάς μάθησης για την τυμηματοποίηση του αγγειακού συστήματος, καθώς οι δημόσιες βάσεις δεδομένων συνήθως δεν περιέχουν περισσότερες από μερικές δεκάδες εικόνες. Στη συνέχεια, είναι δυνατό να πλαισιώσουμε το πρόβλημα της κατάτμησης του αγγείου με δύο τρόπους: ταξινόμηση κάθε patch, όπου το δίκτυο εξάγει την πιθανότητα το κεντρικό εικονοστοιχείο να ανήκει σε ένα αγγείο [31], και τυμηματοποίηση κάθε patch, όπου το δίκτυο εξάγει έναν χάρτη τέτοιων πιθανοτήτων, ένα για κάθε pixel του patch.

Στη στρατηγική της εκπαίδευσης ανά patch (patch-wise training), η εικόνα αναλύεται σε πολλαπλά patches που μπορεί να είναι είτε επικαλυπτόμενα είτε δημιουργούμενα με τυχαίο διαχωρισμό. Ο τυχαίος διαχωρισμός μπορεί να έχει ως αποτέλεσμα μεγαλύτερη διακύμανση μεταξύ των patches και καλύτερη σύγκλιση. Ωστόσο, αυτός ο τυχαίος διαχωρισμός έχει το πρόβλημα της ανισορροπίας των κλάσεων καθώς και δίνει χαμηλότερη ακρίβεια από την επιλογή patches με επικαλυπτόμενο τρόπο. [32]



Σχήμα 3.2: Παράδειγμα ενός θετικού συνόλου (α) και ενός αρνητικού συνόλου από patches (β)

Στην πρώτη μέθοδο της εργασίας αυτής, που βασίζεται στο δίκτυο των Liskowski et al. [31], που από εδώ και στο εξής θα ονομάζουμε Lisk-Net για ευκολία, πραγματοποιήθηκε η εξαγωγή και λήψη πολυάριθμων patches από τις εικόνες. Όσον αφορά τη δειγματοληψία patches από τις εικόνες, τα εξεταζόμενα πειράματα κυμαίνονταν μεταξύ της τυχαίας δειγματοληψίας από το οπτικό πεδίο, της εξαγωγής ενός ισορροπημένου συνόλου patches, σε σχέση με την ετικέτα του κεντρικού εικονοστοιχείου, και της εξαγωγής ενός συνόλου patches σύμφωνα με κάποια αυθαίρετη αναλογία $r \in [0, 1]$ θετικών και αρνητικών δειγμάτων. Παραδείγματα θετικών και αρνητικών patches, σύμφωνα με την οπτική μιας προσέγγισης που βασίζεται σε ταξινόμηση,

παρουσιάζονται στο Σχήμα 3.2 Το μέγεθος του patch, ο αριθμός τους και η στρατηγική της δειγματοληψίας όπως αναφέρονται λεπτομερώς με την περιγραφή κάθε υλοποίησης.

Η διαδικασία εκπαίδευσης ξεκινά με την εξαγωγή patches M-επί-M (το M ορίστηκε στο 27). Στη συνέχεια, πραγματοποιείται επεξεργασία εικόνας, μέσω κανονικοποίησης της συνολικής αντίθεσης και λεύκανσης μηδενικής φάσης. Για κάθε εικόνα βυθιστού εκπαίδευσης, δημιουργήθηκαν τυχαία 20.000 patches σαν αυτή, έτσι ώστε το εκπαιδευτικό σετ έχει 400.000 patches στο DRIVE.

3.1.4 Κανονικοποίηση (Standardization)

Χρησιμοποιώντας τις μελέτες των LeCun et al. [33] και ο Deng Li [34] σχετικά με την απόδοση και τη βελτίωση των αλγορίθμων που βασίζονται στη διάδοση και την κατάβαση κλίσης, εφαρμόστηκαν δύο μετασχηματισμοί δεδομένων για την ενίσχυση της απόδοσης της εκπαίδευσης: κανονικοποίηση δεδομένων και τυποποίηση δεδομένων.

Η κανονικοποίηση ή normalization είναι τόσο απλή όσο η λήψη των αναλογιών (αφοιβαία κανονικοποίηση), ο υπολογισμός των διαφορών (κανονικοποίηση εύρους) και η πολλαπλασιαστική κανονικοποίηση ή η κανονικοποίηση του άξονα Z. Η κανονικοποίηση διασφαλίζει ότι τα δεδομένα κατανέμονται κατά προσέγγιση ομοιόμορφα μεταξύ των εισόδων και των εξόδων του δικτύου. Στην περίπτωση που παρουσιάστηκε, ελήφθη υπόψη ο απλός μετασχηματισμός που δίνεται από την εξίσωση 3.1, ανάγοντας την κλίμακα των δεδομένων στο εύρος [0,1].

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

Η τυποποίηση ή standardization είναι η πιο συχνά χρησιμοποιούμενη τεχνική 3.2, μετασχηματίζοντας τα δεδομένα έτσι ώστε ο αριθμητικός τους μέσος όρος να είναι μηδέν και η τυπική απόκλιση να είναι ένα. Ωστόσο, τόσο η μέση όσο και η τυπική απόκλιση είναι ευαίσθητα σε ακραίες τιμές και αυτή η τεχνική δεν εγγυάται ένα κοινό αριθμητικό εύρος για τις κανονικοποιημένες τιμές.

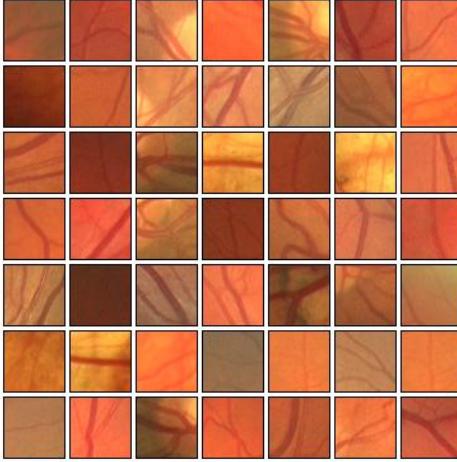
$$x' = \frac{x - \bar{x}}{\sigma} \quad (3.2)$$

3.1.5 Μετασχηματισμοί Προεπεξεργασίας των Εικόνων Εισόδου

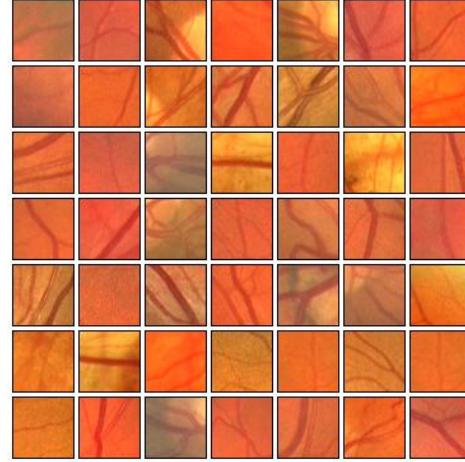
3.1.5.1 Κανονικοποίηση Ολικής Αντίθεσης - Global Contrast Normalization (GCN)

Παρατηρώντας τις εικόνες αμφιβληστροειδούς της βάσης δεδομένων DRIVE, υποδεικνύεται ξεκάθαρα ότι η φωτεινότητα της εικόνας μπορεί να διαφέρει διαμέσω του Field of View (FOV). Προκειμένου να βοηθήσουμε τη διαδικασία μάθησης να αφαιρέσει αυτές τις διαχυμάνσεις και να επικεντρωθεί στην ανίχνευση των αγγείων, εκτελούμε τοπική (ανά patches) κανονικοποίηση φωτεινότητας και αντίθεσης. Κάθε patch κανονικοποιείται (standardization) αφαιρώντας το μέσο όρο και διαιρώντας με την τυπική απόκλιση των στοιχείων του (ανεξάρτητα στο κανάλι R, G και

B), που αντιστοιχίζει επίσης την αρχικά κωδικοποιημένη σε byte φωτεινότητα σε προσημασμένους πραγματικούς αριθμούς. Το Σχήμα 3.4 παρουσιάζει το αποτέλεσμα αυτής της επεξεργασίας.



Σχήμα 3.3: Παράδειγμα συνόλου patches εκπαίδευσης, 27×27 όπως εξήχθησαν από τη βάση DRIVE



Σχήμα 3.4: Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή GCN

3.1.5.2 Λεύκανση Μηδενικής Φάσης - Zero-phase Component Analysis (ZCA Whitening)

Ανάλυση συστατικών μηδενικής φάσης (ZCA Whitening): Σε φυσικές εικόνες, τα γειτονικά pixels εμφανίζουν έντονη συσχέτιση καθώς είναι πιθανό να αντιπροσωπεύουν την ίδια δομή σε μια σκηνή. Επομένως, κατά την εκμάθηση ενός στατιστικού μοντέλου εικόνων, είναι επιθυμητό να αφαιρέσουμε από αυτά τα καθολικά χαρακτηριστικά και να εστιάσουμε στους συσχετισμούς υψηλότερης τάξης. Η κατάργηση των καθολικών συσχετίσεων μπορεί να επιτευχθεί πολλαπλασιάζοντας τον πίνακα δεδομένων με έναν πίνακα λεύκανσης (whitening matrix). Μετά από αυτόν τον συνηθισμένο μετασχηματισμό, καθίσταται αδύνατο να προβλεφθεί η τιμή ενός pixel δεδομένης της τιμής μόνο ενός άλλου pixel.

Η καθοριστική ιδιότητα του μετασχηματισμού ZCA (ή μετασχηματισμός Mahalanobis) είναι ότι οδηγεί σε δεδομένα που έχουν υποστεί λεύκανση και είναι όσο το δυνατόν πιο κοντά στα αρχικά δεδομένα (με την έννοια των ελαχίστων τετραγώνων). Με άλλα λόγια, εάν η σχέση 3.3:

$$\|X - XY^T\|^2 \quad (3.3)$$

πρέπει να ελαχιστοποιηθεί με την προϋπόθεση ότι το XY^T έχει υποστεί λεύκανση, τότε η λύση είναι να ορίσουμε $Y = W_{ZCA}$.

Η μέθοδος υπολογισμού και εφαρμογής του μετασχηματισμού στα δεδομένα περιγράφεται στα παρακάτω βήματα:

- Τα δεδομένα που ανήκουν στο σύνολο εκπαίδευσης και πρόκειται να δωθούν ως είσοδος στο μοντέλο για την διαδικασία εκμάθησης, είναι ήδη κεντραρισμένα (μηδενική μέση τιμή)

και αποθηκεύονται στον πίνακα X : κάθε γραμμή του πίνακα είναι ένα σημείο δεδομένων (ο αριθμός των γραμμών είναι ίσος με τον αριθμό των patches εκπαίδευσης) και κάθε στήλη αντιστοιχεί σε ένα χαρακτηριστικό (ο αριθμός στηλών είναι ίσος με $3 \times 27 \times 27 = 2187$).

- Ο πίνακας σκέδασης (covariance matrix) Σ ορίζεται ως: $\Sigma = \frac{1}{m} \sum_{i=1}^m XX^T$ και $\Sigma = U\Lambda U^T$ όπου U είναι ο πίνακας με στήλες του τα ιδιοδιανύσματα του Σ και Λ ο πίνακας με διαγώνιο τις ιδιοτιμές του. Με βάση τον υπολογισμό αυτών των πινάκων, ο πίνακας του μετασχηματισμού ZCA δίνεται ως εξής:

$$W_{ZCA} = U\Lambda^{-\frac{1}{2}}U^T = \Sigma^{-\frac{1}{2}} \quad (3.4)$$

- Ο μετασχηματισμός αυτός ‘μαθαίνεται’ μόνο από τα δεδομένα εκπαίδευσης (σχέση 3.3) και τελικά εφαρμόζεται και στο σύνολο ελέγχου.

3.1.5.3 Εξισορρόπηση ιστογράμματος - Contrast Limited Adaptive Histogram Equalization (CLAHE)

Το CLAHE είναι μια παραλλαγή της προσαρμοστικής ισοστάθμισης ιστογράμματος που φροντίζει για την υπερβολική ενίσχυση της αντίθεσης. Το CLAHE λειτουργεί σε μικρές περιοχές της εικόνας, πλακίδια, και όχι σε ολόκληρη την εικόνα. Στη συνέχεια, τα γειτονικά πλακίδια συνδυάζονται χρησιμοποιώντας διγραμμική παρεμβολή (binary interpolation) για την αφαίρεση των τεχνητών ορίων. Στην εφαρμογή του μετασχηματισμού αυτού στη μελέτη μας, ορίστηκε ως όριο για τον περιορισμό της αντίθεσης η τιμή 2.0

3.1.5.4 Top-Hat Μετασχηματισμός Πολλαπλής Κλίμακας Multi-Scale Top-Hat Transform (MTHT)

Με τη χρήση μετασχηματισμού ανοίγματος (opening) και κλεισίματος (closing), ο μετασχηματισμός top-hat, συμπεριλαμβανομένου του μετασχηματισμού white top-hat και του μετασχηματισμού black top-hat, που συμβολίζονται με WTH και BTH, ορίζονται ως εξής, αντίστοιχα:

$$WTH(x, y) = f(x, y) - f \circ B(x, y) \quad (3.5)$$

$$BTH(x, y) = f \bullet B(x, y) - f(x, y). \quad (3.6)$$

Το opening εξομαλύνει τις λευκές περιοχές της εικόνας που αντιστοιχούν στο μέγεθος του χρησιμοποιούμενου δομικού στοιχείου B. Έτσι, το WTH χρησιμοποιείται συνήθως για την εξαγωγή λευκών περιοχών της εικόνας. Ομοίως, το closing εξομαλύνει τις μαύρες περιοχές της εικόνας που αντιστοιχούν στο μέγεθος του χρησιμοποιούμενου δομικού στοιχείου B. Έτσι, το BTH χρησιμοποιείται συνήθως για την εξαγωγή μαύρων περιοχών της εικόνας. Μια βασική ιδέα για τη βελτίωση της εικόνας είναι η αύξηση της αντίθεσης μεταξύ των λευκών και μαύρων περιοχών της εικόνας. Ένας τρόπος να γίνει αυτό με βάση τον top-hat μετασχηματισμό είναι η προσθήκη των περιοχών λευκής εικόνας και η αφαίρεση των περιοχών μαύρης εικόνας από την αρχική εικόνα.

Προκειμένου η αρχική εικόνα να βελτιωθεί περισσότερο αποτελεσματικά, είναι απαραίτητο να εξάγουμε τα λευκά και μαύρα χαρακτηριστικά της εικόνας χρησιμοποιώντας πολλαπλές κλίμακες. Έτσι, χρησιμοποιήθηκε ένα εύρος από τιμές για το μέγεθος του δομικού στοιχείου, δίσκου του μετασχηματισμού MHT και ακολούθησε κανονικοποίηση του αποτελέσματος εξαρτώμενη από την κλίμακα.

3.1.6 Επαύξηση Δεδομένων (Data Augmentation)

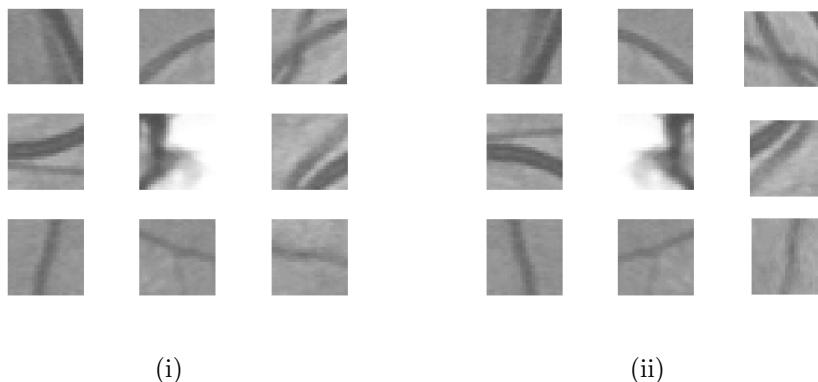
Η προσαύξηση δεδομένων είναι μια από τις πιο απλές μεθόδους βελτίωσης της ικανότητας γενίκευσης των τεχνητών νευρωνικών δικτύων. Η μέθοδος αυτή χρησιμοποιεί ως βάση μια εικόνα του συνόλου δεδομένων και με διάφορες τεχνικές όπως η περιστροφή, η περικοπή κ.ά. συνθέτει νέες εικόνες [35]. Η προσθήκη νεων δειγμάτων στο σύνολο δεδομένων εκπαίδευσης, ειδικά όταν αυτό είναι περιορισμένο, συμβάλλει στην αποφυγή της υπερπροσαρμογής του μοντέλου.

Ένα πράγμα που πρέπει να ληφθεί υπόψη κατά την εκπαίδευση ενός CNN είναι το γεγονός ότι μπορεί να προκύψει υπερπροσαρμογή λόγω του σχετικά μικρού αριθμού δειγμάτων που χρησιμοποιούνται ως είσοδοι σε σύγχριση με τον αριθμό των παραμέτρων του μοντέλου. Για να αυξήσουμε την ποικιλομορφία του συνόλου δεδομένων και, συνεπώς, να μειώσουμε τον κίνδυνο υπερβολικής προσαρμογής των CNNs, εφαρμόσαμε μεθόδους αύξησης δεδομένων. Μια στρατηγική που χρησιμοποιείται για να ξεπεράσει αυτό είναι η εφαρμογή τυχαιοποιημένων μετασχηματισμών στα υπάρχοντα δεδομένα εκπαίδευσης προκειμένου να δημιουργηθεί ουσιαστικά ένα νεότερο και μεγαλύτερο σύνολο εισόδου.

Στη πρώτη μεθοδολογία ταξινόμησης που χρησιμοποιήσαμε, τα δεδομένα επαυξήθηκαν εκτός σύνδεσης, πριν από την εκμάθηση. Κάθε patch, μετά την διαδικασία κανονικοποίησης και λεύκανσης που προηγήθηκε σύμφωνα με την προηγούμενη ενότητα, υποβλήθηκε σε 10 ανεξάρτητους μετασχηματισμούς, ο καθένας αποτελούμενος από τέσσερεις διαφορετικούς τύπους μετασχηματισμών που συνδυάστηκαν τυχαία προκειμένου να παραχθούν δέκα νέες εικόνες από κάθε αρχική (ορισμένα παραδείγματα φαίνονται στο Σχήμα ;). Οι τέσσερις τυχαιοποιημένες ενέργειες περιγράφινται στην ακόλουθη λίστα:

- Κλιμάκωση με συντελεστή μεταξύ 0,7 και 1,2
- Τυχαία περιστροφή κατά μια γωνία μεταξύ [-90, 90]
- Αναστροφή οριζόντια ή κάθετα
- Διόρθωση γάμμα κορεσμού και τιμής (του HSV colorspace) αυξάνοντας τα pixel στην ισχύ σε [0,25, 4].

Τα εύρη των παραμέτρων είχαν προσδιοριστεί σε ένα προκαταρκτικό πείραμα. Οι τιμές των παραμέτρων αντλούνται από αυτά τα εύρη ομοιόμορφα. Εφαρμόζουμε επαυξήσεις σε θραύσματα εικόνας (patches) μεγέθους $2m \times 2m = 49 \times 49$ για να διασφαλίσουμε ότι η κλιμάκωση και η περιστροφή μπορούν να εκτελεστούν χωρίς σφάλματα. Μόλις εφαρμοστούν οι παραπάνω μετασχηματισμοί, τυχόν πλεονάζοντα εικονοστοιχεία απορρίπτονται έτσι ώστε τα προκύπτοντα



Σχήμα 3.5: Αρχικά patches εικόνας (α) και αντίστοιχα αποτελέσματα μετασχηματισμού αύξησης δεδομένων (β)

μπαλώματα να έχουν μέγεθος $m \times m = 27 \times 27$, δηλ. τα ίδια με τα αρχικά. Κάθε patch που προκύπτει λαμβάνει την ίδια ετικέτα κλάσης με το αρχικό patch. Με αυτόν τον τρόπο, αυξάνουμε τον αριθμό των patch εκπαίδευσης κατά 11 φορές. Αυτή η προεπεξεργασία δεν εφαρμόζεται στα patches του συνόλου δοκιμής.

Στη δεύτερη μεθοδολογία τμηματοποίησης που χρησιμοποιήσαμε, προκειμένου να ξεπεραστεί η υπερβολική προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης (overfitting), εφαρμόσαμε λειτουργίες περιστροφής, μετατόπισης και κατοπτρισμού στις εικόνες εισόδου για αύξηση δεδομένων. Η περιστροφή εφαρμόζεται σε σχέση με το κέντρο της εικόνας σε εύρος $-45^\circ < \theta < 45^\circ$. Επειτα, η μετατόπιση εφαρμόζεται τυχαία στο διάστημα που ορίζεται στην 3.7:

$$t_x < w - 448, t_y < h - 448 \quad (3.7)$$

όπου τα t_x και t_y αντιπροσωπεύουν τη μετατόπιση σε κατευθύνσεις x και y , αντίστοιχα και w και h είναι το πλάτος και το ύψος της εικόνας. Αφού μία τυχαία εικόνα έχει δημιουργηθεί, την επαυξάνουμε κατά 8 με συγκεκριμένες περιστροφές ($90^\circ, 90^\circ$) και λειτουργίες αναστροφής. Δημιουργούμε τυχαία 10.000 εικόνες χρησιμοποιώντας αυτήν την τεχνική.

3.2 Αρχιτεκτονικές Νευρωνικών Δικτύων για την Τυματοποίηση των Αγγείων του Αμφιβληστροειδούς

Η patch-wise ταξινόμηση στοχεύει στην αντιστοίχιση μιας κλάσης στο κεντρικό εικονοστοιχείο χρησιμοποιώντας τις πληροφορίες από ολόκληρο το patch ενώ η ταξινόμηση pixels-to-pixels στοχεύει στην ταξινόμηση κάθε εικονοστοιχείου από ολόκληρο το patch εικόνας.

3.2.1 Προσέγγιση Ταξιμόμησης - Patch-wise Classification

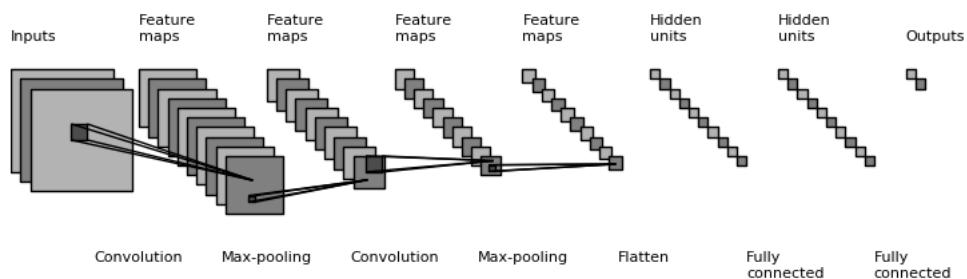
Σε μια προσέγγιση ταξινόμησης των patches, η ιδέα είναι να εφαρμοστεί ένα CNN σε ένα σύνολο patches που εξάγονται από την εικόνα και σε καθένα από αυτά να προβλέψουμε την



Σχήμα 3.6: Οι διαφορές ανάμεσα στο patch-wise classification και στο pixel-wise classification.[36]

ετικέτα ακλάσης του κεντρικού εικονοστοιχείου. Δίκτυα όπως το LeNet [37] και το AlexNet [38] είναι παραδείγματα αυτού του σχεδιασμού, όπου οι είσοδοι σταθερού μεγέθους παράγουν μη χωρικές εξόδους. Στο σχήμα 3.7, παρουσιάζεται μια τυπική αρχιτεκτονική για ταξινόμηση. Είναι σύνηθες να υπάρχουν εναλλασσόμενα βήματα συνέλιξης και λειτουργιών υποδειγματοληψίας (pooling) προκειμένου να καταγράφονται όλο και πιο υψηλότερου επιπέδου χαρακτηριστικά που είναι κατάλληλα για τελικές αποφάσεις. Οι χάρτες χαρακτηριστικών υποβάλλονται σε ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα, από τα οποία εξάγεται η έξοδος. Οι προτεινόμενοι σχεδιασμοί συνήθως διαφέρουν ως προς τον αριθμό των convolutional, pooling, και fully connected επιπέδων, τον αριθμό και το μέγεθος των πυρήνων και τη συμπεριληφθή των επιπέδων dropout και κανονικοποίησης της παρτίδας (batch normalization).

Η πρώτη μεθοδολογία, που χρησιμοποιήθηκε από τους Liskowski και Krawiec [31], Melinscak et al. [39] και Fu et al. [40] είναι μια ταξινόμηση βάσει στοιχείων όπου το CNN εκπαιδεύεται patch προς patch και η έξοδος είναι η πιθανότητα το κεντρικό εικονοστοιχείο (ή ένα μπλοκ κεντρικών εικονοστοιχείων στην περίπτωση της δομημένης πρόβλεψης) να βρίσκεται σε μια συγκεκριμένη κατηγορία (σε αυτήν την περίπτωση μια ταξινόμηση δύο κατηγοριών, καθώς κάθε εικονοστοιχείο μπορεί να είναι αγγείο ή μη αγγείο). Αυτές οι προσεγγίσεις χρησιμοποιούν έναν μεγάλο αριθμό M μικρών τετραγώνων $N \times N$ patches από ακατέργαστο RGB ή προεπεξεργασμένο σύνολο εικόνων εκπαίδευσης ως είσοδο του νευρωνικού δικτύου. Το δίκτυο εκπαιδεύεται και επαληθεύεται με τα εν λόγω δεδομένα και στη συνέχεια ελέγχεται με patches από το σύνολο δεδομένων δοκιμής.



Σχήμα 3.7: Παράδειγμα Αρχιτεκτονικής Συνελικτικού Νευρωνικού Δικτύου

Σε αυτή τη διπλωματική εργασία εφαρμόζεται η προσέγγιση ταξινόμησης σε patches, η οπο-

ία βασίζεται στη δουλειά των Liskowski and Krawiec [31] και χρησιμοποιείται ως baseline για αυτού του είδους τις ταξινομήσεις, καθώς η αρχιτεκτονική του δικτύου και η συνολική διαδικασία που ακολουθεί είναι η πιο ευρέως χρησιμοποιούμενη για αυτό το σκοπό. Αυτή η μέθοδος χρησιμοποιεί ένα βαθύ νευρωνικό δίκτυο (CNN) εκπαιδευμένο σε ένα μεγάλο δείγμα προεπεξεργασμένων δειγμάτων, ώστε να βελτιωθεί η συνολική απόδοση του pipeline. Οι αρχιτεκτονικές που παρουσιάζονται στον πίνακα 3.1 διαφέρουν μόνο ως προς τη χωρική συγκέντρωση. Η διαδικασία Max-Pooling εκτελείται σε ένα παράθυρο 2×2 pixels με βήμα 2 και χρησιμοποιείται μόνο στην υλοποίηση του νευρωνικού Plain. Η υλοποίηση νευρωνικού No-POOL δεν χρησιμοποιεί Max-pooling, καθώς τα πρόσφατα αποτελέσματα έδειξαν ότι τα δίκτυα που δεν κάνουν περαιτέρω δειγματοληψία των χαρακτηριστικών ενδέχεται να έχουν καλύτερη απόδοση όταν εφαρμόζονται σε μικρές εικόνες.

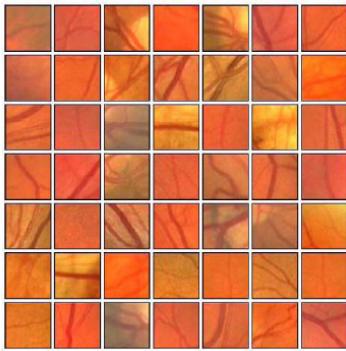
Implementation	Layer	Maps	Kernel size	Stride	Padding
Plain	Convolutional	64	4 x 4	1	0
	Convolutional	64	3 x 3	1	1
	Max-Pooling	-	2 x 2	2	0
	Convolutional	128	3 x 3	1	1
	Convolutional	128	3 x 3	1	1
	Max-Pooling	-	2 x 2	2	0
	Fully Connected	512	-	-	-
	Fully Connected	512	-	-	-
	Fully Connected	2	-	-	-
No-POOL	Convolutional	64	3 x 3	1	1
	Convolutional	64	3 x 3	1	1
	Convolutional	128	3 x 3	1	1
	Convolutional	128	3 x 3	1	1
	Fully Connected	512	-	-	-
	Fully Connected	512	-	-	-
	Fully Connected	2	-	-	-

Πίνακας 3.1: Οι δύο βασικές αρχιτεκτονικές που προτάθηκαν από τους Liskowski and Krawiec [31]

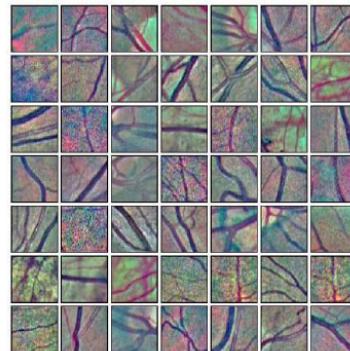
Εκτός από αυτές τις υλοποίησεις, οι συγγραφείς εξέτασαν άλλες τέσσερις, όλες χρησιμοποιώντας την αρχιτεκτονική δικτύου Plain: GCN, ZCA, AUGMENTED και BALANCED. Όλα αποτελούνται από μια στοίβα συνελικτικών επιπέδων που ακολουθούνται από τρία πλήρως συνδεδεμένα επίπεδα (Πίνακας 3.1). Ο κύριος στόχος αυτών των υλοποιήσεων ήταν η ανάλυση της απόδοσης της αρχιτεκτονικής Plain όταν εκπαιδεύονταν σε δεδομένα προεπεξεργασμένα με τις μεθόδους που περιγράφηκαν προηγουμένως (χανονικοποίηση ολικής αντίθεσης, λεύκανση μηδενικής φάσης και επαυξήσεις εικόνας) και ισορροπημένα δεδομένα (με ίση αναλογία κατηγοριών απόφασης). Ένα παράδειγμα των δεδομένων που χρησιμοποιούνται στη διαδικασία εκπαίδευσης είναι ορατό στα σχήματα 3.8, 3.9, 3.10.

Αν και με μια πρώτη ματιά αυτού του τύπου η αρχιτεκτονική φαίνεται να είναι αρκετά καλή για να πραγματοποιήσει τη Σημασιολογική Τμηματοποίηση, παρόλαυτά, έχει ορισμένα σημαντικά μειονεκτήματα:

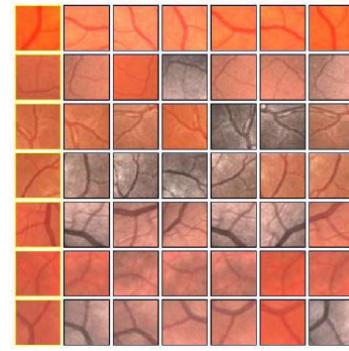
- Αυτή η προσέγγιση περιλαμβάνει πρόβλεψη σε επίπεδο μεμονωμένου εικονοστοιχείου, α-



Σχήμα 3.8: Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή GCN



Σχήμα 3.9: Παράδειγμα συνόλου patches εκπαίδευσης μετά τη μετατροπή ZCA



Σχήμα 3.10: Παράδειγμα συνόλου patches εκπαίδευσης μετά τη διαδικασία επαύξησης

παιτώντας έτσι ένα πυκνό στρώμα (dense) με έναν τεράστιο αριθμό παραμέτρων που πρέπει να μάθουμε, καιθιστώντας το εξαιρετικά δαπανηρό υπολογιστικά.

- Δεν επαναχρησιμοποιούνται χοινά χαρακτηριστικά μεταξύ επικαλυπτόμενων γειτονικών patches, επομένως είναι εξαιρετικά αναποτελεσματική.

3.2.2 Προσέγγιση Τμηματοποίησης - Patch Segmentation or Pixel-wise Classification

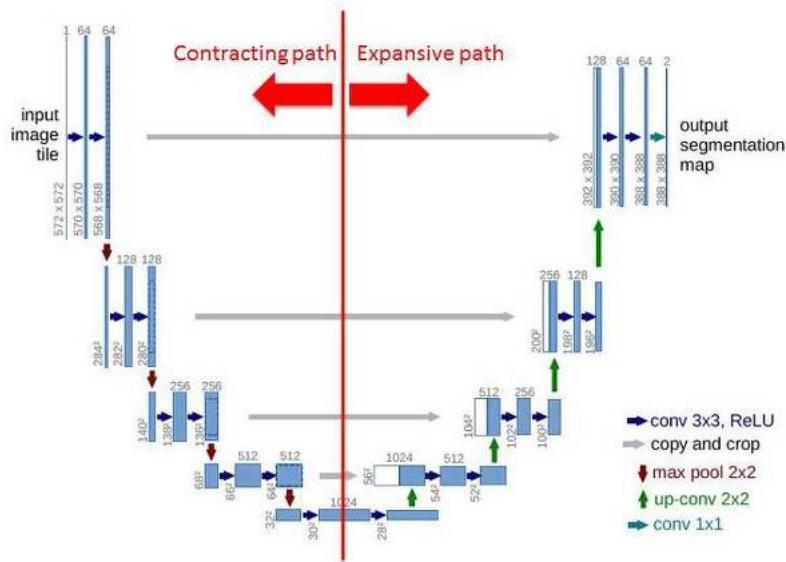
Ένας τρόπος για να αντιμετωπιστούν τα μειονεκτήματα της προηγούμενης αρχιτεκτονικής είναι η στοίβαξη ενός αριθμού επιπέδων συνέλιξης (Convolutional) με παρόμοιο padding για τη διατήρηση της διάστασης και την παραγωγή ενός τελικού χάρτη τμηματοποίησης. Αυτό σημαίνει ότι το μοντέλο θα μάθει την αντιστοιχίση από την εικόνα εισόδου στον αντίστοιχο χάρτη τμηματοποίησης μέσω του διαδοχικού μετασχηματισμού των αντιστοιχίσεων χαρακτηριστικών. Εντούτοις, υπάρχει ένα σημαντικό ζήτημα και με αυτήν την αρχιτεκτονική. Χρησιμοποιούμε το ίδιο padding σε όλα τα επίπεδα Convolution επειδή θα θέλαμε η εικόνα εξόδου να έχει την ίδια διάσταση με την εικόνα εισόδου. Άλλα αυτή η διατήρηση της πλήρους ανάλυσης γίνεται αρκετά υπολογιστικά ακριβή. Θα μπορούσαμε να επιλέξουμε μικρότερο αριθμό επιπέδων, αλλά αυτό θα έβλαπτε την απόδοση κατά πολύ. Δεν αντιμετωπίζουμε αυτό το δίλημμα σε μια εργασία ταξινόμησης επειδή για αυτήν την εργασία μας ενδιαφέρει μόνο η παρουσία ενός μεμονωμένου αντικειμένου ενδιαφέροντος και η απώλεια των πληροφοριών σχετικά με τη θέση του εν λόγω αντικειμένου είναι αβλαβής. Επομένως, μπορούμε περιοδικά να κάνουμε downsampling των εικόνων μέσω συγκέντρωσης. Ωστόσο, αυτό δεν συμβαίνει στο Semantic Segmentation. Εάν θέλετε να τμηματοποιήσετε μια εικόνα του αυτοκινητόδρομου για την εκπαίδευση ενός αυτόνομου οχήματος, δεν θα θέλατε ορισμένα στοιχεία ή οχήματα να περικοπούν από την τελική έξοδο που τροφοδοτείται στο μηχάνημα.

Μια δημοφιλής λύση στο πρόβλημα που αντιμετωπίζει η προηγούμενη Αρχιτεκτονική είναι η χρήση Downsampling και Upsampling σε ένα πλήρως συνελικτικό δίκτυο FCN (2.6). Στο

πρώτο μισό του μοντέλου, μειώνουμε τη χωρική ανάλυση της εικόνας αναπτύσσοντας σύνθετες αντιστοιχίσεις χαρακτηριστικών. Με κάθε συνέλιξη, καταγράφουμε πιο λεπτομερείς πληροφορίες της εικόνας. Σε αυτό το στάδιο, επιτυγχάνουμε εξαιρετικά αποτελεσματική διάκριση μεταξύ διαφορετικών τάξεων. Ωστόσο, οι πληροφορίες σχετικά με την τοποθεσία χάνονται. Για την ανάκτηση των πληροφοριών τοποθεσίας, η μείωση δειγματοληψίας ακολουθείται από μια διαδικασία upsampling η οποία λαμβάνει πολλές εικόνες χαμηλότερης ανάλυσης ως είσοδο και δίνει έναν χάρτη τμηματοποίησης υψηλής ανάλυσης ως έξοδο.[41]

Ένα σημαντικό πρόβλημα με το downsampling εντός δικτύου σε ένα FCN είναι ότι μειώνει την ανάλυση της εισόδου κατά έναν μεγάλο παράγοντα. Έτσι, κατά τη διάρκεια του upsampling γίνεται πολύ δύσκολη η αναπαραγωγή των λεπτότερων λεπτομερειών ακόμη και μετά τη χρήση εξελιγμένων τεχνικών όπως η Transpose Convolution (βλ. 2.6.1), με αποτέλεσμα να έχουμε μια “χονδροειδή” έξοδο. Για να αντιμετωπιστεί αυτό προσθέτονται «συνδέσεις παράλειψης» (skip connections) στο στάδιο Upsampling (βλ. 2.6.2).

Για παράδειγμα, το U-net [42] είναι μια αρχιτεκτονική νευρωνικού δικτύου τύπου κωδικοποιητή αποκωδικοποιητή για τμηματοποίηση εικόνας. Το όνομά του προέρχεται από το μοναδικό του σχήμα, όπου οι χάρτες χαρακτηριστικών της φύσης ανάλυσης στο βήμα της μείωσης δειγματοληψίας τροφοδοτούνται στα συνελικτικά στρώματα στο βήμα της ανερχόμενης δειγματοληψίας (συνδέσεις skip-layer). Η διαδρομή συστολής συλλαμβάνει πληροφορίες συμφραζόμενων μειώνοντας διαδοχικά την ανάλυση των χαρτών χαρακτηριστικών, ενώ η διαδρομή σύνδεσης συνδυάζει χάρτες χαρακτηριστικών από τις διαδρομές συστολής και επέκτασης για να συνδυάσει τοπικές πληροφορίες με ακριβή εντοπισμό (εικόνα 3.11).

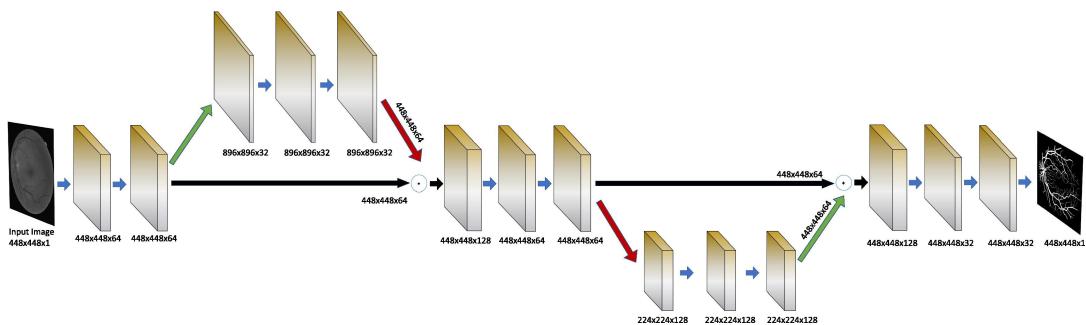


Σχήμα 3.11: Περιγραφή της encoder-decoder αρχιτεκτονικής U-net με την αναγνώριση των διαδρομών συστολής και επέκτασης

Η συστελλόμενη διαδρομή ακολουθεί την τυπική αρχιτεκτονική ενός συνελικτικού δικτύου.

Αποτελείται από την επαναλαμβανόμενη εφαρμογή δύο συνελίξεων 3×3 (χωρίς επιθέματα - un-padded), η καθεμία ακολουθούμενη από μια λειτουργία συγκέντρωσης ReLU και 2×2 max pooling λειτουργία με βήμα 2 για μείωση δειγματοληψίας [42]. Κάθε βήμα στο επεκταμένο μονοπάτι αποτελείται από μια υπερ-δειγματοληψία του χάρτη χαρακτηριστικών που ακολουθείται από μια συνέλιξη 2×2 ('επάνω-συνέλιξη') που μειώνει στο μισό τον αριθμό των καναλιών χαρακτηριστικών, μια συνένωση με τον αντίστοιχο περικομμένο χάρτη χαρακτηριστικών από τη διαδρομή συστολής και δύο συνέλιξεις 3×3 , καθεμία ακολουθούμενη από ένα ReLU.

Παρόμοια με τη λογική που ακολουθείται στο U-Net, ακολουθήσαμε την εργασία των Atli, I., Gedik, O.S. [43] με την υλοποίηση του Sine-Net. Οι Atli et al. πρότειναν ένα μοντέλο το οποίο εφαρμόζει upsampling για τη σύλληψη των λεπτών αγγειακών χαρακτηριστικών και στη συνέχεια τη μείωση δειγματοληψίας (downsampling) για τη σύλληψη παχιών αγγειακών χαρακτηριστικών. Χρησιμοποιήθηκαν διαφορετικές στρατηγικές ώστε να επικεντρωθούν σε διαφορετικές περιπτώσεις αγγειακών χαρακτηριστικών και να επιτύχουν έτσι μεγαλύτερη ποικιλομορφία, κάτι που ήταν αποτελεσματικό και το Sine-Net απέδωσε καλά στην τμηματοποίηση των αγγείων, ειδικά στο δείκτη προσδιοριστικότητας (specificity). Το παραπάνω δομημένο δίκτυο χωδικοποιητή-αποχωδικοποιητή χρησιμοποιεί κλιμακωτή συνέλιξη για την εξαγωγή σημασιολογικών αναπαραστάσεων υψηλού επιπέδου, αλλά χρησιμοποιεί σύνδεση παράβλεψης (skip connection) για τη συνένωση των χαρακτηριστικών χωδικοποιητή και αποχωδικοποιητή. Αυτή η δομή χωδικοποίησης-αποχωδικοποίησης μπορεί να ενισχύσει την αναγνώριση των πληροφοριών για τα όρια του αγγείου. Ωστόσο, εάν το πραγματικό δεκτικό πεδίο (receptive field) του δικτύου είναι ανεπαρκές, θα οδηγήσει σε κακή κατάτμηση των λεπτών αγγείων. Η αρχιτεκτονική του δικτύου αυτού ονομάστηκε έτσι επειδή, με μια αφηρημένη ματιά, μοιάζει με ένα ημιτονοειδές κύμα και παρουσιάζεται στο σχήμα 3.12. Τα μαύρα βέλη στο σχήμα υποδεικνύουν τη μεταφορά χαρτών χαρακτηριστικών από τα ενδιάμεσα επίπεδα στα βαθύτερα στρώματα της αρχιτεκτονικής. Τα πράσινα βέλη δείχνουν τη λειτουργία δειγματοληψίας προς τα πάνω up-sampling και τα κόκκινα βέλη δείχνουν τη δειγματοληψία προς τα κάτω down-sampling.



Σχήμα 3.12: Περιγραφή της encoder-decoder αρχιτεκτονικής Sine-net

Κεφάλαιο 4

Προτεινόμενα Συστήματα

Στο κεφάλαιο αυτό θα αναλυθεί η ανάπτυξη των προτεινόμενων συστημάτων, στο πλαίσιο προσέγγισης του προβλήματος τμηματοποίησης των εικόνων αμφιληστροειδούς σε αγγεία και μη. Θα γίνει μια λεπτομερής περιγραφή των αρχιτεκτονικών που δομήθηκαν αλλά και των διάφορων μεθοδολογιών ενσωμάτωσης του συνόλου δεδομένων που εφαρμόστηκαν στην πράξη. Όπως έχει ήδη αναφερθεί η προσέγγιση του συγχεκριμένου προβλήματος θα μελετηθεί από δύο οπτικές, έτσι η διαδικασία περιγραφής των προτεινόμενων συστημάτων θα γίνει ξεχωριστά για την ταξινόμηση των patches της εικόνας σύμφωνα με το κεντρικό τους pixel και την τμηματοποίηση των patches με την επιμέρους ταξινόμηση όλων των pixels τους. Τελικός στόχος ήταν ο συνδυασμός των δύο συστημάτων που παρουσιάζουν τις καλύτερες επιδόσεις σε ένα ενιαίο σύστημα.

4.1 Συστήματα Ταξινόμησης Τεμαχίων Εικόνας

Η αρχιτεκτονική που χρησιμοποιήθηκε στην προσέγγιση αυτή βασίστηκε στο δίκτυο Lisk-Net [31], όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο. Σε όλες τις αρχιτεκτονικές, μια στοίβα συνελικτικών επιπέδων ακολουθείται από τρία πλήρως συνδεδεμένα επίπεδα (FC): τα δύο πρώτα έχουν 512 μονάδες το καθένα, το τρίτο εκτελεί αμφίδρομη ταξινόμηση και επομένως περιέχει 2 μονάδες (μία για κάθε κατηγορία) (Πίνακας 4.1).

PLAIN	$conv64 \rightarrow conv64 \rightarrow maxpool \rightarrow conv128 \rightarrow conv128 \rightarrow maxpool \rightarrow fc512 \rightarrow fc512 \rightarrow fc2$ $[4 \times 4 \times 1 \times 0] \quad [3 \times 3 \times 1 \times 1] \quad [2 \times 2 \times 2 \times 0] \quad [3 \times 3 \times 1 \times 1] \quad [3 \times 3 \times 1 \times 1] \quad [2 \times 2 \times 2 \times 0]$
No-POOL	$conv64 \rightarrow conv64 \rightarrow conv128 \rightarrow conv128 \rightarrow fc512 \rightarrow fc512 \rightarrow fc2$ $[3 \times 3 \times 1 \times 1] \quad [3 \times 3 \times 1 \times 1] \quad [3 \times 3 \times 1 \times 1] \quad [3 \times 3 \times 1 \times 1]$

Πίνακας 4.1: Αρχιτεκτονική των αξιολογούμενων CNN. Τα ονόματα των επιπέδων ακολουθούνται από αριθμούς χαρτών χαρακτηριστικών. Οι τετράγωνες αγκύλες καθορίζουν το μέγεθος RF, τον διασκελισμό και την επένδυση

4.1.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία - Plain-Net

Η αρχιτεκτονική, όπως φαίνεται στο σχήμα 4.1, ξεκινά με ένα layer εισόδου που ακολουθείται από δύο convolutional layers με 64 φίλτρα το καθένα. Η διάσταση των πυρήνων είναι 4×4 στο πρώτο και 3×3 στο δεύτερο προκειμένου να διαφοροποιηθεί το αγγείο από το φόντο εξάγοντας μικρά και τοπικά χαρακτηριστικά. Η χρήση μιας μικρής τιμής διασκελισμού (stride=1) διασφαλίζει την ανάκτηση της μέγιστης πληροφορίας. Μετά από αυτά, εφαρμόζεται ένα max-pooling layer με ένα μικρό φίλτρο 2×2 , έτσι ώστε τα πιο πολλά υποσχόμενα χαρακτηριστικά να μεταφέρονται στο επόμενο convolutional layer. Αυτή η διαδικασία επαναλαμβάνεται με αύξηση του αριθμού των χαρτών (από 64 σε 128) και διατήρηση του μεγέθους των φίλτρων που χρησιμοποιούνται στα convolutional layers (3×3) και στα pooling layers (2×2). Το μέγεθος του stride παραμένει αμετάβλητο. Τα τελικά στρώματα του δικτύου είναι τρία πλήρως συνδεδεμένα στρώματα, το δύο πρώτα με 512 νευρώνες και το τελευταίο με δύο νευρώνες από τα οποία δίνεται η πιθανότητα κάθε pixel να ανήκει στο φόντο ή στη δομή του αγγείου. Όσον αφορά τις συναρτήσεις ενεργοποίησης, χρησιμοποιήθηκε μια συνάρτηση ενεργοποίησης ReLu στα κρυφά επίπεδα και μια συνάρτηση ενεργοποίησης Sigmoid λήφθηκε υπόψη στο τελευταίο Dense layer, καθώς το πρόβλημα ταξινόμησης εδώ είναι να αποδοθεί μια κατηγορία για κάθε pixel.

Layer	Type	Maps and Neurons	Kernel Size	Kernel Initializer	Stride	Padding	Activation Function
1	Input	3 Maps of 27x27 Neurons	-	-	-	-	-
2	Convolutional	64 Maps of 24x24 Neurons	4x4	Xavier	1	0	ReLu
3	Convolutional	64 Maps of 24x24 Neurons	3x3	Xavier	1	1	ReLu
4	Max Pooling	64 Maps of 12x12 Neurons	2x2	-	2	0	-
5	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLu
6	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLu
7	Max Pooling	64 Maps of 6x6 Neurons	2x2	-	2	0	-
8	Fully Connected	512 neurons	-	Normal	-	-	ReLu
9	Dropout	50% rate	-	-	-	-	-
10	Fully Connected	512 neurons	-	Normal	-	-	ReLu
11	Dropout	50% rate	-	-	-	-	-
12	Fully Connected	2 neurons	-	Normal	-	-	Sigmoid

Σχήμα 4.1: Δομή του μοντέλου Plain Lisk-Net όπως αυτό προτάθηκε

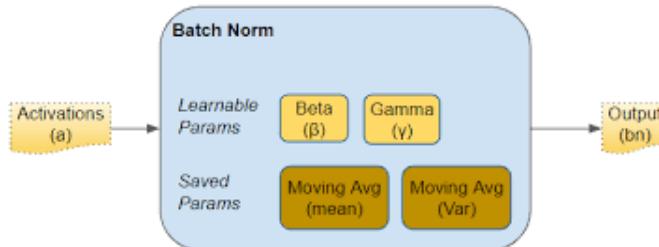
- Batch Normalization layers:

Όπως αναφέρθηκε προηγουμένως, η προεπεξεργασία των δεδομένων που χρησιμοποιούνται στη διαδικασία εκπαίδευσης είναι σημαντική για την απόδοση ενός CNN. Εκτός από αυτό, άλλες στρατηγικές σχετικές με την αρχιτεκτονική δικτύου έχουν δοκιμαστεί για την

επίτευξη του ίδιου στόχου. Μια τέτοια τεχνική είναι η κανονικοποίηση παρτίδας (γνωστή και ως batch normalization) που χρησιμοποιείται συχνά για να γίνουν τα τεχνητά νευρωνικά δίκτυα ταχύτερα και πιο σταθερά μέσω της κανονικοποίησης των εισόδων των επιπέδων με εκ νέου κεντράρισμα και επαναχλιμάκωση. Προτάθηκε από τους Sergey Ioffe, Christian Szegedy [44]. Η κανονικοποίηση παρτίδας είναι μια τεχνική που χρησιμοποιείται για να παρέχονται, σε οποιοδήποτε επίπεδο σε ένα CNN, είσοδοι που έχουν μηδενική μέση τιμή και μοναδιαία διακύμανση. Εάν μπορέσουμε να ομαλοποιήσουμε με κάποιο τρόπο τις ενεργοποιήσεις από κάθε προηγούμενο στρώμα, τότε η κατάβαση κλίσης θα συγκλίνει καλύτερα κατά τη διάρκεια της εκπαίδευσης. Αυτό ακριβώς κάνει για εμάς το επίπεδο Batch Norm που όπως κάθε άλλο δίκτυο, έχει τις δικές του παραμέτρους.

- Δύο παραμέτρους που μπορούμε να τις εκμάθουμε και ονομάζονται βήτα και γάμμα.
- Δύο παράμετροι που δεν μπορούμε να τις εκμάθουμε (Mean Moving Average και Variance Moving Average) αποθηκεύονται ως μέρος της «κατάστασης» του επιπέδου Batch Norm.

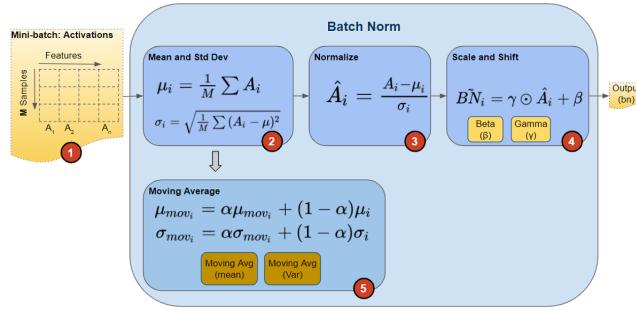
Αυτές οι παράμετροι είναι ανά επίπεδο Batch Norm. Έτσι, αν έχουμε, ας πούμε, τρία κρυφά επίπεδα και τρία επίπεδα Batch Norm στο δίκτυο, θα έχουμε τρεις παραμέτρους βήτα και γάμμα για τα τρία επίπεδα. Ομοίως για τις παραμέτρους Moving Average.



Σχήμα 4.2: Οι παράμετροι του Batch Norm layer

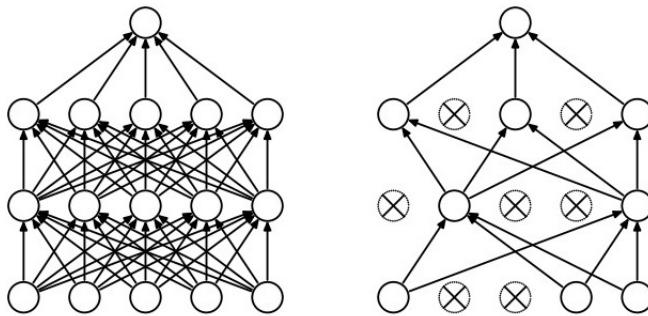
Τα στρώματα αυτά κανονικοποιούν πρώτα τις ενεργοποιήσεις κάθε καναλιού αφαιρώντας τη μέση τιμή της μίνι παρτίδας (mini-batch) και διαιρώντας με την τυπική απόκλιση της μίνι παρτίδας, οι εξισώσεις αυτής της διαδικασίας παρουσιάζονται στο Σχήμα 4.3. Στη συνέχεια, το στρώμα μετατοπίζει την είσοδο κατά β και την κλιμακώνει κατά έναν παράγοντα κλίμακας γ . Και οι δύο αυτές μεταβλητές ενημερώνονται κατά τη διάρκεια της εκπαίδευσης δικτύου. Τα στρώματα κανονικοποίησης παρτίδας χρησιμοποιούνται μεταξύ των γραμμικών και μη γραμμικών επιπέδων, γενικά μεταξύ του συνελικτικού στρώματος και του ReLU. Μειώνοντας τις μη γραμμικότητες μεταξύ των επιπέδων και ενημερώνοντας τις παραμέτρους γ και β , αυτός ο τύπος επιπέδου συμβάλλει στη μείωση της ευαισθησίας του δικτύου στο preprocessing, στην επιτάχυνση της εκπαίδευσης του δικτύου και στην αποφυγή της υπερπροσαρμογής.

- **Dropout layers:**



Σχήμα 4.3: Οι υπολογισμοί που πραγματοποιούνται από το Batch Norm layer

Μία από τις πρώτες προσεγγίσεις που χρησιμοποίησαν στρώματα Dropout παρουσιάστηκε από τους Srivastava et al. [45]. Το Dropout είναι μια τεχνική regularization, όπου τυχαία επιλεγμένοι νευρώνες αγνοούνται ή «αποκλείονται» κατά τη διάρκεια της εκπαίδευσης και έτσι προσεγγίζει την παράλληλη εκπαίδευση μεγάλου αριθμού νευρωνικών δικτύων με διαφορετικές αρχιτεκτονικές, επειδή κάθε ενημέρωση σε ένα επίπεδο κατά τη διάρκεια της εκπαίδευσης εκτελείται με διαφορετική «όψη» του διαμορφωμένου επιπέδου. Επειδή οι έξοδοι ενός στρώματος Dropout υποδειγματίζονται τυχαία, αυτό έχει ως αποτέλεσμα τη μείωση της χωρητικότητας ή την αραιώση του δικτύου κατά τη διάρκεια της εκπαίδευσης. Ως εκ τούτου, ένα ευρύτερο δίκτυο, π.χ. περισσότεροι κόμβοι, ενδέχεται να απαιτούνται κατά τη χρήση Dropout. 4.4



Σχήμα 4.4: Απεικόνιση της διαδικασίας dropout. Στα αριστερά είναι ένα παράδειγμα της τυπικής διαδικασίας και στα δεξιά είναι το αποτέλεσμα του Dropout.

Η ερμηνεία της υπερπαραμέτρου Dropout είναι η πιθανότητα εκπαίδευσης ενός δεδομένου κόμβου σε ένα επίπεδο, όπου το 1.0 σημαίνει ότι δεν υπάρχει αποκλεισμός και το 0.0 σημαίνει ότι δεν υπάρχουν έξοδοι από το επίπεδο. Μια καλή τιμή για την παράμετρο σε ένα κρυφό επίπεδο είναι μεταξύ 0.5 και 0.8. Τα επίπεδα εισόδου χρησιμοποιούν μεγαλύτερο ποσοστό αποκλεισμού, όπως 0.8.

Στη συνέχεια της εισαγωγής στα Batch Normalization στρώματα, θα περιγράψουμε πώς αυτά χρησιμοποιήθηκαν για την επέκταση της αρχιτεκτονικής δικτύου που περιγράφηκε στο σχήμα 4.5.

Η προσθήκη τέτοιων στρωμάτων συμβαίνει συνήθως μετά τα επίπεδα συνέλιξης, ακριβώς πριν την εφαρμογή της μη-γραμμικότητας δηλαδή των activation functions που στην περίπτωσή μας είναι ReLu και πριν τα επίπεδα Max-Pooling. Στις συνελίξεις, έχουμε κοινά φίλτρα που πηγαίνουν κατά μήκος και κατά πλάτος των χαρτών χαρακτηριστικών της εικόνας εισόδου. Αυτά τα φίλτρα είναι τα ίδια σε κάθε χάρτη χαρακτηριστικών. Έτσι, είναι λογικό να ομαλοποιήσουμε την έξοδο, κατά τον ίδιο τρόπο, κοινοποιώντας την κανονικοποίηση στους χάρτες χαρακτηριστικών.

Τα σημεία που αυτά προστέθηκαν φαίνονται στον παρακάτω πίνακα:

Layer	Type	Maps and Neurons	Kernel Size	Kernel Initializer	Stride	Padding	Activation Function
1	Input	3 Maps of 27x27 Neurons	-	-	-	-	-
2	Convolutional	64 Maps of 24x24 Neurons	4x4	Xavier	1	0	ReLU
3	Convolutional	64 Maps of 24x24 Neurons	3x3	Xavier	1	1	ReLU
4	Batch Normalization	-	-	-	-	-	-
5	Max Pooling	64 Maps of 12x12 Neurons	2x2	-	2	0	-
6	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLU
7	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLU
8	Batch Normalization	-	-	-	-	-	-
9	Max Pooling	64 Maps of 6x6 Neurons	2x2	-	2	0	-
10	Fully Connected	512 neurons	-	Normal	-	-	ReLU
11	Dropout	50% rate	-	-	-	-	-
12	Fully Connected	512 neurons	-	Normal	-	-	ReLU
13	Dropout	50% rate	-	-	-	-	-
14	Fully Connected	2 neurons	-	Normal	-	-	Sigmoid

Σχήμα 4.5: Δομή επιπέδων αρχιτεκτονικής Plain-Net με προσθήκη επιπέδων Batch Normalization

4.1.2 Προτεινόμενη αρχιτεκτονική και μεθοδολογία - No POOL-Net

Η αρχιτεκτονική αυτή είναι παρόμοια με την αριτεκτονική που παρουσιάστηκε για το Plain-Net-BN με κύρια διαφορά την απουσία των Max-Pooling layers. Λεπτομερέστερα, όπως φαίνεται στο σχήμα 4.6, η αρχιτεκτονική ξεκινά με ένα layer εισόδου που ακολουθείται από δύο convolutional layers με 64 φίλτρα το καθένα. Η διάσταση των πυρήνων είναι 3×3 και στα δύο πρώτα επίπεδα συνέλιξης αυτή τη φορά, καθώς και η τιμή διασκελισμού (stride=1). Μετά από αυτά μεταφερόμαστε κατευθείαν στα δύο επόμενα convolutional layers, αφού έχει προηγηθεί το Batch Normalization και η προηγηθείσα διαδικασία επαναλαμβάνεται με αύξηση του αριθμού των χαρτών (από 64 σε 128) και διατήρηση του μεγέθους των φίλτρων που χρησιμοποιούνται στα convolutional layers (3×3). Το μέγεθος του stride παραμένει αμετάβλητο. Τα τελικά στρώματα του δικτύου είναι τρία πλήρως συνδεδεμένα στρώματα, το δύο πρώτα με 512 νευρώνες και το τελευταίο με δύο νευρώνες. Όσον αφορά τις συναρτήσεις ενεργοποίησης, χρησιμοποιήθηκε μια συνάρτηση ενεργοποίησης ReLu στα κρυφά επίπεδα και μια συνάρτηση ενεργοποίησης Sigmoid λήφθηκε υπόψη στο τελευταίο Dense layer, καθώς το πρόβλημα ταξινόμησης εδώ είναι να

αποδοθεί μια κατηγορία για κάθε pixel. 4.6

Layer	Type	Maps and Neurons	Kernel Size	Kernel Initializer	Stride	Padding	Activation Function
1	Input	3 Maps of 27x27 Neurons	-	-	-	-	-
2	Convolutional	64 Maps of 24x24 Neurons	3x3	Xavier	1	1	ReLU
3	Convolutional	64 Maps of 24x24 Neurons	3x3	Xavier	1	1	ReLU
4	Batch Normalization	-	-	-	-	-	-
5	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLU
6	Convolutional	128 Maps of 12x12 Neurons	3x3	Xavier	1	1	ReLU
7	Batch Normalization	-	-	-	-	-	-
8	Fully Connected	512 neurons	-	Normal	-	-	ReLU
9	Dropout	50% rate	-	-	-	-	-
10	Fully Connected	512 neurons	-	Normal	-	-	ReLU
11	Dropout	50% rate	-	-	-	-	-
12	Fully Connected	2 neurons	-	Normal	-	-	Sigmoid

Σχήμα 4.6: Δομή επιπέδων αρχικής αρχιτεκτονικής No-POOL-Net

4.2 Συστήματα Κατάτμησης Τεμαχίων Εικόνας

Σε αντίθεση με τα συστήματα ταξινόμησης, η αρχιτεκτονική των συστημάτων τμηματοποίησης απαιτεί συνελικτικές μονάδες κωδικοποιητή και αποκωδικοποιητή. Η μονάδα κωδικοποίησης χρησιμοποιείται για την κωδικοποίηση εικόνων εισόδου σε ένα μεγαλύτερο αριθμό χαρτών με χαμηλότερες διαστάσεις. Η μονάδα αποκωδικοποίησης χρησιμοποιείται για την εκτέλεση λειτουργιών ανοδικής συνέλιξης (deconvolution) για την παραγωγή χαρτών τμηματοποίησης με τις ίδιες διαστάσεις με την αρχική εικόνα εισόδου. Επομένως, η αρχιτεκτονική για εργασίες τμηματοποίησης απαιτεί γενικά σχεδόν διπλάσιο αριθμό παραμέτρων δικτύου σε σύγκριση με την αρχιτεκτονική των εργασιών ταξινόμησης. Έτσι, είναι σημαντικό να σχεδιάσουμε αποτελεσματικές αρχιτεκτονικές DCNN για εργασίες τμηματοποίησης που μπορούν να εξασφαλίσουν καλύτερη απόδοση με λιγότερες παραμέτρους δικτύου.

4.2.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία

Οι Atli et al. [43] προτείνανε μια μέθοδο που βασίζεται σε ένα βαθύ νευρωνικό δίκτυο (CNN) εκπαιδευμένο σε ένα μεγάλο δείγμα παραδειγμάτων προεπεξεργασμένων με τον αλγόριθμο CLAHE (παράγραφος 3.1.5.3) για την αντιμετώπιση της διαφοράς φωτεινότητας μεταξύ των εικόνων και τον αλγόριθμο Multi-Scale Top Hat για την εκτενέστερη ενίσχυση της αντίθεσης και των λεπτομερειών της εικόνας (παράγραφος 3.1.5.4). Μετά από αυτό, για να αποφύγουν την υπερβολική προσαρμογή, χρησιμοποίησαν γεωμετρικούς μετασχηματισμούς όπως τεχνικές περιστροφής, μετατόπισης και κατοπτρισμού για την προσαύξηση των δεδομένων, όπως περιγράφηκε

και στην προηγούμενη ενότητα (κεφάλαιο 3.2.2). Σε αυτή την αρχιτεκτονική και αυτή τη μέθοδο του λεγόμενου Sine-Net βασιστήκαμε για να αναπτύξουμε το επιμυητό σύστημα κατάτμησης των εικόνων αιμοφόρων αγγείων αμφιβληστροειδούς της παρούσας εργασίας.

Η αρχιτεκτονική του δικτύου περιέχει 15 συνελικτικά στρώματα με μέγεθος φίλτρου 3×3 στο καθένα. Η Rectified Linear Unit (ReLU) χρησιμοποιείται ως συνάρτηση ενεργοποίησης μετά από κάθε συνέλιξη εκτός από το τελικό στρώμα. Η ενεργοποίηση του τελικού layer είναι tanh επειδή οι τιμές των προβλέψεων πρέπει να βρίσκονται μεταξύ 0 και 1 (μη γραμμικές). Δεδομένου ότι το προηγούμενο επίπεδο της τελικής ενεργοποίησης είναι ReLU, η είσοδος στο tanh είναι είτε μηδέν είτε κάποια θετική τιμή, γεγονός που αναγκάζει την έξοδο να μεταβεί στην περιοχή 0 και 1. Η συνέλιξη μετατόπισης (transpose-convolution, παράγραφος 2.6.1) για τη λειτουργία δειγματοληψίας προς τα πάνω, καθώς και η δειγματοληψία προς τα κάτω με συνέλιξη ενός βήματος 2, επιτρέπουν στο δίκτυο να συνεχίζει να μαθαίνει ιεραρχικές αναπαραστάσεις ενημερώνοντας τα βάρη, τα οποία δεν υπάρχουν στα max-pooling ή upsampling layers. Ο αριθμός των καναλιών διπλασιάζεται και μειώνεται στο μισό μετά από κάθε λειτουργία δειγματοληψίας προς τα κάτω (υπό-δειγματοληψία) και προς τα πάνω (υπό-δειγματοληψία), αντίστοιχα.

Ο Πίνακας 4.7 παρουσιάζει τις δομές των επιπέδων και τις παραμέτρους της αρχιτεκτονικής CNN. Τα επίπεδα που επισημαίνονται με έντονη γραφή υποδηλώνουν μεταφορές χαρακτηριστικών residual connections από το τρέχον επίπεδο σε βαθύτερα επίπεδα. Η λογική αυτή έχει περιγραφεί στο κεφάλαιο 2.7.

Layer Id	Type	# of channel	Kernel size	Output shape	Activation function
0	Input	1	-	$448 \times 448 \times 1$	-
1	Conv2D	64	3×3	$448 \times 448 \times 64$	ReLU
2	Conv2D	64	3×3	$448 \times 448 \times 64$	ReLU
3	TransConv2Dup	32	3×3	$896 \times 896 \times 32$	ReLU
4	Conv2D	32	3×3	$896 \times 896 \times 32$	ReLU
5	Conv2D	32	3×3	$896 \times 896 \times 32$	ReLU
6	Conv2D_down	64	3×3	$448 \times 448 \times 64$	ReLU
7	Conv2D	64	3×3	$448 \times 448 \times 64$	ReLU
8	Conv2D	64	3×3	$448 \times 448 \times 64$	ReLU
9	Conv2D_down	128	3×3	$224 \times 224 \times 128$	ReLU
10	Conv2D	128	3×3	$224 \times 224 \times 128$	ReLU
11	Conv2D	128	3×3	$224 \times 224 \times 128$	ReLU
12	TransConv2Dup	64	3×3	$448 \times 448 \times 64$	ReLU
13	Conv2D	32	3×3	$448 \times 448 \times 32$	ReLU
14	Conv2D	32	3×3	$448 \times 448 \times 32$	ReLU
15	Conv2D	1 s	1×1	$448 \times 448 \times 1$	ReLU
16	Output	1	-		tanh

Σχήμα 4.7: Δομή του μοντέλου Sine-Net

4.3 Συνδυασμός Συστημάτων - Ensemble Learning

Στην ενότητα αυτή θα γίνει μια προσπάθεια για την ανάπτυξη ενός ενιαίου συστήματος αξιοποίησης και των δύο λογικών αρχιτεκτονικής δικτύου με την τεχνική Ensemble Learning, για την προσέγγιση του προβλήματος κατάτμησης αγγείων από την εικόνα αμφιβληστροειδούς. Για την ανάπτυξη του συστήματος αυτού θα χρησιμοποιηθούν τα βασικά συστήματα, όπως περι-

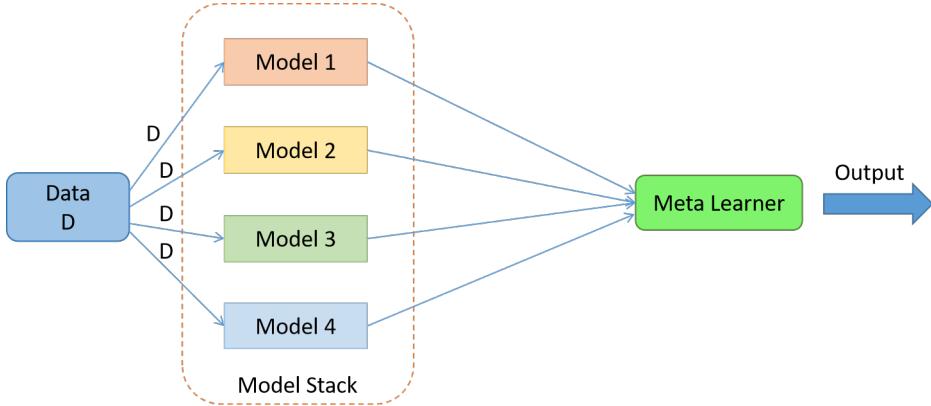
γράφηκαν προηγουμένως, που πετυχάνουν τις βέλτιστες επιδόσεις τόσο με τη μέθοδο patch-wise classification όσο και με τη μέθοδο pixel-wise segmentation.

Ένα σύνολο Ensemble μοντέλων λειτουργεί ακριβώς όπως ένα εκλογικό μηχάνημα. Κάθε μέλος του συνόλου συνεισφέρει εξίσου κάνοντας ψηφοφορία χρησιμοποιώντας τις προβλέψεις του. Ως εκ τούτου, εισάγει μια έννοια της διαφορετικότητας όταν κάθε ένα από τα μοντέλα συνδυάζεται. Αυτή η αίσθηση διαφορετικότητας θα μειώσει τη διακύμανση και θα βελτιώσει την ικανότητα γενίκευσης πέρα από τα δεδομένα εκπαίδευσης. Η μείωση της διακύμανσης των προβλέψεων θα γίνει στο ποσό του λάθους στις προβλέψεις που έγιναν που μπορεί να αποδοθεί στη «διακύμανση». Αυτό δεν συμβαίνει πάντα, αλλά όταν συμβαίνει, αυτή η μείωση της διακύμανσης, με τη σειρά της, οδηγεί σε βελτιωμένη προγνωστική απόδοση.

4.3.1 Προτεινόμενη αρχιτεκτονική και μεθοδολογία

Μία από τις μεθόδους εφαρμογής της Ensemble τεχνικής είναι το Stacking Generalization ή Stacking ή αλλιώς η στοίβαξη.[46] Στη στοίβαξη, ο σκοπός είναι να μειωθεί το σφάλμα γενίκευσης διαφορετικών μοντέλων. Ένας αλγόριθμος παίρνει τις εξόδους των απλών υπομοντέλων ως είσοδο και προσπαθεί να μάθει πώς να συνδυάζει καλύτερα τις προβλέψεις εισόδου για να κάνει μια καλύτερη πρόβλεψη εξόδου.

Η στοίβαξη θεωρεί συχνά ετερογενή αδύναμα εκπαίδευμενα συστήματα («αδύναμους μαθητές» ή weak learners), τα μαθαίνει παράλληλα και τα συνδυάζει εκπαίδευοντας ένα μετα-εκπαίδευμενό (meta-learner) να κάνει μια πρόβλεψη με βάση τις προβλέψεις των διαφορετικών weak learners. Ένας meta-learner εισάγει τις προβλέψεις ως χαρακτηριστικά και ο στόχος είναι οι ground truth τιμές στα δεδομένα (4.8), και επιχειρεί, έτσι, να μάθει πώς να συνδυάζει καλύτερα τις προβλέψεις εισόδου για να κάνει καλύτερη πρόβλεψη εξόδου.

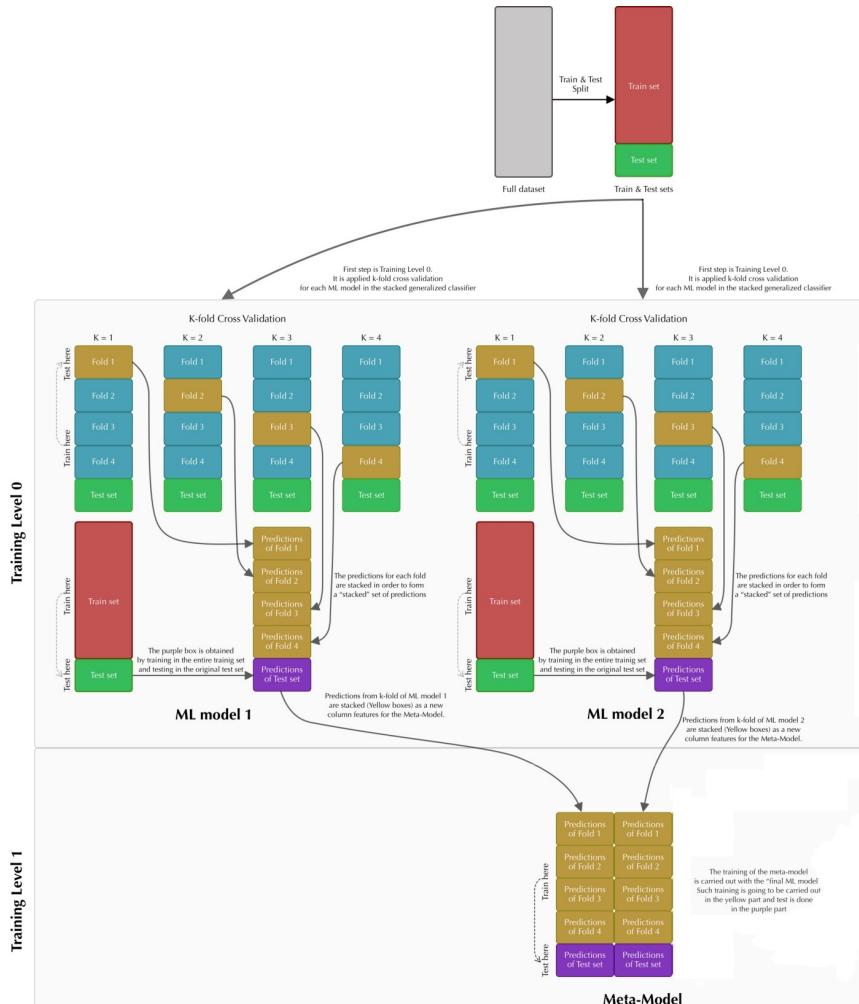


Σχήμα 4.8: Αναπαράσταση της γενικής διαδικασίας του Stacking Ensemble

Η μέθοδος στοίβαξης συνήθως αποτελείται από 2 στάδια εκπαίδευσης, γνωστά ως «επίπεδο 0» και «επίπεδο 1». Ο στόχος του πρώτου σταδίου (επίπεδο 0) είναι η δημιουργία των δεδομένων εκπαίδευσης για το μετα-μοντέλο, αυτό πραγματοποιείται με την εφαρμογή k-fold cross validation για κάθε «αδύναμο μαθητή» που ορίζεται στο πρώτο στάδιο. Οι προβλέψεις καθενός από αυτούς

τους «αδύναμους μαθητές» «στοιβάζονται» προκειμένου να δημιουργηθεί ένα «νέο εκπαιδευτικό σύνολο» (το μετα-μοντέλο). Ο στόχος του δεύτερου σταδίου (επίπεδο 1) είναι να εκπαιδεύσει το μετα-μοντέλο, και μια τέτοια εκπαίδευση πραγματοποιείται μέσω ενός ήδη καθορισμένου «τελικού μαθητή».

Στο σχήμα 4.9 βλέπουμε μια γραφική περιγραφή ενός Stacking Generalization Classifier που αποτελείται από 2 βασικά μοντέλα (αδύναμοι μαθητές) και έναν τελικό εκτιμητή.



Σχήμα 4.9: Αναπαράσταση της της αρχιτεκτονικής που χρησιμοποιήθηκε στην εργασία για το Stacking Ensemble

Κεφάλαιο 5

Πειραματική Διαδικασία

Στο κεφάλαιο αυτό θα αναλυθεί η διαδικασία εκπαίδευσης και επαλήθευσης των συστημάτων που περιγράφηκαν στο κεφάλαιο 4. Αφού έχουν καθοριστεί οι αρχιτεκτονικές και ο τρόπος με τον οποίο θα τροφοδοτηθούν τα δεδομένα στα μοντέλα το επόμενο βήμα που πρέπει να ολοκληρωθεί είναι η εκπαίδευση τους. Επί της ουσίας, ο όρος εκπαίδευση αναφέρεται στην επιλογή κατάλληλων τιμών για τις παραμέτρους του συστήματος (βάρη, πολώσεις), με σκοπό την παραγωγή βέλτιστων εξόδων για γνωστά, και κατ' επέκταση γνωστά, δεδομένα εισόδου. Για την ορθή εκπαίδευση των μοντέλων στην πράξη θα εφαρμοστούν οι αλγόριθμοι Backpropagation (2.4.4.2) και Κατάβασης Κλίσης (2.4.4.3).

Τα δεδομένα σε ένα πρόβλημα επιβλεπόμενης μάθησης, όπως το πρόβλημα που προσπαθεί να αντιμετωπίσει η παρούσα διπλωματική, εμφανίζονται σε ζευγάρια (X, Y), όπου το X αναπαριστά τα δεδομένα εισόδου ενώ το Y την τιμή στόχο (label/target) που προσπαθεί να προβλέψει το μοντέλο. Τα εκάστοτε δεδομένα εισόδου για την σωστή εκπαίδευση και επαλήθευση των μοντέλων χωρίζονται σε τρεις κατηγορίες, τα δεδομένα εκπαίδευσης (training data), τα δεδομένα επαλήθευσης (validation data) και τα δεδομένα ελέγχου (testing data). Τα δεδομένα εκπαίδευσης αποτελούν το μεγαλύτερο μέρος των δεδομένων της εισόδου και όπως είναι προφανές από την ονομασία τους είναι τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση ενός μοντέλου. Τα δεδομένα επαλήθευσης συνήθως συνθέτουν το μικρότερο τμήμα των δεδομένων εισόδου και χρησιμοποιούνται για την επαλήθευση των αποτελεσμάτων σε κάθε στάδιο της εκπαίδευσης. Τέλος, τα δεδομένα ελέγχου χρησιμοποιούνται μετά την ολοκλήρωση της εκπαίδευσης του μοντέλου για την καταγραφή της επίδοσης του.

Δύο ακόμη σημαντικοί παράγοντες για την αποτελεσματική εκπαίδευση ενός μοντέλου είναι οι επιλογή των κατάλληλων τιμών για τον αριθμό των εποχών (epochs) και για το μέγεθος παρτίδας εκπαίδευσης (batch size). Ο αριθμός των εποχών αναφέρεται στον αριθμό των φορών που θα επαναληφθεί η διαδικασία εκπαίδευσης ή στον αριθμό των περασμάτων πάνω σε ολόκληρο το σύνολο των δεδομένων εκπαίδευσης, μέχρι να επιλεγούν οι τελικές τιμές των παραμέτρων του δικτύου. Το μέγεθος παρτίδας αναφέρεται στον αριθμό των δειγμάτων εισόδου που θα χρησιμοποιηθούν ως παρτίδα για τον υπολογισμό της κλίσης και τη χρήση του για την ανανέωση των εσωτερικών παραμέτρων του δικτύου, όπως περιγράφουν οι αλγόριθμοι Backpropagation

και Κατάβασης Κλίσης.

Τελικά, μόλις ολοκληρωθεί η εκπαίδευση των μοντέλων το τελικό βήμα αφορά την αξιολόγηση των εξόδων που αποδίδουν. Για την καταγραφή των αποτελεσμάτων ότια χρησιμοποιηθούν ορισμένες από τις μετρικές αξιολόγησης, όπως περιγράφονται στην ενότητα 2.8, επί των δεδομένων ελέγχου. Η πειραματική διαδικασία ότια αξιολογηθεί χυρίως βάση των μετρικών Accuracy, που περιγράφει των αριθμό των σωστών προβλέψεων προς το σύνολο όλων των προβλέψεων, και Loss, που περιγράφει το σφάλμα της συνάρτησης κόστους εν προκειμένω της Categorical Crossentropy και της Binary Crossentropy. Για την καλύτερη εποπτεία της επίδοσης στη συνέχεια ότια παρουσιαστούν οι τιμές των μετρικών αξιολόγησης αλλά και ότια συγχριθούν οι τιμές που υπολογίζονται κατά την εκπαίδευση και την επαλήθευση.

5.1 Σύστημα Ταξινόμησης Τεμαχίων Εικόνας

Στην ενότητα αυτή ότια περιγραφεί η πειραματική διαδικασία εκπαίδευσης και επαλήθευσης των patch binary classification συστημάτων με σκοπό την επιλογή εκείνου που ότια πετυχαίνει την καλύτερη επίδοση στο πρόβλημα της κατάτμησης των αιμοφόρων αγγείων του αμφιβληστροειδούς. Η διεξαγωγή των πειραμάτων ξεκινά από το απλό μοντέλο *Lisk – Net* με τις *PlainNet* και *No – PlainNet* αρχιτεκτονικές, και καταλήγει με την παρουσίαση ενός πειραματισμού με τη μέθοδο Structured Prediction. Τα διαθέσιμα δεδομένα για το συγκεκριμένο πρόβλημα αποτελούνται από 400,000 έγχρωμα patches εισόδου μεγέθους $27 \times 27 \times 3$ μαζί με τις αντίστοιχες ετικέτες του κεντρικού pixel των patches. Από κάθε εικόνα βυθού εκπαίδευσης δειγματοληπτούμε 20,000 patches με τυχαίο τρόπο ώστε ολόκληρο το training set να περιέχει 400,000 patches. Για τα δεδομένα ελέγχου, δηλαδή το testing set, εξάγουμε όλα τα patches από κάθε εικόνα ελέγχου, έτσι ώστε να έχουμε πρόβλεψη για κάθε κεντρικό pixel και να την αξιολογήσουμε με τη μεγαλύτερη δυνατή ακρίβεια. Τελικά, λαμβάνουμε υπόψην μας μόνο τα patches που ανήκουν εξολοκλήρου στην ενεργή κυκλική περιοχή της εικόνας βυθού, το FOV, οπότε και χάνουμε λίγα από τα εξαγώμενα δείγματα. Στο σχήμα 5.1 παρουσιάζεται η κατανομή των διαθέσιμων δεδομένων σε δεδομένα εκπαίδευσης και ελέγχου καθώς και η κατανομή τους στις δύο κλάσεις, όπως προκύπτει μετά την επαύξηση δεδομένων κατά 11 φορές.

	DRIVE		STARE	
Validation scheme	One-off train + test		Leave-one-out	
Image dimensions	565×584		605×700	
Fundus images	Training 20	Testing 20	Training 19	Testing 1
Total patches	3,857,818	3,993,351	5,326,283	280,330
- positive (vessel)	516,152	538,475	558,261	29,382
- negative	3,341,666	3,454,876	4,768,022	250,948

Σχήμα 5.1: Κατανομή διαθέσιμων δεδομένων DRIVE

5.1.1 Εκπαίδευση Δικτύου

Για την εκπαίδευση του μοντέλου Lisk-Net ωστε εκπαίδευτούν ξεχωριστά τα δύο μοντέλα που προκύπτουν από τις δύο διαφορετικές μεθοδολογίες-αρχιτεκτονικές (Plain Net & No-Pool Net), όπως περιγράφονται στην ενότητα 4.1. Για την εκπαίδευση του δικτύου χρησιμοποιείται ως συνάρτηση κόστους (loss function) η συνάρτηση binary crossentropy (βλέπε 2.28). Ως βελτιστοποιητής χρησιμοποιήθηκε ο SGD με ρυθμό μάθησης (learning rate) ίσο με 0.001 και τις παραμέτρους οριμής β_{eta_1} και β_{eta_2} ρυθμισμένες στα 0.9 και 0.999 αντίστοιχα. Ο αριθμός των εποχών εκπαίδευσης επιλέχθηκε μεταξύ 20 και 30 εποχές και το μέγεθος τεμαχίου ίσο με 128 ή 64. Για την επιλογή των βέλτιστων υπερπαραμέτρων εκπαίδευσης εφαρμόστηκε η τεχνική της αναζήτησης πλέγματος (grid search). Από τα δεδομένα εκπαίδευσης το 20% ωστε χρησιμοποιηθεί ως validation data για την επαλήθευση της εκπαίδευσης.

Για την υλοποίηση των βαθιών νευρωνικών δικτύων της κατηγορίας αυτής και του κεφαλαίου αυτού χρησιμοποιήθηκε η high-level διεπαφή προγραμματιστικών εφαρμογών (API) Keras [47].

Ακολουθούν οι τιμές που έλαβαν οι παράμετροι του δικτύου:

- Epochs = 30, 25, 20
- Batch size = 128 (Plain-Net-BN, No-Pool-Net-BN) ή 64
- Optimizer = SGD
 - Learning rate = 10^{-3}
 - Momentum = 0.9
 - Weight decay = $5 * 10^{-4}$
 - nesterov = True
- Learning rate reduction = *ReduceLROnPlateau()*
 - Gamma = 0.1
 - Step size = 6
 - Step decay = 0.1
 - Learning rate reduction if epoch is multiple of 3

5.1.1.1 Callbacks

Το Callback είναι ένα σύνολο λειτουργιών που ωστε εφαρμοστούν σε συγκεκριμένα στάδια της διαδικασίας εκπαίδευσης. Ένα callback ορίζεται και χρησιμοποιείται όταν θέλουμε να αυτοματοποιήσουμε ορισμένες εργασίες μετά από κάθε εκπαίδευση/εποχή που μας βοηθούν να έχουμε τον έλεγχο της εκπαίδευτικής διαδικασίας. Οι λειτουργίες callback που χρησιμοποιήθηκαν κατά την εκπαίδευση του συγκεκριμένου μοντέλου είναι:

- **ModelCheckpoint:** Χρησιμοποιείται για την αποθήκευση του καλύτερου μοντέλου ή μοντέλων (ή των βαρών του) μετά από κάθε εποχή, για μελλοντική χρήση σε προβλέψεις.

- monitor: συμβάν με βάση το οποίο θέλετε να αποθηκεύσετε το μοντέλο σας. Αν είναι το val_loss τότε θα έχουμε το καλύτερο μοντέλο όπου το validation loss είναι το ελάχιστο από όλες τις εποχές. Αν είναι το val_acc τότε θα έχουμε το καλύτερο μοντέλο όπου το validation accuracy είναι το υψηλότερο μεταξύ όλων των εποχών.
- save_best_only: Εάν είναι ίσο με True, τότε μόνο το καλύτερο μοντέλο θα αποθηκευτεί. Εάν είναι False τότε το μοντέλο θα αποθηκεύεται μετά από κάθε εποχή.
- mode: Γενικά χρησιμοποιείται το "auto". Εάν η λειτουργία έχει οριστεί σε "min" και monitor = "val_loss", τότε το μοντέλο θα αποθηκευτεί μόνο εάν μειωθεί η τιμή val_loss. Εάν mode = "max" και monitor = "val_acc", τότε το μοντέλο θα αποθηκευτεί μόνο εάν βελτιωθεί το val_acc.
- save_weights_only: Εάν είναι True, τότε μόνο τα βάρη του μοντέλου θα αποθηκευτούν. Διαφορετικά, ολόκληρο το μοντέλο θα αποθηκευτεί.

Στη δική μας περίπτωση ολόκληρο το δίκτυο αποθηκεύεται σε κάθε νέο lowest level του validation loss. Τα οφέλη αυτής της προσέγγισης είναι πολλαπλά καθώς μετά την ολοκλήρωση της εκπαίδευσης μπορούμε να δοκιμάσουμε την απόδοση του μοντέλου στο testing set σε διαφορετικά στάδια. Επιπλέον, μπορούμε να συνεχίσουμε την εκπαίδευση του μοντέλου μας. Σε αντίθεση με το early stopping δεν χρειάζεται να τερματιστεί η εκπαίδευση του δικτύου για την αποφυγή overfitting και έτσι μπορούμε να δούμε σε ποιο σημείο θα φτάσει το αριανές για να αξιολογήσουμε τις μεθόδους κανονικοποίησης του συστήματος.

- **ReduceLRonPlateau:** Παρακολουθεί μια συγκεκριμένη μετρική και μειώνει τον ρυθμό εκμάθησης κατά ένα δεδομένο ποσό όταν αυτή η μετρική σταματήσει να βελτιώνεται πέρα από ένα επίπεδο υπομονής (patience).

- patience: Αριθμός εποχών χωρίς βελτίωση στο παρακολουθούμενο γεγονός μετά από τον οποίο μειώνουμε το ρυθμό εκπαίδευσης.
- factor: Παράγοντας κατά τον οποίο θα μειωθεί ο ρυθμός εκμάθησης: $lr_{new} = lr_{old} * factor$

Συγκεκριμένα, στη δική μας περίπτωση, ο ρυθμός εκμάθησης υποδεκαπλασιάζεται (factor = 0.1) μετά από 2 εποχές (patience=2) στασιμότητας του validation loss (monitor = "val_loss").

5.1.1.2 Κλάση ImageDataGenerator

Η κλάση ImageDataGenerator παρέχεται από το Keras και αποτελεί ένα πολύ χρήσιμο εργαλείο κατά την εκπαίδευση CNNs. Στα πλαίσια της εργασίας δημιουργήθηκε μια customized εκδοχή αυτής της κλάσης για την "in-place" επαύξηση των δεδομένων εκπαίδευσης (data augmentation) σε πραγματικό χρόνο. Οπως αναλύσαμε στο θεωρητικό μέρος της εργασίας, η μέθοδος data augmentation αποτελεί μέτρο αντιμετώπισης του overfitting. Η CustomDataGenerator

λαμβάνει δεσμίδες batches δεδομένων εκπαίδευσης και αφού εφαρμόσει μία σειρά τυχαίων μετασχηματισμών (transformations) σε κάθε εικόνα του batch, τροφοδοτεί τις παραγόμενες εικόνες στο δίκτυο. Οι μετατροπές που θα εφαρμοστούν αποτελούν παραμέτρους της CustomData-Generator (rotation, horizontal_flip, vertical_flip, random_gamma, resize, βλ. 3.1.6). Έτσι το δίκτυο σε κάθε εποχή λαμβάνει ως είσοδο καινούργιες, ελαφρώς τροποποιημένες εικόνες, και εν συνεχεία εξάγει πιο εύρωστα (robust) χαρακτηριστικά.

5.1.2 Προκλήσεις κατά την πειραματική διαδικασία και βελτιστοποίηση του δικτύου

5.1.2.1 Ανισοροπία Κλάσεων (Class Imbalance)

Είναι σύνηθες στην επεξεργασία ιατρικής εικόνας η ανατομία που μας ενδιαφέρει να καταλαμβάνει μόνο ένα πολύ μικρό μέρος της εικόνας. Ως εκ τούτου, τα περισσότερα από τα εξαγόμενα patches ανήκουν στην περιοχή του φόντου, ενώ τα μικρά όργανα (ή ανωμαλίες) έχουν μεγαλύτερη σημασία. Η εκπαίδευση ενός δικτύου με τέτοια δεδομένα οδηγεί συχνά στη μεροληφτία του εκπαιδευμένου δικτύου ως προς το φόντο και στον εγκλωβισμό στα τοπικά ελάχιστα. Μια δημοφιλής λύση για αυτό το ζήτημα είναι η επαναστάθμιση δείγματος, όπου εφαρμόζεται υψηλότερο βάρος στα patches του φόντου κατά τη διάρκεια της εκπαίδευσης. Η αυτόματη τροποποίηση της επαναστάθμισης δείγματος έχει αναπτυχθεί χρησιμοποιώντας για παράδειγμα το Dice coefficient. Ωστόσο, η αποτελεσματικότητα είναι περιορισμένη στην αντιμετώπιση της ακραίας ανισορροπίας κλάσεων. Η εκπαίδευση Patch-wise σε συνδυασμό με την επιλογή patches μπορεί να βοηθήσει στην αντιμετώπιση του ζητήματος της ανισορροπίας κλάσεων.

Οι μηχανισμοί ελέγχου του προβλήματος της ανισοροπίας που δοκιμάστηκαν ή χρησιμοποιήθηκαν είτε κατά τη δημιουργία του σετ εκπαίδευσης, είτε κατά την διαδικασία εκπαίδευσης των δικτύων Lisk-Net που παρουσιάστηκαν στις προηγούμενες ενότητες, παρουσιάζονται παρακάτω:

- Έλεγχος της εξαγωγής των δεδομένων εκπαίδευσης ώστε τα κεντρικά pixels των training patches να ανήκουν ισάριθμα και στις δύο κλάσεις (vessel, non-vessel).
- Έλεγχος των βαρών που αποδίδονται στις δύο κλάσεις κατά την εκπαίδευση.
- Χρήση μετρικών απόδοσης που αντιλαμβάνονται καλύτερα class imbalance των δεδομένων, όπως το Weighted Average Accuracy και Jaccard Index που εξηγούνται στην επόμενη ενότητα.

Στην παρούσα εργασία, για το σύστημα Lisk_Net δοκιμάστηκαν δύο από τους προαναφερθέντες τρόπους. Αρχικά, κατά την εξαγωγή των patches από κάθε εικόνα, γινόταν έλεγχος του αριθμού των κεντρικών τους εικονοστοιχείων για το αν ανήκουν στη θετική ή στην αρνητική κλάση, ώστε ο τελικός πλυνθημός των δύο κλάσεων να είναι ο ίδιος (3.2.1). Η συνάρτηση extract_random_patches που υλοποιήθηκε, εξασφάλισε ότι για το μέγιστο αριθμό θετικών λευκών center pixels που ανήκουν σε αγγείο, που μπορεί να έχει μια εικόνα σύμφωνα με τη ground truth αντίστοιχη εικόνα, εξάγουμε τα αντίστοιχα θετικά patches και τον ίδιο αριθμό

αρνητικών-μαύρων patches. Πάντα με την προϋπόθεση ότι το patch εμπεριέχεται πλήρως στο αντίστοιχο FOV.

Στη συνέχεια, κατά τη διαδικασία της εκπαίδευσης, δοκιμάστηκε η στάθμιση της συνάρτησης κόστους ή αλλιώς η διαδικασία απόδοσης class weight για μια συγκεκριμένη κλάση ανάλογη με την εκπροσώπηση της. Για το λόγο αυτό, η συνάρτηση *add_sample_weights* υπολόγισε τα δύο αντίστοιχα βάρη των κλάσεων w_0 και w_1 ως εξής:

$$w_0 = \frac{\#total_pixels}{2 \times \#negative_pixels} \quad (5.1) \quad w_1 = \frac{\#total_pixels}{2 \times \#positive_pixels} \quad (5.2)$$

Το μοντέλο τιμωρεί τα λάθη σε δείγματα της κλάσης[i] με *class_weight*[i] αντί για 1. Άρα υψηλότερο βάρος κλάσης σημαίνει ότι θέλουμε να δώσουμε μεγαλύτερη έμφαση σε μια τάξη, που στην περίπτωσή μας είναι η τάξη θετικών-μηδενικών pixels που αντιστοιχούν σε αγγεία. Η κλιμάκωση κατά $total/2$ συμβάλλει στη διατήρηση του loss σε παρόμοια τάξη μεγέθους και το άθροισμα των βαρών όλων των παραδειγμάτων παραμένει το ίδιο.

Η μέθοδος αντικετώπισης του class imbalance που επιλέχθηκε ως η καλύτερη ήταν η ισοσταθμισένη εξαγωγή patches από τις δύο κλάσεις.

5.1.3 Αξιολόγηση Συστήματος

5.1.3.1 Συμπληρωματικές Μετρικές

Η επιλογή της κατάλληλης μετρικής για την αξιολόγηση των συστημάτων αποτελεί πρόκληση ιδιαίτερα σε προβλήματα ταξινόμησης με imbalanced dataset. Συγκεκριμένα, σε προβλήματα διάγνωσης καρκίνου συνηθίζεται τα καλοήθη δείγματα να υπερέχουν αριθμητικά έναντι των κακοήθων δειγμάτων. Αν για παράδειγμα σε ένα τέτοιο πρόβλημα επιλέξουμε ως μετρική την ακρίβεια (accuracy) και η αναλογία καλοήθων κακοήθων δειγμάτων είναι 10 προς 1, μπορεί το μοντέλο μας να πετύχει ικανοποιητικό σκορ (90%) ταξινομώντας όλα τα δείγματα ως καλοήθη. Στην πράξη όμως, αυτός ο ταξινομητής δεν έχει καμία χρησιμότητα, αντιθέτως, θα μπορούσε να χαρακτηριστεί ακόμα και επικίνδυνος, καθώς αποτυγχάνει στον εντοπισμό της παθογένειας. Γίνεται λοιπόν αντιληπτό ότι η επιλογή μιας ακατάλληλης μετρικής για την αξιολόγηση των συστημάτων μπορεί να λειτουργήσει παραπλανητικά.

Για τον παραπάνω λόγο, οι επιπλέον μετρικές που χρησιμοποιήθηκαν (πέραν όσων αναφέρθηκαν στο κεφάλαιο 2.8) για την αξιολόγηση των συστημάτων μας είναι:

- Intersection-Over-Union (Jaccard Index): είναι μια μέθοδος ποσοτικοποίησης του ποσοστού επικάλυψης μεταξύ της μάσκας-στόχου και της πρόβλεψης-εξόδου, που δίνεται από την εξίσωση:

$$IoU = \frac{\text{Area of Overlap (predicted segmentation, ground truth)}}{\text{Area of Union (predicted segmentation, ground truth)}} \quad (5.3)$$

Για δυαδική (δύο κλάσεις) τμηματοποίηση, η μέση *IoU* της εικόνας υπολογίζεται λαμβάνοντας το μέσο όρο των *IoU* κάθε κατηγορίας.

- Weighted average: υπολογίζεται μέσω της συνάρτησης `classification_report` της βιβλιοθήκης `sklearn.metrics`. Η μέση τιμή που προκύπτει αντιπροσωπεύει τη συνεισφορά κάθε κλάσης όπως αυτή σταθμίζεται με τον αριθμό των παραδειγμάτων στη συγκεκριμένη κλάση.

5.1.3.2 Βέλτιστος Συντονισμός Κατωφλίου Απόφασης στο Στάδιο Ελέγχου

Πολλές μέθοδοι μηχανικής μάθησης είναι σε θέση να προβλέψουν την πιθανότητα ή το βαθμό συμμετοχής στην κάθε κλάση. Αυτό είναι γενικά επιθυμητό, καθώς προσφέρει μια μέτρηση της εμπιστοσύνης ή της αβεβαιότητας μιας πρόβλεψης, ενώ δίνει επίσης περισσότερες λεπτομέρειες από την απλή πρόβλεψη της ετικέτας κλάσης. Μερικές εργασίες ταξινόμησης χρειάζονται ακριβή πρόβλεψη της ετικέτας κλάσης. Για παράδειγμα, ένα μοντέλο που προβλέπει πιθανότητες μπορεί να φαίνεται ότι έχει πολύ κακή απόδοση όταν αξιολογείται σύμφωνα με τις ετικέτες κλάσης του. Αυτό συμβαίνει επειδή παρόλο που οι προβλεπόμενες πιθανότητες μπορεί να επιδεικνύουν επιδεξιότητα, πρέπει να ερμηνεύονται σύμφωνα με το κατάλληλο όριο πρίν τη μετατροπή τους σε διαχριτές ετικέτες κλάσης. Το όριο διέπει τη μετατροπή μια προβλεπόμενης πιθανότητας σε ετικέτα κλάσης. Για κανονικοποιημένες προβλεπόμενες πιθανότητες στην περιοχή από 0 έως 1, το κατώφλι έχει οριστεί στο 0,5 από προεπιλογή. Το προεπιλεγμένο κατώφλι μπορεί να μην είναι ο καλύτερος τρόπος ερμήνευσης της πιθανότητας και αυτό συνήθως οφείλεται στο ότι η κατανομή των κλάσεων είναι σημαντικά στρεβλή (μη συμμετρική) και το κόστος ενός τύπου εσφαλμένης ταξινόμησης (False Negative) είναι πιο σημαντικό από ένα άλλο είδος εσφαλμένης ταξινόμησης (False Positive).

Η απόκτηση βέλτιστων ορίων πιθανότητας χρησιμοποιώντας τις καμπύλες Precision-Recall ή ROC Curves είναι ένας τρόπος μεγιστοποίησης της προγνωστικής ικανότητας του μοντέλου μηχανικής εκμάθησης. Μια καμπύλη Precision-Recall είναι μια διαγνωστική γραφική παράσταση που αξιολογεί ένα σύνολο προβλέψεων πιθανοτήτων σε ένα σύνολο δεδομένων δοκιμής (test set) αλλά εστιάζει αποκλειστικά στην απόδοση του ταξινομητή στη θετική κλάση μειοψηφίας.^[48]

Κατά το στάδιο αξιολόγησης του συστήματος στα δεδομένα ελέγχου, πριν τον υπολογισμό όλων των μετρικών, χρίθηκε αναγκαία η διερεύνηση του βέλτιστου κατωφλίου απόφασης. Πιο συγκεκριμένα, υπολογίζεται ένα όριο για κάθε batch εικόνων, χρησιμοποιώντας τη συνάρτηση `sklearn.metrics.precision_recall_curve` που παίρνει ως είσοδο τις πραγματικές τιμές εξόδου και τις πιθανότητες για την θετική κλάση και επιστρέφει τα precision, recall και threshold values. Τελικά, θέλουμε να προσδιορίσουμε αυτό το όριο που δίνει την πάνω δεξιά γωνία της καμπύλης. Αυτό υπολογίζεται ως εξής:

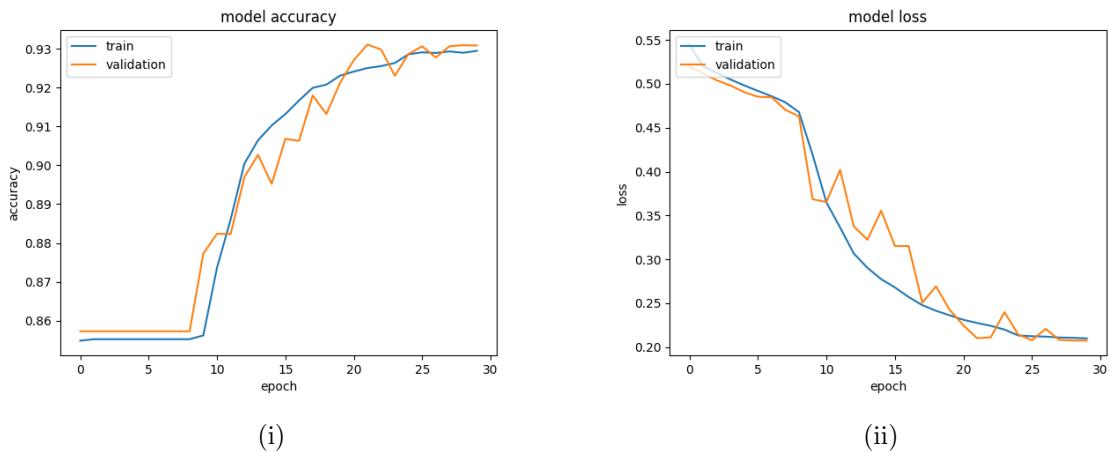
$$f\text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{index} = \text{argmax}(f\text{score})$$

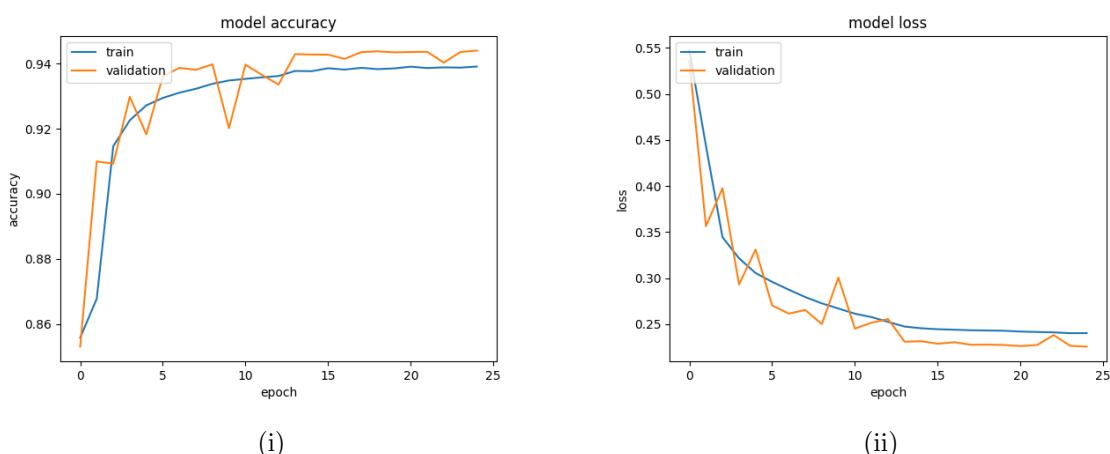
$$\text{threshold_optimal} = \text{thresholds}[\text{index}]$$

5.1.3.3 Αποτελέσματα Αξιολόγησης στα Δεδομένα Εκπαίδευσης

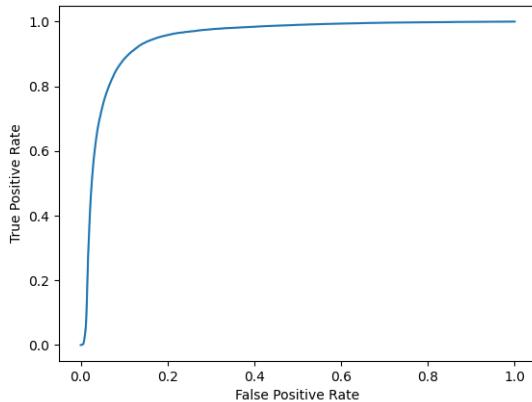
Στην ενότητα αυτή, απεικονίζεται η σύγκριση των τιμών ακρίβειας εκπαίδευσης (Training Accuracy) και ακρίβειας επαλήθευσης (Validation Accuracy) καθώς και η σύγκριση των τιμών σφάλματος (Loss) εκπαίδευσης και επαλήθευσης κατά την πάροδο των εποχών εκπαίδευσης (5.2). Το πρώτο σχήμα προκύπτει από τη χρήση της αρχικής αρχιτεκτονικής όπως προτάθηκε (4.1) ενώ το δεύτερο μετά την προσθήκη των επιπέδων Batch Normalization *Plain_Net_BN* (4.5). Ακολουθεί η καμπύλη ROC της τελευταίας στο σχήμα 5.14(ii) που δείχνει πώς αλλάζει η ευαισθησία όταν αλλάζει η ειδικότητα και βλέπουμε ότι έχουμε μια πολύ καλή απόδοση εφόσον τείνει να μοιάζει με ορθή γωνία.



Σχήμα 5.2: Accuracy του *Plain_Net* για 300,000 training patches και batch size = 64

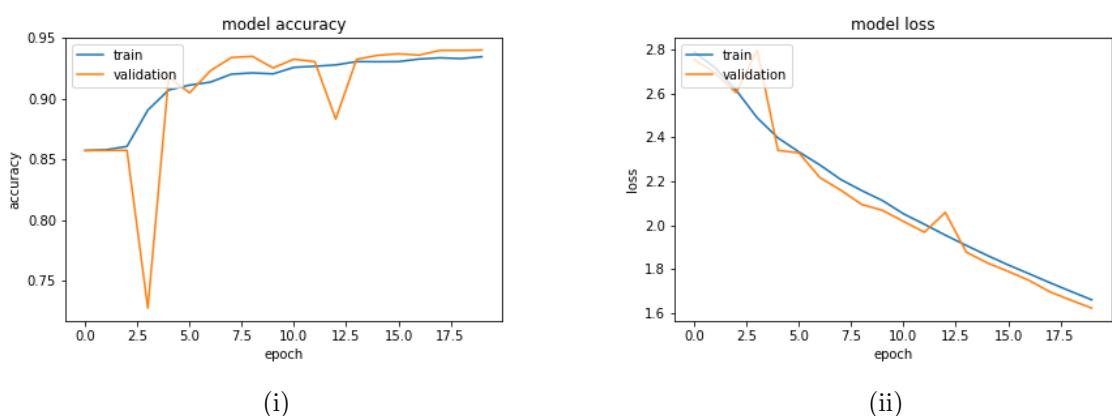


Σχήμα 5.3: Accuracy του *Plain_Net* με *Batch_Normalization* για 300,000 training patches και batch size = 128

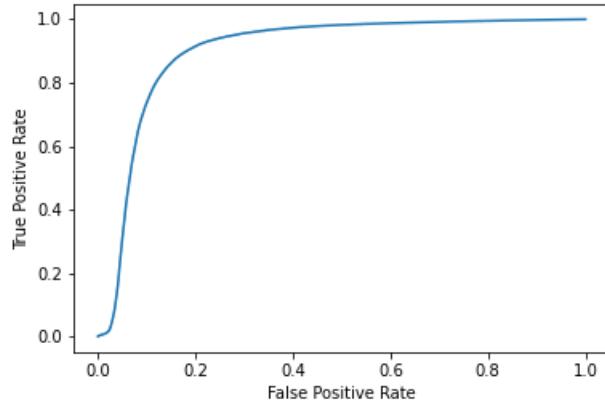


Σχήμα 5.4: ROC του Plain_Net με Batch_Normalization για 300,000 training patches και batch size = 128

Για το μοντέλο και δίκτυο No-Pool-Net, απεικονίζεται η σύγκριση των τιμών ακρίβειας εκπαίδευσης (Training Accuracy) και ακρίβειας επαλήθευσης (Validation Accuracy) καθώς και η σύγκριση των τιμών σφάλματος (Loss) κατά την πάροδο των εποχών εκπαίδευσης (σχήμα 5.5), καθώς και η καμπύλη ROC για τα δεδομένα εκπαίδευσης (σχήμα 5.6).



Σχήμα 5.5: Accuracy του No_Pool_Net με Batch_Normalization για 300,000 training patches και batch size = 64



Σχήμα 5.6: ROC του *No_Pool_Net* με *Batch_Normalization* για 300,000 training patches και batch size = 64

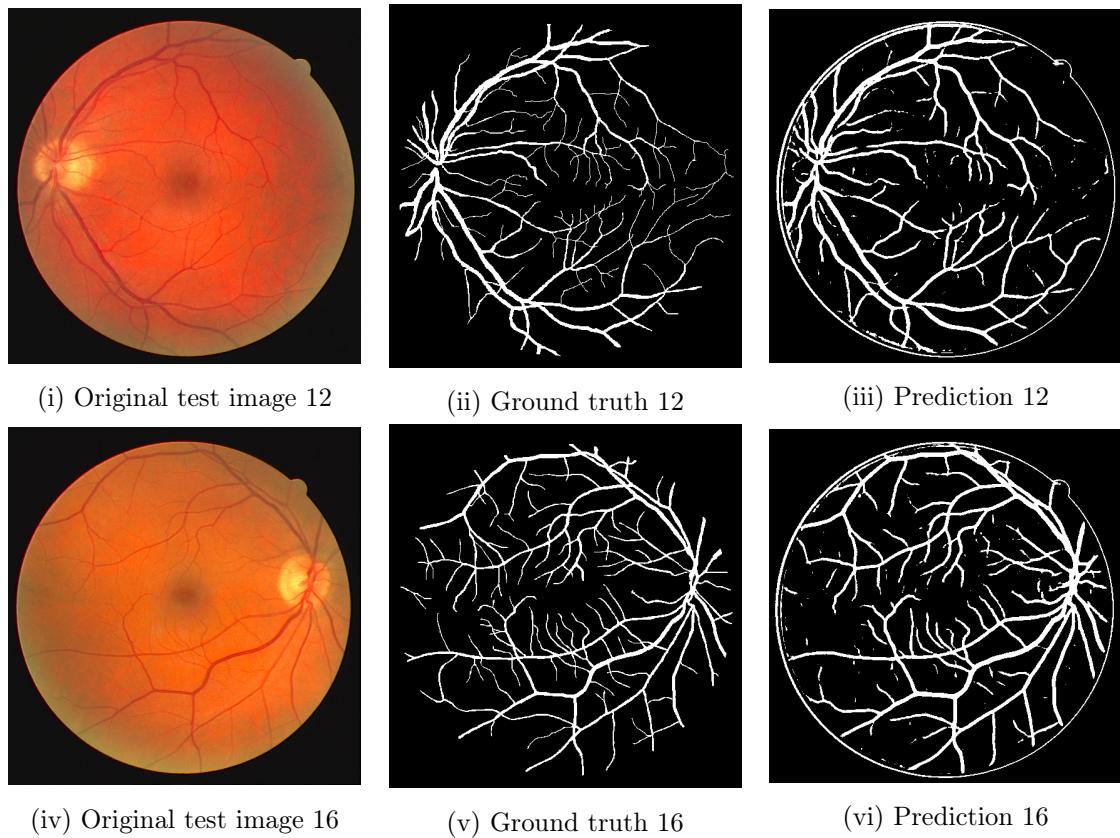
Συγχριτικά, για τα παραπάνω μοντέλα οδηγούμαστε, στο τέλος της εκπαίδευσης, στις παραχάτω βέλτιστες τιμές validation accuracy, και validation loss:

Model	Validation Loss	Validation Accuracy
Plain-Net	0.2158	0.9303
Plain-Net-BN	0.2402	0.9441
No-Pool-Net	1.5	0.9446

Πίνακας 5.1: Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net

5.1.3.4 Αποτελέσματα Αξιολόγησης στα Δεδομένα Ελέγχου

Στο σημείο αυτό, θα δούμε την επόδοση και τα αποτελέσματα των παραπάνω μοντέλων στα δεδομένα ελέγχου. Στον πίνακα 5.2 παρουσιάζονται, για την προτεινόμενη βιβλιογραφική μέθοδο Plain-Net αρχικά, οι τελικές τιμές επίδοσης που πετυχαίνει το εκπαιδευμένο μοντέλο σε κάθε μία από τις εικόνες που ανήκουν στα δεδομένα ελέγχου. Στο σχήμα 5.7, παραθέτονται κάποια παραδείγματα πρόβλεψης κατάτμησης των αγγείων σε δύο από τις εικόνες ελέγχου σε σύγκριση με τις δεδομένες εικόνες αληθείας ground truth από το βελτιστοποιημένο μοντέλο Plain-Net-BN. Στη συνέχεια, στους πίνακες 5.8 παρουσιάζονται και τα classification reports των παραδειγμάτων αυτών ώστε να αξιολογηθεί καλύτερα η απόδοση του συστήματος στα 'άγνωστα' δεδομένα.



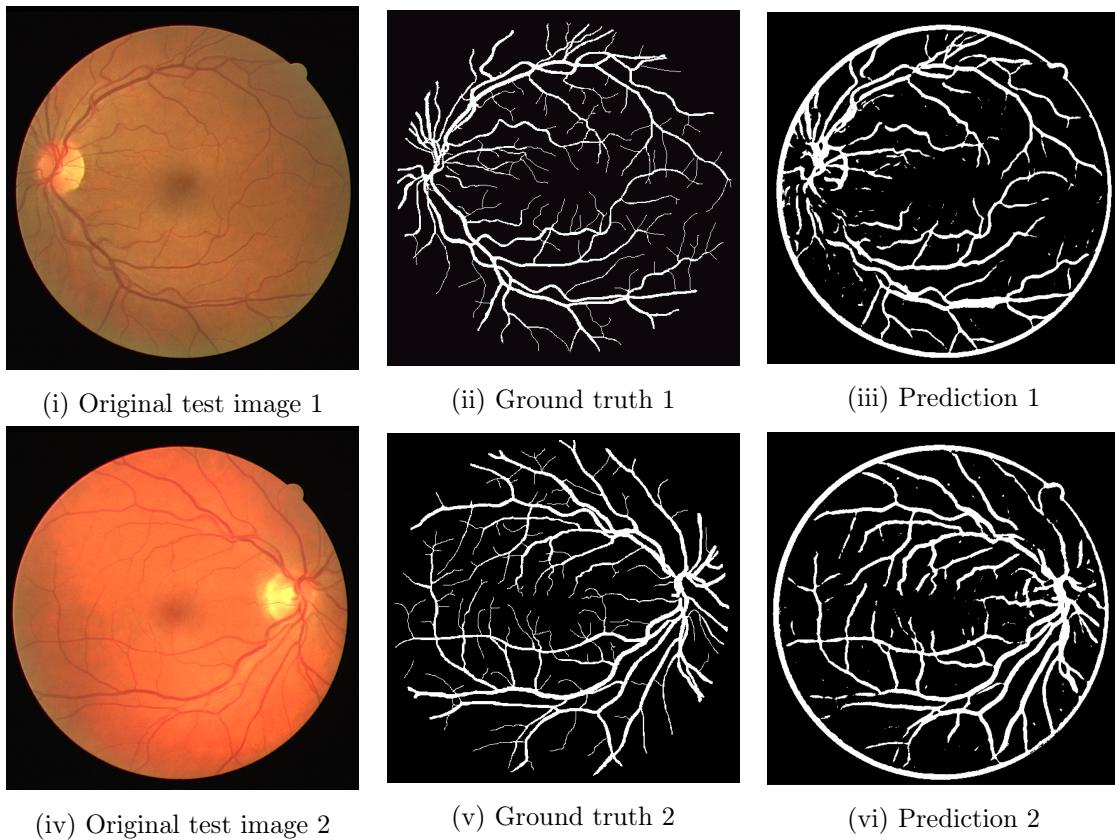
Σχήμα 5.7: Ground truth segmentation testing εικόνας και αντίστοιχο αποτέλεσμα πρόβλεψης από το μοντέλο Plain_Net_BN

	precision	recall	f1-score	support
0	0.9682	0.9409	0.9543	200920
1	0.6335	0.7678	0.6942	26740
accuracy	0.9206	0.9206	0.9206	0.9206
macro avg	0.8009	0.8543	0.8243	227660
weighted avg	0.9289	0.9206	0.9238	227660

	precision	recall	f1-score	support
0	0.9683	0.9442	0.9561	199115
1	0.6681	0.7842	0.7214	28490
accuracy	0.9242	0.9242	0.9242	0.9242
macro avg	0.8182	0.8642	0.8388	227605
weighted avg	0.9307	0.9242	0.9268	227605

Σχήμα 5.8: Classification Reports των εικόνων από τα δεδομένα ελέγχου του σχήματος 5.7

Στο σημείο αυτό, θα παρουσιαστούν οι αντίστοιχες μετρικές και αποτελέσματα για το μοντέλο No_Pool_Net, ξεκινώντας με κάποια παραδείγματα πρόβλεψης κατάτμησης των αγγείων σε δύο από τις εικόνες ελέγχου σε σύγκριση με τις δεδομένες τιμές αληθείας ground truth (5.9). Έπειτα, ακολουθούν τα classification reports των παραδειγμάτων αυτών για την περιγραφή της επίδοσης σε νέες εικόνες για το δίκτυο στους πίνακες 5.10.



Σχήμα 5.9: Ground truth segmentation testing εικόνας και αντίστοιχο αποτέλεσμα πρόβλεψης από το μοντέλο *No_Pool_Net*

	precision	recall	f1-score	support
0	0.9597	0.882	0.9192	191364
1	0.5412	0.7899	0.6423	33723
accuracy	0.8682	0.8682	0.8682	0.8682
macro avg	0.7505	0.836	0.7808	225087
weighted avg	0.8971	0.8682	0.8777	225087

	precision	recall	f1-score	support
0	0.9643	0.8872	0.9241	194965
1	0.5113	0.7827	0.6185	29412
accuracy	0.8735	0.8735	0.8735	0.8735
macro avg	0.7378	0.8349	0.7713	224377
weighted avg	0.905	0.8735	0.8841	224377

Σχήμα 5.10: Classification Reports των εικόνων από τα δεδομένα ελέγχου του σχήματος 5.9

Τέλος, θα παρουσιαστούν οι τιμές επίδοσης στα δεδομένα ελέγχου, συγχριτικά, και για τις δύο μεθόδους του Plain-Net και για το No-Pool-Net στον πίνακα 5.3. Οι τιμές έχουν υπολογιστεί ως μέσος όρος μεταξύ των τιμών των 20 εικόνων του test set.

<i>Test Image</i>	<i>IOU</i>	<i>Weighted Avg Acc</i>	<i>SP</i>	<i>SE</i>	<i>Precision</i>
1	0.864	0.917	0.932	0.804	0.610
2	0.857	0.912	0.923	0.838	0.608
3	0.872	0.924	0.947	0.761	0.670
4	0.854	0.908	0.915	0.856	0.565
5	0.853	0.905	0.909	0.875	0.535
6	0.870	0.922	0.935	0.836	0.660
7	0.844	0.903	0.912	0.840	0.588
8	0.875	0.931	0.967	0.722	0.796
9	0.847	0.906	0.913	0.860	0.602
10	0.858	0.916	0.939	0.774	0.677
11	0.855	0.911	0.919	0.859	0.615
12	0.867	0.916	0.919	0.886	0.559
13	0.871	0.924	0.942	0.809	0.688
14	0.855	0.919	0.977	0.519	0.763
15	0.832	0.895	0.905	0.836	0.593
16	0.856	0.913	0.926	0.826	0.630
17	0.862	0.915	0.927	0.829	0.607
18	0.883	0.933	0.966	0.691	0.729
19	0.862	0.923	0.967	0.662	0.774
20	0.845	0.901	0.900	0.906	0.554

Πίνακας 5.2: Οι τιμές που λήφθηκαν για μία σειρά μετρικών στο στάδιο αξιολόγησης του Plain-Net

Μοντέλο	Acc %	SP	SE	Pr
Plain-Net	0.9147	0.9320	0.7995	0.6412
Plain-Net-BN	0.9226	0.9503	0.7139	0.6597
No-Pool-Net	0.9135	0.9389	0.7452	0.6494

Πίνακας 5.3: Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net

Η υλοποίηση No-POOL-Net δεν χρησιμοποιεί Max-pooling, και σύμφωνα με παλιότερα αποτελέσματα, τα δίκτυα που δεν κάνουν περαιτέρω δειγματοληψία των χαρακτηριστικών ενδέχεται να έχουν καλύτερη απόδοση όταν εφαρμόζονται σε μικρές εικόνες. Παρόλαυτα, το δίκτυο που είχε τη βέλτιστη απόδοση από τα τρία, ήταν το Plain-Net με Batch Normalization.

5.2 Σύστημα Κατάτμησης Τεμαχίων Εικόνας

Στην ενότητα αυτή ωστε πειριγραφεί η πειραματική διαδικασία εκπαίδευσης και επαλήθευσης ενός image segmentation συστήματος, που αποτελείται από ένα πλήρως συνελιτικό δίκτυο (FCNN), με σκοπό την επιλογή των παραμέτρων και των τεχνικών εκείνων που ωστε πειρέρουν την καλύτερη επίδοση του μοντέλου στο πρόβλημα της κατάτμησης των αιμοφόρων αγγείων του αμφιβληστροειδούς. Η διεξαγωγή των πειραμάτων βασίζεται στην αρχιτεκτονική *Sine – Net* και στη λογική που παρουσιάστηκε από τους I. Atli, O.S. Gedik [43] (Κεφάλαιο 4.2).

Τα διαθέσιμα δεδομένα εκπαίδευσης για το συγκεκριμένο πρόβλημα αποτελούνται από τις 20 εικόνες του training set, από τις οποίες παίρνουμε ως είσοδο τις προεπεξεργασμένες εικόνες που έχουν υποστεί data augmentation. Η διαδικασία του data augmentation που ακολουθήθηκε για την επαύξηση των εικόνων περιγράφηκε στην ενότητα 3.1.6 στην παράγραφο σχετικά με τη δεύτερη μεθοδολογία της εργασίας, και υλοποιήθηκε όπως ωστε πειριγραφεί στην επόμενη παράγραφο.

5.2.1 Εκπαίδευση Δικτύου

Για την εκπαίδευση του μοντέλου Sine-Net ωστε χρησιμοποιηθεί η αρχιτεκτονική του δικτύου που περιγράφηκε στην ενότητα 4.2. Για την εκπαίδευση του δικτύου χρησιμοποιήθηκαν δύο βελτιστοποιητές: αρχικά o Stochastic Gradient Descent (SGD) για 60 εποχές και στη συνέχεια o βελτιστοποιητής Adam για άλλες 40 εποχές, οπότε και το δίκτυο εκπαιδεύτηκε συνολικά για 100 εποχές. Σε κάθε εποχή ανακατανείμαμε το training set με *shuffle* = *True* αλλά όχι το 10% αυτού, το οποίο χρησιμοποιήθηκε ως validation set.

Η προεπεξεργασία δεδομένων ξεκινάει με τον μετασχηματισμό-ενίσχυση των εικόνων εκπαίδευσης με την εφαρμογή του αλγορίθμου MHT και στη συνέχεια του αλγορίθμου CLAHE σε κάθε εικόνα ξεχωριστά. Στη συνέχεια, τα δεδομένα υπόκεινται σε μια ακολουθία μετασχηματισμών επαύξησης που συμβαίνει αρχικά offline, σε δύο ξεχωριστά βήματα.

Στο πρώτο βήμα η κάθε εικόνα υπόκεινται 1 φορά σε τυχαίο συνδυασμό μετασχηματισμών περιστροφής, μεταπότισης και κατοπτρισμού με τυχαίες μεταβλητές, ως εξής:

```
transformations_step1 = Compose([
    Rotate((-45°, 45°), interpolation = linear),
    Affine(translate_px = (0, 117), interpolation = linear),
    OneOf(VerticalFlip, HorizontalFlip)]) (5.4)
```

Στο δεύτερο βήμα η κάθε εικόνα υπόκεινται 8 φορές στο συνδυασμό μετασχηματισμών περιστροφής και κατοπτρισμού, με τυχαία γωνία περιστροφής και τυχαίο άξονα κατοπτρισμού, ως εξής:

$$\begin{aligned}
 transformations_{step2} = Compose([\\
 & Rotate((-90^\circ, 90^\circ), interpolation = linear, \\
 & OneOf(VerticalFlip(), HorizontalFlip())])
 \end{aligned} \tag{5.5}$$

Μετά το πέρας της διαδικασίας δημιουργίας μεταποιημένων εικόνων, οι εικόνες θα κανονικοποιηθούν με τον αλγόριθμο Standard Scaling ώστε να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση.

Τελικά, κατά την τροφοδότηση των εικόνων σε mini-batches μέσω ενός DataLoader στην είσοδο του δικτύου για την εκπαίδευσή του, πραγματοποιούνται εκ νέου τυχαίοι συνδυασμοί μετασχηματισμών από αυτούς που περιγράφηκαν παραπάνω. Αυτό έχει ως αποτέλεσμα το δίκτυο να βλέπει ένα συνολικό αριθμό εικόνων ίσο με τον αριθμό όσων δημιουργήθηκαν στο offline augmentation πολλαπλασιασμένο με τον αριθμό των εποχών. Να σημειωθεί ότι σε όλους τους μετασχηματισμούς περιστροφής και μετατόπισης χρησιμοποιείται γραμμική συνάρτηση interpolation ή αλλιώς παρεμβολή ώστε να μην χάνεται πολύτιμη πληροφορία από τυχόν pixels της εικόνας που κατά την περιστροφή της βρεθούν εκτός. Η παρεμβολή λειτουργεί χρησιμοποιώντας γνωστές τιμές των γειτονικών pixels για την εκτίμηση προσεγγιστικών τιμών έντασης στα άγνωστα pixels.

Όσον αφορά στο στάδιο της αξιολόγησης στα δεδομένα του test set, πριν την τροφοδότηση αυτών στο δίκτυο *Sine – Net*, προηγουμένως έχουμε χωρίσει κάθε εικόνα μεγέθους 565×565 στον ελάχιστον αριθμό patches, μεγέθους 448×448 , όσο δηλαδή και το απαραίτημα μέγεθος της εισόδου του δικτύου. Ο αριθμός των δημιουργούμενων patches είναι τέσσερα (4) και έτσι, για τις επικαλυπτόμενες περιοχές που προκύπτουν μεταξύ τους, παίρνουμε το σταθμισμένο μέσο όρο των τεσσάρων προβλέψεων που δίνει ως έξοδο το δίκτυο μας.

Για την υλοποίηση των βαθιών νευρωνικών δικτύων της κατηγορίας αυτής και του κεφαλαίου αυτού χρησιμοποιήθηκε το PyTorch. Το PyTorch είναι μια βελτιστοποιημένη βιβλιοθήκη μηχανικής εκμάθησης ανοιχτού κώδικα, από tensors (ταυστές) Deep Learning, που βασίζεται στην Python και στη βιβλιοθήκη Torch και χρησιμοποιείται κυρίως για εφαρμογές που χρησιμοποιούν GPU και CPU [49].

Ακολουθούν οι τιμές που έλαβαν οι παράμετροι του δικτύου:

- Epochs = 100 (SGD: 60 & Adam: 40)
- Batch size = 10
- Optimizer = SGD
 - Learning rate = 10^{-4}
 - Momentum = 0.9
 - nesterov = True
- Optimizer = Adam

- Learning rate = 10^{-6}
- Hyperparameters $\beta_1 = 0.9$ & $\beta_2 = 0.999$
- epsilon = 10^{-4}
- Learning rate reduction = *ReduceLRonPlateau()*
 - monitor="val_loss"
 - patience = 20
 - Step decay = 0.1

Οι συναρτήσεις κόστους που δοκιμάστηκαν κατά την πειραματική διαδικασία θα παρουσιαστούν στην επόμενη παράγραφο.

5.2.2 Προκλήσεις κατά την πειραματική διαδικασία και βελτιστοποίηση του δικτύου

5.2.2.1 Ανισοροπία Κλάσεων (Class Imbalance)

Σε συνέχεια της ενότητας 5.1.2.1 ο μηχανισμός ελέγχου του προβλήματος της ανισοροπίας, που δοκιμάστηκε κατά την διαδικασία εκπαίδευσης του δικτύου Sine-Net, ήταν η χρήση μετρικών σφάλματος, όπως η Weighted Binary Cross Entropy, Dice Loss, Tversky Loss, Focal Loss ή πιθανοί συνδυασμοί τους.

Η κύρια δυσκολία της κανονικής binary cross-entropy (BCE) είναι ότι λαμβάνει υπόψη όλα τα εικονοστοιχεία εξίσου κατά τον υπολογισμό της απώλειας. Σε μια μάσκα όπου το 90% των εικονοστοιχείων είναι 0 και μόνο το 10% είναι 1, το δίκτυο λαμβάνει χαμηλή απώλεια ακόμα και αν χάσει όλα τα 1, πράγμα που σημαίνει ότι το δίκτυο δεν μαθαίνει τίποτα. Στην παρούσα εργασία, για το σύστημα Sine_Net δοκιμάστηκαν 4 από τους προαναφερθέντες τρόπους:

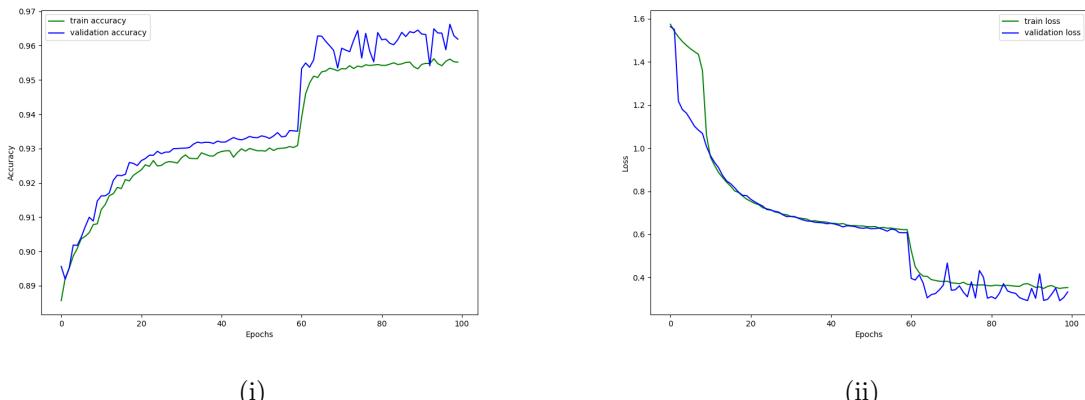
- **Dice Binary Cross Entropy Loss (5.11):** Γραμμικός συνδυασμός των Dice loss και Binary Crossentropy loss with logits. Εδώ χρησιμοποιήθηκε η απλή πρόσθεσή τους. Το Dice Coefficient, που έχει ενσωματωθεί στο Dice Loss, είναι ένας ευρέως χρησιμοποιούμενος τρόπος για τον υπολογισμό της ομοιότητας ή επικάλυψης μεταξύ της πρόβλεψης και του ground truth, παρόμοιος με την ευριστική *Intersection – over – Union*. Σε αντίθεση με το BCE, λαμβάνει υπόψη μόνο την κλάση τημηματοποίησης και όχι την κλάση του φόντου.
- **Dice Binary Cross Entropy Loss with Balanced Sampler (5.12):** Τύπος loss function με το παραπάνω αλλά το αποτέλεσμα περιλαμβάνει και τη βοήθεια από balanced sampling στα δείγματα εισόδου.
- **Weighted Dice Binary Cross Entropy Loss (5.13):** Προσπαθεί να λύσει το πρόβλημα του BCE σταθμίζοντας περισσότερο τη θετική (μικρότερη) κλάση, ώστε να είναι δυσκολότερο να την αγνοήσει.

- **Tversky Loss:** Ο δείκτης Τερσκψ (TI) είναι ένα μέτρο ασύμμετρης ομοιότητας που είναι μια γενίκευση του Dice Coefficient και του δείκτη Θαξαρδ. Όταν η παράμετρος $a > b$, τότε τα Φαλσε Νεγατιες τιμωρούνται περισσότερο, γι' αυτό και θέσαμε: $b = 0.90985$ & $a = 0.09015$.
- **Focal Loss:** Εφαρμόζει έναν διαμορφωτικό όρο στο cross-entropy loss προκειμένου να επικεντρώνει η μάθηση στα σοβαρώς λανθασμένα παραδείγματα. Είναι ένα δυναμικά κλιμακούμενο cross-entropy loss, όπου ο συντελεστής κλιμάκωσης μειώνεται στο μηδέν καθώς αυξάνεται η εμπιστοσύνη στη σωστή τάξη.

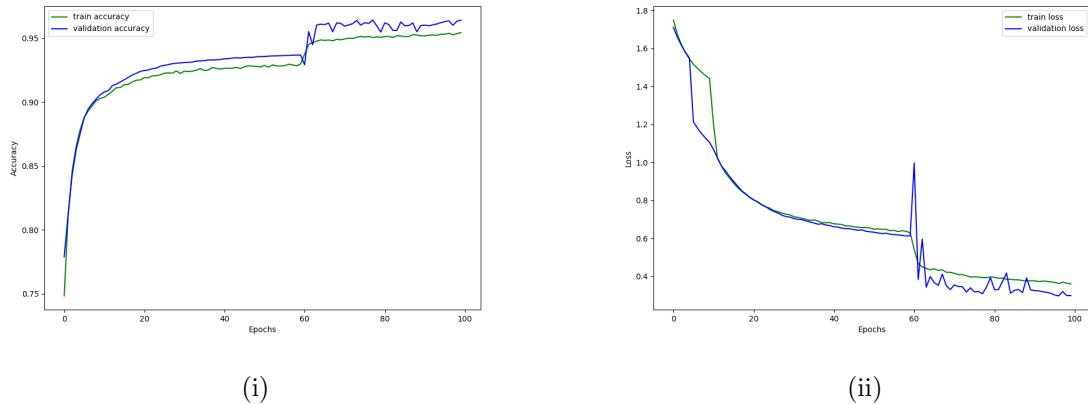
5.2.3 Αξιολόγηση Συστήματος

5.2.3.1 Αποτελέσματα Αξιολόγησης στα Δεδομένα Εκπαίδευσης

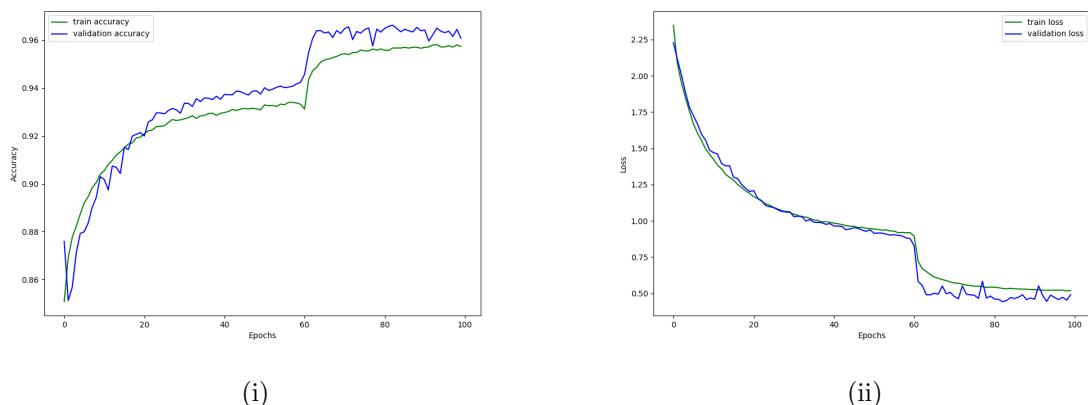
Στη συνέχεια, θα παρουσιάσουμε τις γραφικές παραστάσεις Accuracy και Loss από τα πειράματα που πραγματοποιήσαμε μεταβάλοντας τη συνάρτηση κόστους loss function και τη δειγματοληψία των κλάσεων στα δεδομένα εκπαίδευσης ώστε να λαμβάνονται δείγματα από αυτές ισότιμα και να προκύπτει, έτσι, ένα balanced training set.



Σχήμα 5.11: Accuracy (i) και Loss (ii) του Sine_Net_1 με Dice Binary CrossEntropy Loss

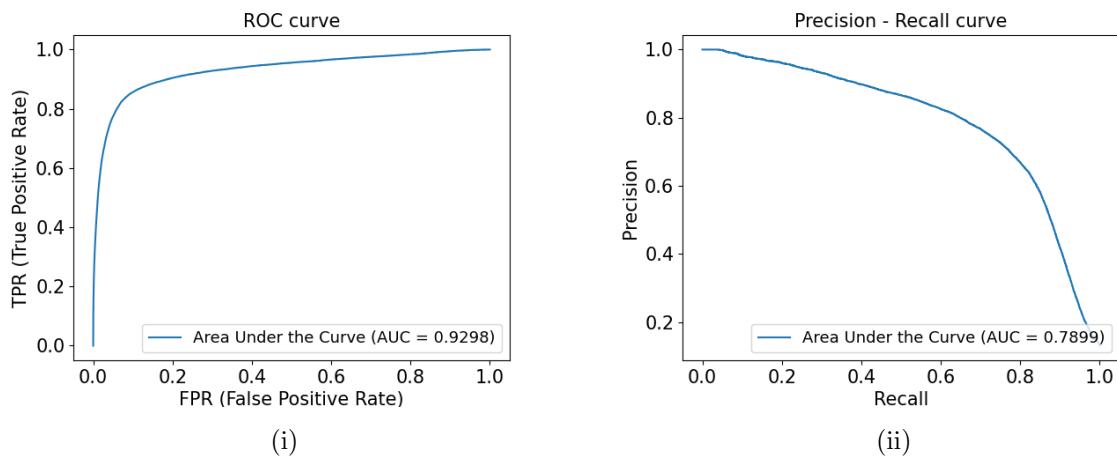


Σχήμα 5.12: Accuracy (i) και Loss (ii) του *Sine_Net_2* με Balanced Sampler και Dice Binary CrossEntropy Loss



Σχήμα 5.13: Accuracy (i) και Loss (ii) του *Sine_Net_3* με Weighted Dice Binary CrossEntropy Loss

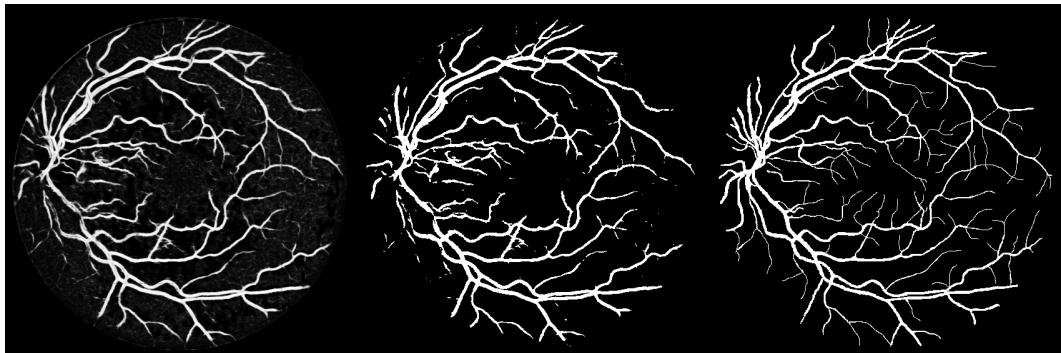
Όπως παρατηρούμε η συνάρτηση loss που έφερε τα καλύτερα αποτελέσματα όταν η Dice Binary Cross Entropy loss του γραφήματος *Sine_Net_1*. Για το μοντέλο αυτό, η καμπύλη ROC φαίνεται παρακάτω:



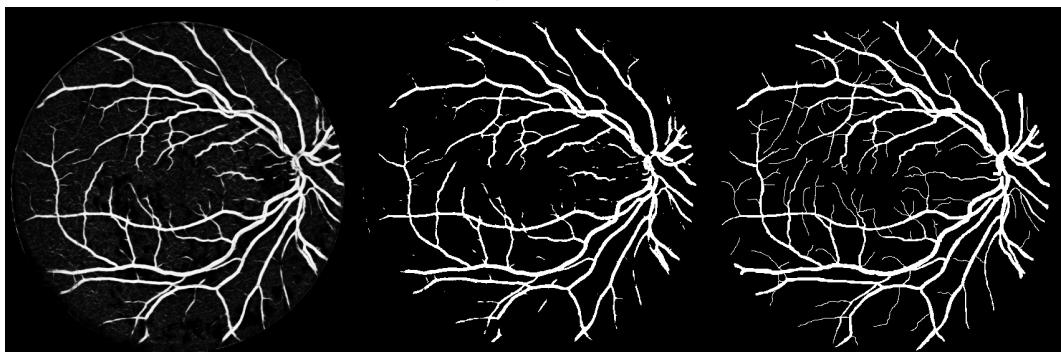
Σχήμα 5.14: ROC-Curve (i) PR-Curve (ii) του *Sine_Net*

5.2.3.2 Αποτελέσματα Αξιολόγησης στα Δεδομένα Ελέγχου

Στα παρακάτω γραφήματα (5.4, 5.5 και 5.15) παρουσιάζονται τα αποτελέσματα που πέτυχε το μοντέλο *Sine – Net* στα άγνωστα δεδομένα του test set:



(i) Prediction output (i) Binary prediction output (ii) και Ground truth (iii) του μοντέλου *Sine_Net* για την εικόνα 1 του test set



(ii) Prediction output (i) Binary prediction output (ii) και Ground truth (iii) του μοντέλου *Sine_Net* για την εικόνα 2 του test set

Σχήμα 5.15

<i>Test Image</i>	<i>AUC ROC</i>	<i>AUC PR</i>	<i>F1 score</i>	<i>Accuracy</i>	<i>SE</i>	<i>SP</i>	<i>Precision</i>
1	0.930	0.790	0.737	0.932	0.729	0.962	0.744
2	0.934	0.854	0.781	0.939	0.724	0.977	0.849
3	0.905	0.762	0.692	0.920	0.616	0.972	0.788
4	0.906	0.801	0.733	0.937	0.650	0.981	0.840
5	0.906	0.781	0.699	0.931	0.593	0.984	0.851
6	0.892	0.760	0.688	0.925	0.588	0.980	0.828
7	0.913	0.806	0.735	0.938	0.645	0.983	0.854
8	0.886	0.741	0.661	0.930	0.540	0.987	0.853
9	0.873	0.721	0.641	0.931	0.522	0.986	0.830
10	0.915	0.781	0.723	0.939	0.663	0.977	0.795
11	0.919	0.799	0.734	0.936	0.687	0.973	0.789
12	0.900	0.725	0.658	0.925	0.573	0.976	0.772
13	0.902	0.780	0.704	0.927	0.611	0.979	0.829
14	0.939	0.801	0.741	0.941	0.709	0.973	0.775
15	0.940	0.798	0.744	0.947	0.743	0.971	0.745
16	0.925	0.792	0.725	0.933	0.672	0.973	0.786
17	0.905	0.739	0.666	0.929	0.574	0.979	0.794
18	0.902	0.725	0.680	0.932	0.631	0.971	0.737
19	0.944	0.842	0.777	0.945	0.792	0.966	0.762
20	0.915	0.737	0.691	0.938	0.647	0.973	0.740

Πίνακας 5.4: Οι τιμές που λήφθηκαν για μία σειρά μετρικών στο στάδιο αξιολόγησης του Sine-Net

Μοντέλο	Acc	SP	SE	Pr
Plain-Net	0.9147	0.9320	0.7995	0.6412
Plain-Net-BN	0.9226	0.9503	0.7139	0.6597
No-Pool-Net	0.9135	0.9389	0.7452	0.6494
Sine-Net	0.9338	0.9760	0.6454	0.7981

Πίνακας 5.5: Τιμές επίδοσης μοντέλων Plain-Net, Plain-Net-BN και No-Pool-Net και Sine-Net

5.2.3.3 Βελτιστοποίηση Αποτελεσμάτων Κατάτμησης μέσω Διαδικασίας Post Processing

Λόγω της πολυπλοκότητας της εικόνας του βυθού του αμφιβληστροειδούς, υπάρχουν ορισμένα προβλήματα με την ακριβή κατάτμηση των αγγείων του αμφιβληστροειδούς. Δύο τέτοια βασικά ζητήματα είναι η απώλεια των λεπτών αγγείων και η λανθασμένη ανίχνευση σε ‘πολύπλοκες’ περιοχές. Στην ενότητα αυτή θα παρουσιαστεί μια προσπάθεια που έγινε για την βελτιστοποίηση

των τμηματοποιημένων εικόνων που προκύπτουν από τις προβλέψεις του δικτύου *Sine – Net* στις εικόνες του test set. Η διαδικασία βασίστηκε στη χρήση μετασχηματισμών και επεξεργασίας εικόνας και στη δουλειά των Nguyen et al. [50] και Zhou et al. [51].

Η υποκείμενη τεχνική της προτεινόμενης μεθόδου είναι ένας γραφικός συνδυασμός ανιχνευτών γραμμής line detectors σε διαφορετικές κλίμακες για την παραγωγή της τμηματοποίησης του αγγείου για κάθε εικόνα αμφιβληστροειδούς. Ένας βασικός ανιχνευτής γραμμής χρησιμοποιεί ένα σύνολο από κατά προσέγγιση περιστρεφόμενες ευθείες γραμμές για να ανιχνεύσει τα αγγεία σε διαφορετικές γωνίες. Η διαφορά μεταξύ του μέσου επιπέδου γκρι της νικητήριας γραμμής (της γραμμής με το μέγιστο μέσο γκρι επίπεδο) και του μέσου επιπέδου γκρι του περιβάλλοντος παραθύρου παρέχει ένα μέτρο της «αγγειότητας» κάθε εικονοστοιχείου εικόνας.

Ο βασικός ανιχνευτής γραμμής λειτουργεί στο ανεστραμμένο πράσινο χανάλι μιας εικόνας αμφιβληστροειδούς. Σε κάθε ύφεση pixel, ορίζεται ένα παράθυρο μεγέθους $W \times W$ pixel και το μέσο επίπεδο γκρι υπολογίζεται ως I_{avg}^W . Ορίζονται δώδεκα γραμμές μήκους W εικονοστοιχείων που προσανατολίζονται σε 12 διαφορετικές κατευθύνσεις (γωνιακή ανάλυση 15°) που διέρχονται από τα κεντρικά εικονοστοιχεία και υπολογίζεται η μέση τιμή των επιπέδων γκρι των εικονοστοιχείων που βρίσκονται κατά μήκος κάθε γραμμής. Η γραμμή με τη μέγιστη τιμή ονομάζεται «γραμμή που κερδίζει» και η τιμή της ορίζεται ως I_{max}^W . Η απόχριση γραμμής σε ένα pixel υπολογίζεται στη συνέχεια ως:

$$R_W = I_{max}^W - I_{avg}^W \quad (5.6)$$

Έπειτα, γενικεύουμε τον βασικό ανιχνευτή γραμμής μεταβάλλοντας το μήκος των ευθυγραμμισμένων γραμμών. Με την αλλαγή των τιμών του μήκους L , επιτυγχάνονται ανιχνευτές γραμμής σε διαφορετικές κλίμακες. Ο γενικευμένος ανιχνευτής γραμμής ορίζεται ως:

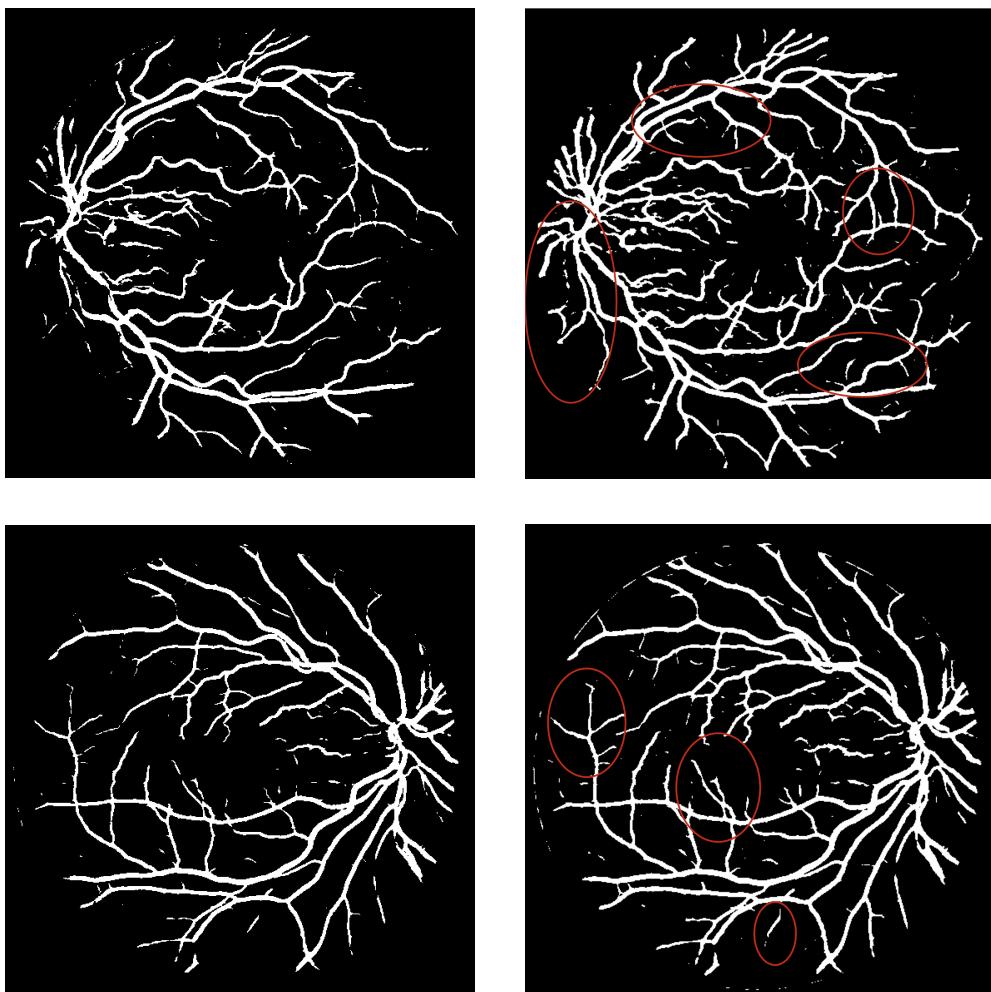
$$R_W^L = I_{max}^L - I_{avg}^L \quad (5.7)$$

Στο στάδιο του συνδυασμού, αποδίδεται το ίδιο βάρος για κάθε κλίμακα και η τελική τμηματοποίηση είναι ο γραφικός συνδυασμός αποκρίσεων γραμμής διαφορετικών κλιμάκων. Με το συνδυασμό αυτό διατηρούνται οι δυνατότητες των μεμονωμένων ανιχνευτών ενώ εξαλείφονται τα μειονεκτήματά τους. Η απόχριση σε κάθε εικονοστοιχείο εικόνας ορίζεται ως:

$$R_{combined} = \frac{1}{n_L + 1} \sum_L R_W^L + I_{igc} \quad (5.8)$$

όπου n_L είναι ο αριθμός των κλιμάκων που χρησιμοποιούνται, R_W^L είναι η απόχριση του ανιχνευτή γραμμής στην κλίμακα L και I_{igc} είναι η τιμή του ανεστραμμένου πράσινου χαναλιού στο αντίστοιχο pixel.

Τα αποτελέσματα της μεθόδου αυτής όταν εφαρμόστηκε στις εικόνες εξόδου-πρόβλεψης του μοντέλου *Sine – Net* φαίνονται παρακάτω:



Σχήμα 5.16: Binary reduction output (i) και Line Detector output για τις εικόνες 1 (επάνω) και 2 (κάτω) του test set με επισημασμένες περιοχές με αισθητή βελτιστοποίηση.

5.3 Συνδυασμός Συστημάτων

Στην ενότητα αυτή θα διεξαχθεί μια λίγο διαφορετική διαδικασία εκπαίδευσης από τις προηγούμενες που παρουσιάστηκε και στο προηγούμενο κεφάλαιο, στην ενότητα 4.3. Η επιλογή να δοκιμαστεί η μέθοδος αυτή βασίστηκε στην ικανότητα των μεθόδων αυτών να συνδυάζουν την επίδοση πολλαπλών επιμέρους μοντέλων σε μία πιο δυνατή προγνωστική οντότητα που επιτυγχάνει καλύτερα αποτελέσματα. Τα βασικά επιμέρους μοντέλα που επιλέχθηκαν ως ‘‘αδύναμοι μαθητές’’, από όσα περιγράφηκαν στα προηγούμενα κεφάλαια, είναι το *Plain – Net – BN* και το *Sine – Net*, ένα από κάθε κατηγορία αρχιτεκτονικής, patch-based classification και FCN segmentation. Επιπλέον, τα δύο αυτά μοντέλα ήταν τα πιο υπολογιστικά διαχειρίσιμα που κατανάλωσαν τους λιγότερους πόρους. Το Model Ensemble που χρησιμοποιήθηκε στην παρούσα μελέτη είναι ένας Stacking Regressor.

Στην πράξη, χρησιμοποιώντας τη μέθοδο k-fold-cross-validation, το training set επιμερίστηκε σε 4 ίσα τμήματα ($k = 4$). Κάθε βασικό μοντέλο εκπαιδεύτηκε σειριακά σε κάθε ένα από τα 3 τμήματα, ενώ ένα από τα τμήματα αυτά χρησιμοποιήθηκε ως validation set. Τα 2 set προβλέψεων που προέκυψαν από τα out-of-fold κομμάτια, υπέστησαν στοίβαξη ή stacking και αποτέλεσαν το σύνολο εκπαίδευσης για το μετα-μοντέλο ταξινομητή στο επίπεδο-1. Αφού εκπαιδεύσαμε τα βασικά μοντέλα σε όλο το training set και τα επαληθεύσαμε, στη φάση testing σε όλο το original test set, η στοίβαση των προβλέψεων που προκλύπτουν αποτέλεσαν τα test δεδομένα του τελικού ταξινομητή.

Λόγω της υπολογιστικής πολυπλοκότητας, της ταυτόχρονης χρήσης βιβλιοθηκών όπως η Keras, και η Pytorch και των απαιτήσεων σε χωρητικότητα RAM, το Ensembling πραγματοποιήθηκε σε πολύ μικρότερο όγκο δεδομένων με αποτέλεσμα τα αποτελέσματα που επέφερε να μην είναι αξιοσημείωτα.

5.4 Σύγκριση Μεθόδων με το State-of-the-Art

Στο σημείο αυτό θα δούμε συνολικά τις επιδόσεις όλων των μεθόδων που έχουν περιγραφεί σε σύγκριση μετάξυ τους και σε σύγκριση με κάποιες από τις State-of-the-Art μεθόδους και αρχιτεκτονικές.

Database	Method	Accuracy	Specificity	Sensitivity
DRIVE	Plain-Net	0.9147	0.9320	0.7995
	Plain-Net-BN	0.9226	0.9503	0.7139
	No-Pool-Net	0.9135	0.9389	0.7452
	Sine-Net	0.9338	0.9760	0.6454
	Melinscak et al. [39]	0.9466	-	0.7276
	Fu et al. [40]	0.9470	-	0.7294
	Wang et al. [7]	0.9767	0.8173	0.9733
	Li et al. [52]	0.9527	0.7569	0.9816
	Liskowski et al. [31]	0.9535	0.7811	0.9807
	Mo et al. [53]	0.9521	0.7779	0.978
	Jiang et al. [54]	0.9624	0.754	0.9825
	Feng et al. [55]	0.9528	0.7625	0.9809
	Hu et al. [56]	0.9533	0.7772	0.9793
	Human	0.9637	0.7757	0.9819
	Atli et al.(pp) [43]	0.9685	0.826	0.9824

Πίνακας 5.6: Τιμές επίδοσης μοντέλων και σύγκριση με Στατε-οφ-τηε-Αρτ

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Επεκτάσεις

Στα προηγούμενα κεφάλαια έγινε μια προσπάθεια προσέγγισης του προβλήματος της κατάτμησης των αιμοφόρων αγγείων στην εικόνα αμφιβληστροειδούς, εφαρμόζοντας τις αρχές και τη λογική της βαθιάς μηχανικής μάθησης και συγκεκριμένα της επιβλεπόμενης μάθησης. Παρουσιάστηκαν υλοποιήσεις συστημάτων βαθιάς μηχανικής μάθησης σε συνδυασμό με κάποιες μεθοδολογίες επεξεργασίας και επαύξησης δεδομένων βασισμένες σε τεχνικές ψηφιακής επεξεργασίας εικόνας. Η διαδικασία ολοκληρώθηκε με την εκπαίδευση και την αξιολόγηση των συστημάτων που υλοποιήθηκαν για κάθε μια από τις διαφορετικές προσεγγίσεις. Στο παρόν κεφάλαιο συνοψίζεται το έργο που επιτελέσθηκε στη προκείμενη διπλωματική εργασία και καταγράφονται τα αποτελέσματα που εξήχθησαν, δίνοντας στη συνέχεια προτάσεις για μελλοντική έρευνα και ανάπτυξη.

6.1 Συμπεράσματα

Για την ανάπτυξη ενός συστήματος Βαθιάς Μηχανικής Μάθησης που θα πετυχαίνει μεγάλες επιδόσεις στο πρόβλημα της τμηματοποίησης των αγγείων στις εικόνες αμφιβληστροειδούς, πραγματοποιήθηκε σε πρώτο στάδιο εκτενής μελέτη παρεμφερών εργασιών και ερευνών με σκοπό την επιλογή των κατευθύνσεων δράσης. Από τη μελέτη αυτή προέκυψε ότι υπάρχουν δύο βασικές οικογένειες αρχιτεκτονικών νευρωνικών δικτύων πάνω στις οποίες στηρίχτηκαν οι περισσότερες μελέτες με επιμέρους τροποποιήσεις ή επεκτάσεις. Από τη μία έχουμε τη μεθοδολογία Ταξινόμησης Τεμαχίων Εικόνας (Patch Classification) με χρήση αρχιτεκτονικών Convolutional Neural Networks με τελικά Dense Layers και από την άλλη τη μεθοδολογία τμηματοποίησης τεμαχίων εικόνας (Patch Segmentation) με pixel-to-pixel classification και χρήση αρχιτεκτονικών Fully Convolutional Neural Networks. Επιπλέον, χρίναμε ότι με την ανάπτυξη ενός συστήματος που εφαρμόζει ένα συνδυασμό αυτών των δύο λογικών με τη μέθοδο της μηχανικής μάθησης Ensemble Learning θα είμαστε σε θέση να πετύχουμε καλύτερες επιδόσεις σε σχέση με τα συστήματα που αξιοποιούν ξεχωριστά κάθε μια από αυτές τις αναλύσεις.

Με γνώμονα τα παραπάνω πραγματοποιήθηκε αρχικά ξεχωριστή μελέτη και διαχείριση τριών βασικών προτεινόμενων αρχιτεκτονικών, δύο από την πρώτη κατηγορία, βασισμένες στην εργασία των Liskowski et al. [31], ονόματι Plain-Net και No-Pool-Net και μία από τη δεύτερη κατηγορία βασισμένη στην εργασία των Atli et al. [43] με όνομα Sine-Net. Αφού δοκιμάστηκαν οι προτεινόμενες αρχιτεκτονικές και τροποποιήθηκαν ή επεκτάθηκαν ώστε να δώσουν τα βέλτιστα δυνατά αποτελέσματα, σύμφωνα με όσα περιγράφηκαν στο κεφάλαιο 5, επιλέχθηκαν τα δύο βέλτιστα συστήματα, ένα από κάθε κατηγορία.

Η πρώτη προσπάθεια μελέτης που πραγματοποιήθηκε αφορούσε την δημιουργία ενός συστήματος CNN που να κατηγοριοποιεί patches. Από αυτή την μελέτη προέκυψαν τρεις αρχιτεκτονικές (*Plain-Net*, *Plain-Net-BN*, *No-Pool-Net*). Πιο συγκεκριμένα, στην αρχή δοκιμάστηκε η αρχιτεκτονική και μεθοδολογία επαύξησης δεδομένων όπως αυτές προτάθηκαν στη βιβλιογραφική εργασία *Plain - Net*, με αρκετά ικανοποιητικά επίπεδα επίδοσης, όχι τόσο καλά όσο στη βιβλιογραφία όπως ήταν όμως αναμενόμενο. Στη συνέχεια, στο μοντέλο *Plain - Net - BN* προστέθηκαν δύο (2) ενδιάμεσα επίπεδα Batch Normalization και είδαμε ότι η επίδοση βελτιώθηκε σημαντικά. Τελος, εξετάστηκε και η μέθοδος *No - Pool - Net* με επιπλέον προσθήκη ενδιάμεσων Batch Normalization layers και πάλι με ικανοποιητική επίδοση που δε ξεπέρασε όμως τις προηγούμενες και έτσι η *Plain - Net - BN* επιλέχθηκε ως η πιο αποδοτική από τις μενόδους της πρώτης λογικής.

Δύο βασικοί λόγοι που οδηγούν στα υψηλά επίπεδα απόδοσης και τελικά στην επιλογή αυτού του μοντέλου, είναι πρώτον ότι η ομαλοποίηση των εισόδων στα επίπεδα του νευρωνικού δικτύου ώστε να λάβουν ένα παρόμοιο εύρος τιμών, μέσω της κανονικοποίηση παρτίδας Batch Normalization, μπορεί να επιταχύνει τη μάθηση. Δεύτερον, το Batch Normalization μειώνει την εσωτερική μετατόπιση συμμεταβλητών (internal covariate shift) του δικτύου. Η εσωτερική μετατόπιση συμμεταβλητής είναι μια αλλαγή στην κατανομή της εισόδου ενός εσωτερικού στρώματος του νευρωνικού δικτύου. Για τους νευρώνες σε ένα εσωτερικό στρώμα, οι είσοδοι που λαμβάνονται (από το προηγούμενο στρώμα) αλλάζουν συνεχώς. Αυτό οφείλεται στους πολλαπλούς υπολογισμούς που έγιναν πριν από αυτό και στα βάρη στη διαδικασία εκπαίδευσης. Η εφαρμογή Batch Normalization διασφαλίζει ότι ο μέσος όρος και η τυπική απόκλιση των εισόδων του επιπέδου θα παραμένουν πάντα οι ίδιες. Έτσι, το μέγεθος της αλλαγής στην κατανομή της εισόδου των στρωμάτων μειώνεται. Τα βαθύτερα στρώματα έχουν πιο εύρωστο έδαφος για το ποιες θα είναι οι τιμές εισόδου, κάτι που βοηθά κατά τη διαδικασία εκμάθησης. Τέλος, φαίνεται ότι το Batch Normalization έχει ένα αποτέλεσμα κανονικοποίησης (regularization). Επειδή υπολογίζεται σε μικρές παρτίδες (mini-batches) και όχι σε ολόκληρο το σύνολο δεδομένων, η κατανομή δεδομένων του μοντέλου που παρατηρείται κάθε φορά έχει κάποιο θόρυβο. Αυτό μπορεί να λειτουργήσει ως κανονικοποιητής, το οποίο μπορεί να βοηθήσει να ξεπεραστεί το overfitting. Ωστόσο, ο θόρυβος που προστίθεται είναι αρκετά μικρός και έτσι, γενικά δεν αρκεί για να κανονικοποιήσει σωστά από μόνο του και γι' αυτό χρησιμοποιείται μαζί με το Dropout.

Η δεύτερη διαδικασία μελέτης σχετίζεται με τα Fully Convolutional Neural Networks και τις αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή. Αντικείμενο μελέτης σε αυτό το βήμα ήταν η τροφοδοσία με τεχνικές προεπεξεργασίας και επαύξησης και η επιλογή της κατάλληλης συνάρ-

τησης υπολογισμού του loss, ώστε το δίκτυο αυτό να πετύχει την καλύτερη επίδοση. Σύμφωνα με τα γραφήματα αποτελεσμάτων (5.4, 5.5 και 5.15), το δίκτυο *Sine – Net* ξεπερνά σε απόδοση τα προηγούμενα δίκτυα της προηγούμενης κατηγορίας *Lisk – Net* και αυτό μπορεί να αποδοθεί στα πλεονεκτήματα που παρουσιάζει η αρχιτεκτονική αυτή τύπου FCNN.

Το πρώτο πλεονέκτημα είναι ότι δεν περιλαμβάνει πλήρως συνδεδεμένα στρώματα στο τέλος για την ταξινόμηση αγγείων. Αυτή η ιδιότητα βελτιώνει τον χρόνο απόκρισης της εκπαίδευσης και της πρόβλεψης, επειδή γενικά πραγματοποιούνται έντονοι πολλαπλασιασμοί στα πλήρως συνδεδεμένα επίπεδα. Η πρώτη μέθοδος χρησιμοποιεί μια προσέγγιση ‘υσρόμενου παραθύρου’ γεγονός που προκαλεί πολλαπλούς περιττούς υπολογισμούς των γειτονικών εικονοστοιχείων, υποβαθμίζοντας έτσι ην απόδοση του χρόνου εκτέλεσης. Επίσης, δεδομένου ότι ο αριθμός των επιπέδων είναι μικρότερος στο *Sine_Net*, ο χρόνος τμηματοποίησης είναι μικρότερος. Γενικότερα, το *Sine_Net* είναι υπολογιστικά αποδοτικό καθώς προβλέπει όλα τα pixels ταυτόχρονα αντί να χρησιμοποιεί ένα patch για την πρόβλεψη μόνο του κεντρικού pixel. Το δεύτερο πλεονέκτημα είναι ότι η τοπική συνδεσιμότητα μειώνεται σημαντικά λόγω του μεγαλύτερου μεγέθους των 448×448 patches που δίνονται ως είσοδος. Αυτό βελτιώνει τον χρόνο πρόβλεψης για εικόνες σε αντίθεση με μικρότερου μεγέθους patches. Εντούτοις, η συνεχής αύξηση του μεγέθους των patches δεν είναι εφικτή επειδή το δίκτυο τείνει να μειώνει την απόδοση της τμηματοποίησης και χρειάζεται περισσότερη μνήμη. Το τρίτο πλεονέκτημα είναι ότι το δίκτυο μας συνεχίζει να μαθαίνει χαρακτηριστικά κατά τη διάρκεια των λειτουργιών down-sampling και up-sampling, όπως αναφέρεται και στην παράγραφο 4.2.1. Συμπεριλαμβάνοντας residuals για να περάσουμε περισσότερα χαρακτηριστικά στο βαθύτερο επίπεδο της αρχιτεκτονικής από χαμηλότερα επίπεδα, το δίκτυο μπορεί να ανακτήσει χαρακτηριστικά που μπορεί να χαθούν κατά τη διαδικασία δειγματοληψίας προς τα πάνω ή προς τα κάτω. Με αυτόν τον τρόπο αυξάνεται η ικανότητα εκμάθησης του δικτύου.

Ακολούθησε η δοκιμή μας με υπόδομο post-processing, που βασίστηκε στον αλγόριθμο Weighted Line Detector και ήταν αποτελεσματική στην αντιμετώπιση του προβλήματος εξαφάνισης των λεπτών αγγείων και στην εξάλειψη της λανθασμένης κατάτμησης ορισμένων ‘δύσκολων’ αγγειακών περιοχών, κάτι που αντικατοπτρίζεται επίσης από το προηγούμενο χαμηλό σκορ Sensitivity που επιτυγχάνεται από τα νευρωνικά δίκτυά μας. Τα οπικά αποτελέσματα στις εικόνες ελέγχου έδειξαν αξιοσημείωτες βελτιώσεις.

Η τελική μελέτη της διπλωματικής εργασίας επικεντρώθηκε στον τρόπο με τον οποίο θα καταφέρουν να συνδυαστούν τα μοντέλα *Plain_Net_BN* και *Sine_Net* ώστε να αποδώσουν ένα κοινό καλύτερο αποτέλεσμα. Το κίνητρο για την απόφαση αυτή ήταν η διαφορετική φύση και ποικιλομορφία των μεθολογιών που εξυπηρετούν διαφορετικούς σκοπούς, όπως εξηγήσαμε. Η απόφαση που πάρθηκε για τον σκοπό αυτό ήταν τα μοντέλα να εκπαιδεύονται σειριακά σε μικρότερα κομμάτια του training set και να επαληθεύονται σε εξίσου μικρά κομμάτια validation set, ενώ οι προβλέψεις τους θα τροφοδοντούνται σε ένα τελικό ταξινομητή, ο οποίος θα δέχεται τις εξόδους τους και θα πραγματοποιεί μια κοινή πρόβλεψη. Παρόλαυτά, η υπολογιστικές απαιτήσεις της παραπάνω μεθόδου δεν επέτρεψαν την εκπαίδευση για επαρκή δεδομένα και επαρκείς επαναλήψεις ώστε να παρουσιαστεί αξιοσημείωτη βελτίωση των μεμονωμένων αποτελεσμάτων.

6.2 Μελλοντικές Επεκτάσεις και Προτάσεις

Παρόλο που καμία αρχιτεκτονική μεμονωμένα δεν κατάφερε να ξεπεράσει σε απόδοση όλες τις μετρικές στις state-of-the-art μεθόδους, η χρήση διαφορετικών μοντέλων λειτούργησε εποικοδομητικά καθώς μέσω της τριβής με το πρόβλημα προέκυψαν συμπεράσματα που μας βοήθησαν να βελτιώσουμε τα σύστηματα και να σκεφτούμε πολλά υποσχόμενες μελλοντικές προεκτάσεις.

Ένας από τους σημαντικότερους παράγοντες στην αποτελεσματική ανάπτυξη ενός συστήματος Βαθιάς Μηχανικής Μάθησης είναι το μέγεθος του συνόλου δεδομένων που θα χρησιμοποιηθεί. Για το πρόβλημα επιβλεπόμενης μάθησης που προσπαθήσαμε να επιλύσουμε, το μεγαλύτερο μεμονωμένο (από μια πηγή) dataset που μπορέσαμε να εντοπίσουμε με επαρκή βιβλιογραφία αποτελούνταν από 20 εικόνες, αρκετά μικρό για τις απαιτήσεις ενός συστήματος βαθιάς μάθησης, παρόλες τις τεχνικές επαύξησης. Έτσι, ένας πρώτος στόχος θα ήταν η συλλογή ενός μεγαλύτερου και πιθανώς καλύτερα ισορροπημένου συνόλου δεδομένων ή και η χρήση Transfer Learning.

Μια ακόμα κατεύθυνση για μελλοντική ανάπτυξη θα μπορούσε να αποτελέσει η καλύτερη επεξεργασία και διαχείριση της εικόνας αμφιβληστροειδούς. Ένα παράδειγμα είναι η καλύτερη επιλογή των παραμέτρων των μετασχηματισμών που εφαρμόστηκαν, όπως το μετασχηματισμό gamma τους συντελεστές scaling, τις παραμέτρους των μετασχηματισμών CLAHE και MTHT, το μέγεθος των patches κ.α.. Με την εισαγωγή αρκετών εικόνων εκπαίδευσης, εισάγονται και διαφορετικοί τύποι θορύβου με αποτέλεσμα το μοντέλο να μάθει να τις διαφοροποιεί και να ξεπερνάει και αυτές τις προκλήσεις.

Επιπλέον, όπως είναι γνωστό, ο χώρος αναζήτησης των βέλτιστων υπερπαραμέτρων στη βαθιά μηχανική μάθηση είναι πολύ μεγάλος. Συνεπώς η εκτενέστερη μελέτη των βέλτιστων τιμών, ειδικά για το learning rate, θα μπορούσε να βελτιώσει τα αποτελέσματα.

Μια ακόμα σημαντική παράμετρος που θα συνέβαλε καθοριστικά στη βέλτιστοποίηση της απόδοσης και των αποτελεσμάτων είναι η άρση του περιορισμού των πόρων, δηλαδή η χρήση ισχυρών υπολογιστικών συστημάτων θα επέτρεπε την αύξηση της ανάλυσης των εικόνων και ταυτόχρονα την κλιμάκωση των μοντέλων. Με αυτό τον τρόπο, θα μπορέσουμε να εξετάσουμε και τα αποτελέσματα της ευοίωνης μεθόδου Ensemble Learning.

Τέλος, εφικτή θα ήταν η αξιοποίηση και η εφαρμογή του συστήματος που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας και σε άλλα προβλήματα. Παραδείγματος χάρη, μια άλλη χρήση του συστήματος θα μπορούσε να είναι σε επιπλέον προβλήματα που απασχολούν το ερευνητικό πεδίο Medical Image Segmentation για υποβοήθηση διάγνωσης μέσω υπολογιστή, όπως είναι η αναγνώριση και εξαγωγή άλλου είδους παθολογιών από εικόνα, είτε σε εικόνα αμφιβληστροειδούς είτε σε πολλές άλλες εικόνες όπως MRI, PET, κ.α..

Βιβλιογραφία

- [1] Shahzad Akbar et al. “Automated techniques for blood vessels segmentation through fundus retinal images: A review”. In: *Microscopy research and technique* 82.2 (2019), pp. 153–170.
- [2] Michael D Abràmoff, Mona K Garvin, and Milan Sonka. “Retinal imaging and image analysis”. In: *IEEE reviews in biomedical engineering* 3 (2010), pp. 169–208.
- [3] Yun Jiang et al. “Automatic retinal blood vessel segmentation based on fully convolutional neural networks”. In: *Symmetry* 11.9 (2019), p. 1112.
- [4] Ana Salazar-Gonzalez et al. “Segmentation of the blood vessels and optic disk in retinal images”. In: *IEEE journal of biomedical and health informatics* 18.6 (2014), pp. 1874–1886.
- [5] Zhengyuan Liu. “Retinal vessel segmentation based on fully convolutional networks”. In: *arXiv preprint arXiv:1911.09915* (2019).
- [6] *DRIVE: Digital Retinal Images for Vessel Extraction*. URL: <https://drive.grand-challenge.org/>.
- [7] Shuangling Wang et al. “Hierarchical retinal blood vessel segmentation based on feature and ensemble learning”. In: *Neurocomputing* 149 (2015), pp. 708–717.
- [8] Chang Wang et al. “Dense U-net Based on Patch-Based Learning for Retinal Vessel Segmentation”. In: *Entropy* 21 (2019).
- [9] Tao Li et al. “Applications of deep learning in fundus images: A review”. In: *Medical Image Analysis* (2021), p. 101971.
- [10] Renu Jogi. *Basic Ophthalmology*. Jaypee Brothers Medical Publishers, 2008.
- [11] Niall Patton et al. “Retinal image analysis: Concepts, applications and potential”. In: *Progress in Retinal and Eye Research* 25 (2006), pp. 99–127.
- [12] Cong Sun et al. “Retinal vascular caliber: systemic, environmental, and genetic associations.” In: *Survey of ophthalmology* 54 1 (2009), pp. 74–95.
- [13] Nishtha Panwar et al. “Fundus Photography in the 21st Century-A Review of Recent Technological Advances and Their Implications for Worldwide Healthcare”. In: *Telemedicine and e-Health* 22.3 (2016), pp. 198–208.

- [14] M. Niemeijer et al. *DRIVE: digital retinal images for vessel extraction*. 2004.
- [15] Maryam Badar, Muhammad Haris, and Anam Fatima. “Application of deep learning for retinal image analysis: A review”. In: *Computer Science Review* 35 (2020), p. 100203.
- [16] Arthur L. Samuell. “Some Studies in Machine Learning Using the Game of Checkers”. In: *IBM Journal of Research and Development* 3 (1959), pp. 210–229.
- [17] Thomas Mitchell. *Machine learning*. McGraw-Hill Pub. Co. (ISE Editions), 1997.
- [18] Warren S. McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259). URL: <https://doi.org/10.1007/BF02478259>.
- [19] F. Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain”. In: *Psychological Review* (1958), pp. 65–386.
- [20] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1969.
- [21] Simon Haykin. “Neural Networks and Learning Machines”. In: 2010.
- [22] Facundo Bre, Juan M. Gimenez, and Víctor D. Fachinotti. “Prediction of wind pressure coefficients on building surfaces using artificial neural networks”. In: *Energy and Buildings* 158 (2018), pp. 1429–1441. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2017.11.045>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778817325501>.
- [23] Ronald J. Williams David E. Rumelhart Geoffrey E. Hinton. “Learning representations by back-propagating errors”. In: *Nature* 323 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). URL: <https://doi.org/10.1038/323533a0>.
- [24] Agnes Lydia and Sagayaraj Francis. “Adagrad - An Optimizer for Stochastic Gradient Descent”. In: (May 2019).
- [25] Matthew D. Zeiler. *ADADELTA: An Adaptive Learning Rate Method*. 2012. arXiv: [1212.5701 \[cs.LG\]](https://arxiv.org/abs/1212.5701).
- [26] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: [1412.6980 \[cs.LG\]](https://arxiv.org/abs/1412.6980).
- [27] T.N. Wiesel D.H. Hubel. “Receptive fields and functional architecture of monkey striate cortex”. In: *J Physiol* 195 (Mar. 1968), pp. 215–243. DOI: [10.1113/jphysiol.1968.sp008455](https://doi.org/10.1113/jphysiol.1968.sp008455).
- [28] Dr.Kunihiko Fukushima. “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position”. In: *Biological Cybernetics* 36 (1980), pp. 193–202. DOI: [doi:10.1007/BF00344251](https://doi.org/10.1007/BF00344251).

- [29] Geoffrey E. Hinton Alex Krizhevsky Ilya Sutskever. “ImageNet classification with deep convolutional neural networks”. In: *Communication of the ACM* 60 (2012), pp. 84–90. DOI: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [30] J. J. Hopfield. “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”. In: *Proceedings of the National Academy of Sciences of the United States of America* 79.8 (1982), pp. 2554–2558. ISSN: 00278424. URL: <http://www.jstor.org/stable/12175>.
- [31] Paweł Liskowski and Krzysztof Krawiec. “Segmenting Retinal Blood Vessels With Deep Neural Networks”. In: *IEEE Transactions on Medical Imaging* 35 (2016), pp. 2369–2380.
- [32] Mohammad Hesam Hesamian et al. “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges”. In: *Journal of Digital Imaging* 32.4 (May 2019), pp. 582–596. DOI: [10.1007/s10278-019-00227-x](https://doi.org/10.1007/s10278-019-00227-x). URL: <https://doi.org/10.1007/s10278-019-00227-x>.
- [33] Y. Lecun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791). URL: <https://doi.org/10.1109/5.726791>.
- [34] Li Deng. “A tutorial survey of architectures, algorithms, and applications for deep learning – ERRATUM”. In: *APSIPA Transactions on Signal and Information Processing* 3.1 (2014). DOI: [10.1017/atsip.2014.4](https://doi.org/10.1017/atsip.2014.4). URL: <https://doi.org/10.1017/atsip.2014.4>.
- [35] Agnieszka Mikołajczyk and Michał Grochowski. “Data augmentation for improving deep learning in image classification problem”. In: *2018 International Interdisciplinary PhD Workshop (IIPhDW)* (2018), pp. 117–122.
- [36] Liang Zou et al. “Spectral–Spatial Exploration for Hyperspectral Image Classification via the Fusion of Fully Convolutional Networks”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 659–674. DOI: [10.1109/jstars.2020.2968179](https://doi.org/10.1109/jstars.2020.2968179). URL: <https://doi.org/10.1109/jstars.2020.2968179>.
- [37] Y. Lecun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
- [39] Martina Melinak, Pavle Prentasic, and Sven Lonarić. “Retinal Vessel Segmentation using Deep Neural Networks”. In: *VISAPP*. 2015.

- [40] H. Fu et al. “Retinal vessel segmentation via deep learning network and fully-connected conditional random fields”. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (2016), pp. 698–701.
- [41] Rahul Purohit. *Fully Convolutional Network (Semantic Segmentation)*. 2020. URL: <https://www.mygreatlearning.com/blog/fcn-fully-convolutional-network-segmentation/>.
- [42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* abs/1505.04597 (2015). arXiv: [1505.04597](https://arxiv.org/abs/1505.04597). URL: <http://arxiv.org/abs/1505.04597>.
- [43] İbrahim Atli and Osman Serdar Gedik. “Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation”. In: *Engineering Science and Technology, an International Journal* 24.2 (2021), pp. 271–283. ISSN: 2215-0986. DOI: <https://doi.org/10.1016/j.jestch.2020.07.008>. URL: <https://www.sciencedirect.com/science/article/pii/S221509862030330X>.
- [44] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. DOI: [10.48550/ARXIV.1502.03167](https://arxiv.org/abs/1502.03167). URL: <https://arxiv.org/abs/1502.03167>.
- [45] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: <http://jmlr.org/papers/v15/srivastava14a.html>.
- [46] David H. Wolpert. “Stacked generalization”. In: *Neural Networks* 5.2 (Jan. 1992), pp. 241–259. DOI: [10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1). URL: [https://doi.org/10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1).
- [47] Francois Chollet et al. *Keras*. 2015. URL: <https://github.com/fchollet/keras>.
- [48] *what is classification threshold - deepchecks* 2022. URL: <https://deepchecks.com/glossary/classification-threshold/>.
- [49] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [50] Uyen T.V. Nguyen et al. “An effective retinal blood vessel segmentation method using multi-scale line detection”. In: *Pattern Recognition* 46.3 (Mar. 2013), pp. 703–715. DOI: [10.1016/j.patcog.2012.08.009](https://doi.org/10.1016/j.patcog.2012.08.009). URL: <https://doi.org/10.1016/j.patcog.2012.08.009>.

- [51] Chao Zhou, Xiaogang Zhang, and Hua Chen. “A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden Markov model”. In: *Computer Methods and Programs in Biomedicine* 187 (Apr. 2020), p. 105231. DOI: [10.1016/j.cmpb.2019.105231](https://doi.org/10.1016/j.cmpb.2019.105231). URL: <https://doi.org/10.1016/j.cmpb.2019.105231>.
- [52] Qiaoliang Li et al. “A cross-modality learning approach for vessel segmentation in retinal images”. In: *IEEE Transactions on Medical Imaging* 35.1 (2016), pp. 109–118. DOI: [10.1109/tmi.2015.2457891](https://doi.org/10.1109/tmi.2015.2457891).
- [53] Juan Mo and Lei Zhang. “Multi-level deep supervised networks for retinal vessel segmentation”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.12 (2017), pp. 2181–2193. DOI: [10.1007/s11548-017-1619-0](https://doi.org/10.1007/s11548-017-1619-0).
- [54] Zhexin Jiang et al. “Retinal blood vessel segmentation using fully convolutional network with transfer learning”. In: *Computerized Medical Imaging and Graphics* 68 (2018), pp. 1–15. DOI: [10.1016/j.compmedimag.2018.04.005](https://doi.org/10.1016/j.compmedimag.2018.04.005).
- [55] Shouting Feng et al. “CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features”. In: *Neurocomputing* 392 (2020), pp. 268–276. DOI: [10.1016/j.neucom.2018.10.098](https://doi.org/10.1016/j.neucom.2018.10.098).
- [56] Kai Hu et al. “Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function”. In: *Neurocomputing* 309 (2018), pp. 179–191. DOI: [10.1016/j.neucom.2018.05.011](https://doi.org/10.1016/j.neucom.2018.05.011).
- [57] M Elena Martinez-Perez et al. “Segmentation of blood vessels from red-free and fluorescein retinal images”. In: *Medical image analysis* 11.1 (2007), pp. 47–61.
- [58] Lauri Laaksonen. “Spectral retinal image processing and analysis for ophthalmology”. In: 2016.
- [59] Ann Nowe, Peter Vrancx, and Yann-Michaël De Hauwere. “Game Theory and Multi-agent Reinforcement Learning”. In: Jan. 2012, p. 30. ISBN: 978-3-642-27645-3. DOI: [10.1007/978-3-642-27645-3_14](https://doi.org/10.1007/978-3-642-27645-3_14).
- [60] Sven Arnhold. “Data restructuring as formal preprocessing for machine learning with neural networks”. In: 2015.

6.2.0.0.0.1