



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών  
και Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής  
και Υπολογιστών

## Κατηγοριοποίηση Ιστοπαθολογικών Εικόνων Καρκίνου του Μαστού με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΕΛΕΝΗ Δ. ΣΤΡΑΪΤΟΥΡΗ**

**Επιβλέπων :** Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπων :** Γεώργιος Σιόλας

Εργαστηριακό Διδακτικό Προσωπικό Ε.Μ.Π.

Αθήνα, Νοέμβριος 2020





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών  
και Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής  
και Υπολογιστών

## Κατηγοριοποίηση Ιστοπαθολογικών Εικόνων Καρκίνου του Μαστού με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΛΕΝΗ Δ. ΣΤΡΑΪΤΟΥΡΗ

**Επιβλέπων :** Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπων :** Γεώργιος Σιόλας  
Εργαστηριακό Διδακτικό Προσωπικό Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 24η Νοεμβρίου 2020.

.....  
Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

.....  
Γεώργιος Στάμου  
Αν. Καθηγητής Ε.Μ.Π.

.....  
Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2020

.....  
**Ελένη Δ. Στραϊτούρη**

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

Copyright © Ελένη Δ. Στραϊτούρη, 2020.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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.

## Περίληψη

Η νευραλγική σημασία της έγκαιρης κι έγκυρης διάγνωσης για την επιτυχή αντιμετώπιση πολύ σοβαρών ασθενειών, όπως ο καρκίνος του μαστού, έχει στρέψει ένα μεγάλο μέρος του ερευνητικού ενδιαφέροντος στην ανάπτυξη ευφύων συστημάτων υποβοήθησης ιατρικής διάγνωσης. Ειδικά για τη διάγνωση καρκίνου του μαστού μέσω της διαδικασίας της βιοψίας, όπου ύποπτα δείγματα ιστών εξετάζονται σε μικροσκόπιο, το πρόβλημα της ανάπτυξης ενός τέτοιου συστήματος υποβοήθησης, ως πρόβλημα μηχανικής μάθησης ανάγεται στο ευρύτερο πρόβλημα της κατηγοριοποίησης εικόνων. Σε αυτό το πλαίσιο, αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η μελέτη του προβλήματος αυτού, δηλαδή της κατηγοριοποίησης ιστοπαθολογικών εικόνων μικροσκοπίας καρκίνου του μαστού, με την εφαρμογή τεχνικών μηχανικής μάθησης και συγκεκριμένα με την εφαρμογή μεταφοράς μάθησης, σε πολυεπίπεδες συνελκτικές αρχιτεκτονικές διαφορετικών τεχνοτροπιών. Εξετάζονται διαφορετικές πτυχές του προβλήματος και συγκεκριμένα συγκρίνεται η επίδοση όλων των υπό μελέτη αρχιτεκτονικών για το πρόβλημα της κατηγοριοποίησης τόσο σε δύο όσο και περισσότερες κλάσεις, ενώ ακόμη αξιολογείται η επίδοσή τους σε δεδομένα ίδιας αλλά και διαφορετικής μεγέθυνσης, μελετώντας παράλληλα και την εφαρμογή τεχνικών επαύξησης δεδομένων για τα δεδομένα κοινής μεγέθυνσης. Από τα εξαιρετικά πειραματικά αποτελέσματα για το δυαδικό πρόβλημα, προέκυψε πράγματι ότι η μεταφορά μάθησης μπορεί να εφαρμοστεί επιτυχώς για τη δημιουργία ενός συστήματος υποβοήθησης ιατρικής διάγνωσης με υψηλή αξιοπιστία.

## Λέξεις κλειδιά

Μεταφορά μάθησης, Βαθιά συνελκτικά νευρωνικά δίκτυα, Επιβλεπόμενη Μάθηση, VGG16, DenseNet, Δίκτυα Squeeze-and-Excitation, Επαύξηση δεδομένων.



## **Abstract**

The key role of an early and accurate diagnosis for very serious diseases, such as breast cancer, has turned a great part of research interest towards the development of intelligent computer-aided medical diagnosis systems. Especially for breast cancer diagnosis, with biopsy process, in which suspicious tissue samples are examined under a microscope, the problem of the development of such system, as a machine learning problem, is translated to the general problem of image classification. In this context, the objective of the present diploma thesis is the study of this problem, namely breast cancer histopathological image classification with machine learning techniques and specifically with the application of transfer learning in deep convolutional differently structured architectures. Different aspects of the problem are examined and more specifically, the performance of all the architectures under study is compared, as far as not only the binary but also the multi-classification problem are concerned, whereas their performance is also evaluated over data at the same or different magnification, studying in parallel the application of data augmentation techniques for data at the same magnification. The exceptional experimental results for the binary problem have led to the conclusion that transfer learning can indeed be applied successfully for the development of a highly trustworthy computer-aided diagnosis system.

## **Key words**

Transfer learning, Deep convolutional neural networks, Supervised learning, VGG16, DenseNet, Squeeze-and-Excitation networks, Data augmentation





## Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Ανδρέα-Γεώργιο Σταφυλοπάτη για την ευκαιρία που μου έδωσε να ασχοληθώ με το παρόν επιστημονικό πεδίο, όπως επίσης και τους κ. Στέφανο Κόλλια, και κ. Γεώργιο Στάμου για την συμμετοχή τους στην εξεταστική επιτροπή. Ιδιαίτερες ευχαριστίες αξίζουν επίσης στον κ. Γεώργιο Σιόλα για τη συνεχή υποστήριξη και άρτια συνεργασία καθόλη τη διάρκεια εκπόνησης της παρούσας εργασίας. Τέλος, ένα μεγάλο ευχαριστώ από καρδιάς στους φίλους μου, στην αγαπημένη μου οικογένεια και ειδικά στη λατρεμένη μου αδερφή, που μου στάθηκαν και ήταν πάντα στο πλευρό μου όλα αυτά τα χρόνια.

Ελένη Δ. Στραϊτούρη,

Αθήνα, 24η Νοεμβρίου 2020



# Περιεχόμενα

Περίληψη . . . . .	5
Abstract . . . . .	7
Ευχαριστίες . . . . .	9
Περιεχόμενα . . . . .	11
Κατάλογος πινάκων . . . . .	13
Κατάλογος σχημάτων . . . . .	15
<b>1. Εισαγωγή . . . . .</b>	<b>17</b>
1.1 Εισαγωγικές Έννοιες . . . . .	17
1.2 Αντικείμενο της εργασίας . . . . .	18
1.3 Δομή της εργασίας . . . . .	18
<b>2. Θεωρητικό υπόβαθρο . . . . .</b>	<b>21</b>
2.1 Μηχανική Μάθηση . . . . .	21
2.1.1 Ορισμός . . . . .	21
2.1.2 Είδη . . . . .	21
2.2 Νευρωνικά Δίκτυα . . . . .	22
2.2.1 Δομικά Στοιχεία . . . . .	24
2.2.2 Εκπαίδευση . . . . .	27
2.3 Μετρικές Αξιολόγησης . . . . .	29
2.4 Μεταφορά μάθησης . . . . .	31
2.4.1 Στρατηγικές . . . . .	32
2.4.2 Δείγματα Αρχιτεκτονικών . . . . .	33
<b>3. Ανασκόπηση Ερευνητικού Έργου . . . . .</b>	<b>37</b>
3.1 Δυαδική Κατηγοριοποίηση Κοινής Μεγέθυνσης . . . . .	37
3.2 Δυαδική Κατηγοριοποίηση Ανεξαρτήτου Μεγέθυνσης . . . . .	39
3.3 Κατηγοριοποίηση Πολλών Κλάσεων Κοινής Μεγέθυνσης . . . . .	40
3.4 Κατηγοριοποίηση Πολλών Κλάσεων Ανεξαρτήτου Μεγέθυνσης . . . . .	42

<b>4. Σύνολο Δεδομένων</b>	45
4.1 Περιγραφή	45
4.2 Αποθήκευση	47
4.3 Προεπεξεργασία	48
4.3.1 Βασική Προεπεξεργασία	48
4.3.2 Επαύξηση Δεδομένων	48
<b>5. Πειραματικές Διατάξεις</b>	51
5.1 VGG16	52
5.1.1 Αρχιτεκτονική	52
5.1.2 Εκπαίδευση	53
5.2 DenseNet121	53
5.2.1 Αρχιτεκτονική	53
5.2.2 Εκπαίδευση	54
5.3 IDSNet	54
5.3.1 Αρχιτεκτονική	54
5.3.2 Εκπαίδευση	55
<b>6. Πειραματικά αποτελέσματα</b>	57
6.1 Δυαδική Κατηγοριοποίηση Κοινής Μεγέθυνσης	57
6.1.1 Μεγέθυνση 40X	57
6.1.2 Μεγέθυνση 100X	58
6.1.3 Μεγέθυνση 200X	58
6.1.4 Μεγέθυνση 400X	59
6.2 Κατηγοριοποίηση Πολλών Κλάσεων Κοινής Μεγέθυνσης	59
6.2.1 Μεγέθυνση 40X	59
6.2.2 Μεγέθυνση 100X	60
6.2.3 Μεγέθυνση 200X	60
6.2.4 Μεγέθυνση 400X	61
6.3 Δυαδική Κατηγοριοποίηση Ανεξαρτήτου Μεγέθυνσης	61
6.4 Κατηγοριοποίηση Πολλών Κλάσεων Ανεξαρτήτου Μεγέθυνσης	62
<b>7. Συμπεράσματα - Μελλοντικές επεκτάσεις</b>	63
7.1 Συμπεράσματα	63
7.2 Μελλοντικές επεκτάσεις	64
<b>Βιβλιογραφία</b>	65

## Κατάλογος πινάκων

4.1	Κατανομή καλοήθων υπότυπων . . . . .	46
4.2	Κατανομή κακοήθων υπότυπων . . . . .	46
6.1	Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης $40\times$ . . . . .	57
6.2	Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης $100\times$ . . . . .	58
6.3	Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης $200\times$ . . . . .	58
6.4	Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης $400\times$ . . . . .	59
6.5	Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης $40\times$ .	59
6.6	Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης $100\times$	60
6.7	Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης $200\times$	61
6.8	Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης $400\times$	61
6.9	Αποτελέσματα δυαδικής κατηγοριοποίησης ανεξαρτήτου μεγέθυνσης . . . . .	61
6.10	Αποτελέσματα πολυ-κατηγοριοποίησης ανεξαρτήτου μεγέθυνσης . . . . .	62



## Κατάλογος σχημάτων

2.1	Βιολογικός και Τεχνητός Νευρώνας . . . . .	23
2.2	Πολυστρωματικό Perceptron για το μη διαχωρίσιμο πρόβλημα XOR . . . . .	23
2.3	Συνέλιξη . . . . .	25
2.4	Η αρχιτεκτονική AlexNet . . . . .	33
2.5	Το στρώμα Inception . . . . .	34
2.6	Οι residual δομικές μονάδες ταυτότητας-identity (αριστερά) και προβολής-projection (δεξιά) . . . . .	35
2.7	Δείγμα αρχιτεκτονικής DenseNet . . . . .	35
2.8	Δομική Μονάδα Squeeze-and-Excitation . . . . .	36
4.1	Καλοήθη δείγματα σε μεγέθυνση $40\times$ . . . . .	46
4.2	Κακοήθη δείγματα σε μεγέθυνση $40\times$ . . . . .	47
5.1	Η αρχιτεκτονική VGG16 . . . . .	52
5.2	Η αρχιτεκτονική IDSNet . . . . .	55





## Κεφάλαιο 1

### Εισαγωγή

#### 1.1 Εισαγωγικές Έννοιες

Μία από τις σοβαρότερες ασθένειες που καλείται να αντιμετωπίσει η παγκόσμια ιατρική κοινότητα, η οποία ευθύνεται για εκατομμύρια θανάτους παγκοσμίως σε ετήσια βάση είναι ο καρκίνος. Η ασθένεια αυτή μπορεί να προσβάλλει οποιοδήποτε τύπο ιστού, άρα και οποιοδήποτε όργανο, και χαρακτηρίζεται από την ανεξέλεγκτη αναπαραγωγή δυσλειτουργικών κυττάρων και την εισβολή τους σε γειτονικές περιοχές του οργανισμού, καθώς και την πιθανή εξάπλωσή τους σε διαφορετικά όργανα. Σύμφωνα με στατιστικές μελέτες για το 2018 του παγκόσμιου οργανισμού υγείας ([Worldwide Health Organisation - WHO](#)) η ασθένεια αυτή αποτέλεσε τη δεύτερη συχνότερη αιτία θανάτου, προκαλώντας βάσει εκτιμήσεων την απώλεια 9.6 εκατομμυρίων ανθρώπινων ζωών. Η ίδια μελέτη αναδεικνύει ότι μεταξύ των συχνότερων τύπων καρκίνου βρίσκεται ο καρκίνος του μαστού με 2.09 εκατομμύρια διαγνώσεις, ενώ συγκαταλέγεται και στους πιο θανατηφόρους τύπους με 627 χιλιάδες θανάτους.

Ο καρκίνος του μαστού, ειδικά για το γυναικείο πληθυσμό, παρουσιάζει συγκριτικά με άλλους τύπους πολύ υψηλή θνησιμότητα, με μεγάλο μερίδιο ευθύνης να φέρει η μη έγκαιρη διάγνωση, ειδικά για χώρες μέσου και χαμηλού εισοδήματος, όπου η πρόσβαση σε διαγνωστικές υπηρεσίες υγείας είναι περιορισμένη. Δεδομένης της νευραλγικής σημασίας της έγκαιρης κι έγκυρης διάγνωσης, έχουν αναπτυχθεί πολλές διαφορετικές μέθοδοι εντοπισμού και διάγνωσης, οι οποίες συχνά μπορεί να εφαρμόζονται και συνδυαστικά. Μεταξύ αυτών οι συχνότερα εφαρμοζόμενες, όσον αφορά τον εντοπισμό και απεικόνιση ύποπτων αλλοιώσεων, είναι η ψηφιακή μαστογραφία, η μαγνητική τομογραφία και ο υπέρηχος, ενώ η πιο αξιόπιστη μέθοδος διάγνωσης είναι η βιοψία, η οποία περιλαμβάνει την αφαίρεση δειγμάτων των προβληματικών ιστών και την εξέτασή τους στο μικροσκόπιο.

Η εξαγωγή κυτταρικών δειγμάτων, στα πλαίσια βιοψίας μπορεί να γίνει με διάφορους τρόπους, όπως για παράδειγμα με λεπτή βελόνη, με κόπτουσα (μεγαλύτερης διαμέτρου) βελόνα, με ανοιχτή ή χειρουργική αφαίρεση, ενώ σε κάθε περίπτωση απαιτείται η ύπαρξη εξειδικευμένου ιατρικού προσωπικού για την εξαγωγή και προετοιμασία των δειγμάτων για μικροσκοπία. Η εξέταση των δειγμάτων στο μικροσκόπιο και η εξαγωγή της τελικής γνώματευσης πραγματοποιείται από πεπειραμένους κι εξειδικευμένους ιστοπαθολόγους και αποτελεί μία εξαιρετικά χρονοβόρα διαδικασία. Δεδομένων των πολύ υψηλών αυτών απαιτήσεων, άνθρωποι παράγοντες όπως η κούραση ή η έλλειψη επαρκούς προσοχής επηρεάζουν την ιστοπαθολογική ανάλυση και μπορεί εύκολα να οδηγήσουν σε εσφαλμένη διάγνωση. Στο σημείο αυτό, αναδύεται η ανάγκη ανάπτυξης συστημάτων υποβοήθησης ιατρικής διάγνωσης, προκειμένου να αποτελέσουν ένα ισχυρό συμβουλευτικό εργαλείο για την ιατρική κοινότητα και να συνεισφέρουν στην ελαχιστοποίηση της πιθανότητας διαγνωστικού σφάλματος. Βασικό πόλο έλξης της ερευνητικής κοινότητας για την ανάπτυξη τέτοιων συστημάτων, λαμβάνοντας υπόψη την

εντυπωσιακή άνθιση που έχει γνωρίσει τα τελευταία χρόνια, τις εκπληκτικές δυνατότητες και το απίστευτα ευρύ πεδίο εφαρμογών της, έχει αποτελέσει φυσικά η μηχανική μάθηση.

Βασικές προϋποθέσεις για τη μελέτη κι ανάπτυξη συστημάτων υποβοήθησης ιατρικής διάγνωσης καρκίνου του μαστού από ιστοπαθολογικές εικόνες, με τεχνικές μηχανικής μάθησης είναι αφενός η δημιουργία και διάθεση εκτενών συνόλων δεδομένων με τέτοιες εικόνες, αφετέρου η ορθή μοντελοποίηση του προβλήματος. Όσον αφορά τη συλλογή και δημιουργία ενός συνόλου δεδομένων ιστοπαθολογικών εικόνων καρκίνου του μαστού, πρέπει να τονιστεί ότι πρόκειται για διαδικασία εξαιρετικά χρονοβόρα κι απαιτητική. Ωστόσο, αν και περιορισμένα σε αριθμό, τα τελευταία χρόνια έχουν δημιουργηθεί τέτοια σύνολα, με τα πιο γνωστά από αυτά να διατίθενται είτε στα πλαίσια διαγωνισμών όπως ο Grand Challenge on Breast Cancer Histology images (BACH) [1] και ο διαγωνισμός CAMELYON [2], είτε από πρωτοβουλίες ερευνητικών κέντρων και πανεπιστημίων όπως το σύνολο δεδομένων BreakHis [3], το οποίο και χρησιμοποιήθηκε στην παρούσα εργασία. Αναφορικά με τη μοντελοποίηση του προβλήματος στο πεδίο της μηχανικής μάθησης, η προσέγγιση που συνήθως ακολουθείται συνίσταται στην αναγωγή του στο ευρύτερο πρόβλημα της κατηγοριοποίησης εικόνων, με τις ιστολογικές εικόνες μικροσκοπίας να αποτελούν τα δεδομένα του προβλήματος, τα οποία διακρίνονται ως κακοήθη σε περίπτωση που περιέχουν καρκινικά κύτταρα, ή καλοήθη διαφορετικά. Επέκταση της παραπάνω προσέγγισης αποτελεί η διάκριση των εικόνων με βάση το υποείδος της καλοήθους ή κακοήθους αλλοίωσης που παρατηρείται, δηλαδή η κατηγοριοποίηση σε πολλές κλάσεις.

## 1.2 Αντικείμενο της εργασίας

Αντικείμενο της παρούσας διπλωματικής αποτελεί η μελέτη του προβλήματος της κατηγοριοποίησης ιστοπαθολογικών εικόνων καρκίνου του μαστού, μέσω τεχνικών μηχανικής μάθησης. Ειδικότερα, εξετάζεται η εφαρμογή της τεχνικής μεταφοράς μάθησης, η οποία συνίσταται στην αξιοποίηση ήδη αποκτηθείσας γνώσης ενός συστήματος σχετικά με την αναγνώριση οπτικών αναπαραστάσεων γενικής φύσεως, προκειμένου να μπορεί αυτό να ειδικευθεί στην διάκριση εικόνων εξειδικευμένης φύσεως κι εν προκειμένω ιστοπαθολογικών εικόνων καρκίνου του μαστού. Η μελέτη της εφαρμογής μεταφοράς μάθησης στο υπό εξέταση πρόβλημα πραγματοποιείται μέσω της εκπαίδευσης κι αξιολόγησης τριών διαφορετικών αρχιτεκτονικών βαθιάς μάθησης στο σύνολο δεδομένων BreakHis [3]. Το συγκεκριμένο σύνολο δεδομένων δίνει τη δυνατότητα μελέτης τόσο του προβλήματος δυαδικής κατηγοριοποίησης, όπου η διάκριση γίνεται μεταξύ καρκινικών και μη δειγμάτων, όσο και του προβλήματος κατηγοριοποίησης στα υποείδη της κυτταρικής αλλοίωσης κάθε δείγματος. Παράλληλα, η παροχή εικόνων μικροσκοπίας σε διαφορετικές μεγεθύνσεις προσθέτει μία ακόμη διάσταση στο πρόβλημα, υπό την έννοια της εξέτασης αποκλειστικά δεδομένων κοινής μεγέθυνσης ή δεδομένων όλων των μεγεθύνσεων. Οι μελετηθείσες αρχιτεκτονικές εφαρμόστηκαν κι αξιολογήθηκαν για κάθε έκφανση του προβλήματος με βάση τα παραπάνω, ενώ ειδικά για τα προβλήματα με δεδομένα κοινής μεγέθυνσης, εξετάστηκε και η εφαρμογή τεχνικών επαύξησης του αρχικού συνόλου εκπαίδευσης.

## 1.3 Δομή της εργασίας

Η εργασία αποτελείται συνολικά από 7 κεφάλαια. Σε αυτά αναλύονται τόσο το απαραίτητο θεωρητικό υπόβαθρο όσο και οι λεπτομέρειες υλοποίησης των μοντέλων και διεξαγωγής των πειραματικών

δοκιμών. Συγκεκριμένα στο κεφάλαιο 2 αναλύονται οι βασικές θεωρητικές έννοιες για την κατανόηση των αρχιτεκτονικών και της πειραματικής διαδικασίας ενώ στο κεφάλαιο 3 παρουσιάζονται οι κυριότερες μελέτες που διατίθενται στη βιβλιογραφία ειδικά για το υπό εξέταση σύνολο δεδομένων. Σχετικά με την πειραματική διαδικασία, στο κεφάλαιο 4 περιγράφεται αναλυτικά το σύνολο δεδομένων όπως και η επεξεργασία που υπέστη, στο κεφάλαιο 5 εξετάζονται οι υπό μελέτη αρχιτεκτονικές και στο κεφάλαιο 6 παρατίθενται τα πειραματικά αποτελέσματα. Στο κεφάλαιο 7 που είναι και το τελευταίο εξάγονται τα συνολικά συμπεράσματα που προέκυψαν από την πειραματική διαδικασία, ενώ τέλος προτείνονται μελλοντικές κατευθύνσεις μελέτης.



## Κεφάλαιο 2

# Θεωρητικό υπόβαθρο

## 2.1 Μηχανική Μάθηση

### 2.1.1 Ορισμός

Ως Μηχανική Μάθηση ορίζεται η διαδικασία αυτόματης βελτίωσης ορισμένων λειτουργιών ενός υπολογιστικού συστήματος μέσω της τροφοδότησης αυτού με ένα σύνολο εμπειριών [4]. Οι εμπειρίες αυτές μπορούν να αποτελούνται από οποιοδήποτε είδος ερεθισμάτων και πληροφοριών αρκεί αυτές να δύνανται να μετασχηματιστούν με τέτοιο τρόπο, ώστε να αποτελούν την είσοδο ενός υπολογιστικού αλγορίθμου. Το επιστημονικό πεδίο της Μηχανικής Μάθησης εντάσσεται στον ευρύτερο τομέα της τεχνητής νοημοσύνης και σκοπό έχει τη μελέτη κι ανάπτυξη ευφυών συστημάτων, ικανών να αναγνωρίζουν πρότυπα, να εξάγουν συμπεράσματα, να προβλέπουν δράσεις, προκειμένου να αποτελέσουν ένα ισχυρό εργαλείο με ένα γιγαντιαίο πλήθος εφαρμογών σε τομείς όπως η βιομηχανία, η υγεία κ.ά.

### 2.1.2 Είδη

Ανάλογα με τις βασικές λειτουργίες που επιτελεί η μηχανική μάθηση διακρίνονται σε διάφορες κατηγορίες οι κυριότερες εκ των οποίων είναι οι εξής τρεις: *επιβλεπόμενη μάθηση*, *μη επιβλεπόμενη μάθηση* κι *ενισχυτική μάθηση*.

**Επιβλεπόμενη Μάθηση** Στην κατηγορία αυτή εντάσσονται οι αλγόριθμοι με στόχο τη μάθηση μίας συνάρτησης μέσω ενός συνόλου δεδομένων που περιέχουν για κάθε είσοδο την αντίστοιχη επιθυμητή έξοδο. Το σύνολο αυτό αποτελεί το σύνολο εκπαίδευσης, ενώ η επιτυχία του αλγορίθμου κρίνεται από την ικανότητά του να προβλέπει τη σωστή έξοδο για άγνωστα δεδομένα εισόδου που ανήκουν σε ένα σύνολο ελέγχου. Το πρόβλημα της κατηγοριοποίησης ενός συνόλου στιγμιότυπων ενός προβλήματος, καθένα από τα οποία ανήκει σε μία ή περισσότερες διακριτές κατηγορίες, το πλήθος των οποίων είναι πεπερασμένο, μπορεί εύκολα να προσεγγιστεί από τέτοιους αλγορίθμους, με τον περιορισμό ότι για κάθε στιγμιότυπο (δεδομένο εισόδου) πρέπει να είναι γνωστή η κατηγορία/-ίες του, ή αλλιώς η ετικέτα/-ες του (επιθυμητή έξοδος).

**Μη Επιβλεπόμενη Μάθηση** Στο είδος αυτό συγκαταλέγονται αλγόριθμοι, οι οποίοι έχουν ως είσοδο δεδομένα για τα οποία η επιθυμητή έξοδος δεν είναι γνωστή. Σκοπό τους αποτελεί η εύρεση και η ανάδειξη της εγγενούς δομής του συνόλου των δεδομένων, όπως ο εντοπισμός ομάδων ή κοινοτήτων που μπορεί να σχηματίζουν αυτά. Η ομοιότητα ή η ανομοιότητα μεταξύ δύο δεδομένων εισόδου

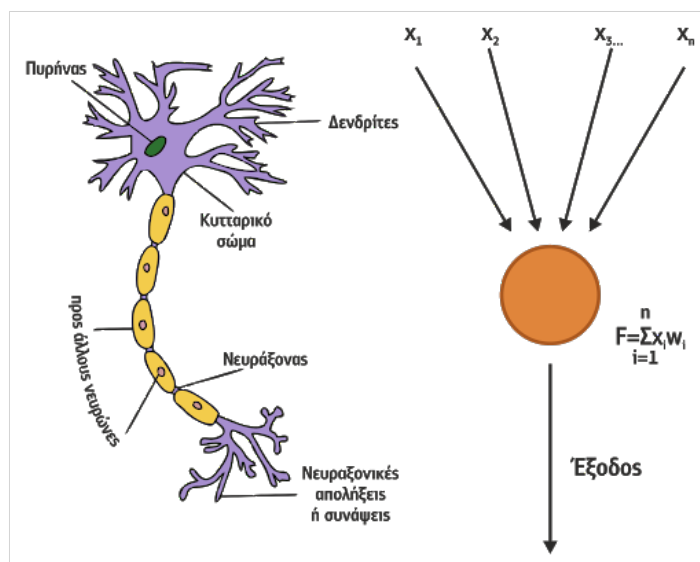
με βάση ορισμένα κριτήρια είναι εκείνη που καθορίζει εάν αυτά αποτελούν μέλη της ίδιας ή διαφορετικής ομάδας. Μεταξύ επιβλεπόμενης και μη επιβλεπόμενης μάθησης εντάσσεται και η κατηγορία των αλγορίθμων ημι-επιβλεπόμενης μάθησης, για την οποία είναι γνωστή η επιθυμητή έξοδος μόνο για ένα μέρος των δεδομένων εισόδου. Τέτοιοι αλγόριθμοι δημιουργήθηκαν για την εκμετάλλευση των πλεονεκτημάτων των δύο μεθόδων, προκειμένου να βελτιώσουν αφενός την ικανότητα μάθησης και αφετέρου να αντιμετωπίσουν προβλήματα, στα οποία είναι δυσχερής η παροχή των επιθυμητών δεδομένων εξόδου για όλα τα στιγμιότυπα εισόδου.

**Ενισχυτική Μάθηση** Οι αλγόριθμοι ενισχυτικής μάθησης υφίστανται στα πλαίσια ενός περιβάλλοντος (environment) στο οποίο δρουν ως πράκτορες (agents), λαμβάνοντας ως εισόδους καταστάσεις του περιβάλλοντος αυτού, από τις οποίες λαμβάνουν μία επιβράβευση (reward). Ανάλογα με την κατάσταση, οι αλγόριθμοι επιλέγουν δράσεις (actions), μέσω των οποίων μεταβαίνουν σε νέα κατάσταση, με στόχο τη μεγιστοποίηση της επιβράβευσης. Χαρακτηριστικό τους αποτελεί το γεγονός ότι δεν υπάρχει προκαθορισμένο σύνολο εκπαίδευσης, αφού οι καταστάσεις από τις οποίες θα μεταβεί ο αλγόριθμος καθορίζονται από την αλληλουχία των δράσεων που θα επιλέξει. Ένα επιτυχημένο σύστημα ενισχυτικής μάθησης καλείται να προσαρμόζεται ανάλογα με τις καταστάσεις, επιλέγοντας τη βέλτιστη ακολουθία δράσεων έτσι ώστε να επιτυγχάνεται η μέγιστη συνολική επιβράβευση. Η προσαρμοστικότητα αυτή καθιστά αυτούς τους αλγορίθμους ιδανικούς σε ένα δυναμικά μεταβαλλόμενα πολυπαραγοντικά προβλήματα αυξημένης πολυπλοκότητας όπως η αυτόνομη οδήγηση, οι ρομποτικές εφαρμογές, παίγνια υπολογιστή εναντίον ανθρώπου, κ.ά. .

## 2.2 Νευρωνικά Δίκτυα

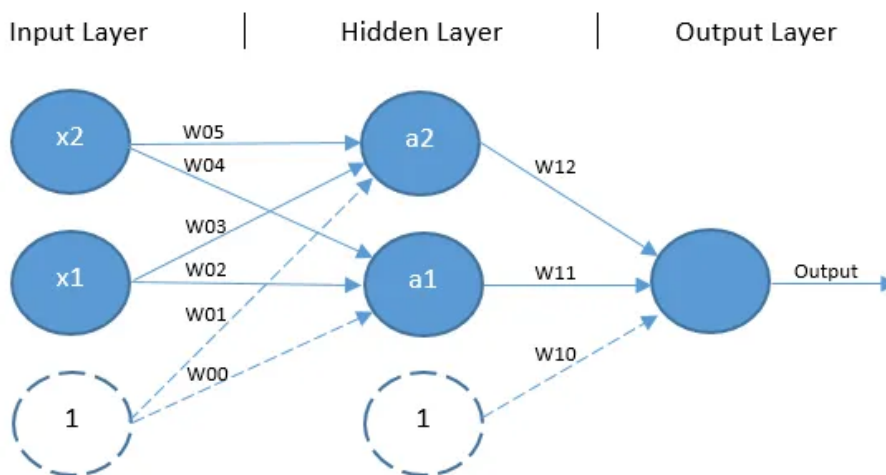
Καρδιά της υλοποίησης της συντριπτικής πλειοψηφίας των συστημάτων μηχανικής μάθησης αποτελούν τα τεχνητά νευρωνικά δίκτυα, πηγή έμπνευσης των οποίων αποτέλεσε ο τρόπος λειτουργίας των βιολογικών νευρώνων. Ένας βιολογικός νευρώνας αφού δεχθεί κάποια σήματα εισόδου στους δενδρίτες (βλ. σχήμα 2.1), τα αθροίζει στον πυρήνα του κι εφόσον αυτά υπερβαίνουν κάποιο κατώφλι διαβιβάζονται μέσω του άξονα και των συνάψεων σε γειτονικούς νευρώνες. Ομοίως, οι τεχνητοί νευρώνες δέχονται ένα διάνυσμα εισόδου, αθροίζουν τα στοιχεία του, αφού αυτά πολλαπλασιαστούν με ορισμένα βάρη, και μεταβιβάζουν το άθροισμα αυτό στους γειτονικούς νευρώνες εφόσον υπερβαίνει μία τιμή κατώφλιου. Η λειτουργία της απόφασης μεταβίβασης σήματος σε γειτονικούς νευρώνες αναπαρίσταται ως μία μαθηματική συνάρτηση, που χαρακτηρίζεται ως συνάρτηση ενεργοποίησης. Οι τεχνητοί νευρώνες συνιστούν τη βασική δομική μονάδα των τεχνητών νευρωνικών δικτύων, τα οποία οργανώνονται σε ένα ή περισσότερα επίπεδα, τα οποία μπορούν να περιλαμβάνουν έναν ή περισσότερους νευρώνες. Την απλούστερη μορφή ενός τεχνητού νευρωνικού δικτύου ενός επιπέδου κι ενός νευρώνα, εισήγαγε ο Rosenblatt το 1958 με το όνομα Perceptron [5], ενώ την έννοια του τεχνητού νευρώνα ως μαθηματικό υπολογιστικό μοντέλο εισήγαγαν πρώτοι οι McCulloch και Pitts το 1943 [6].

Το απλό μοντέλο Perceptron αποδείχθηκε εξαιρετικά αποτελεσματικό, βρίσκοντας λύση σε πεπερασμένο αριθμό επαναλήψεων, σε προβλήματα ταξινόμησης γραμμικά διαχωρίσιμων στιγμιότυπων, υπό την έννοια ότι, δεδομένης της αναπαράστασης των δεδομένων εισόδου ως διανύσματα πεπερασμένων διαστάσεων στο χώρο, υπάρχει ένα υπερεπίπεδο το οποίο διαχωρίζει όλα τα στιγμιότυπα της μίας κλάσης από την άλλη. Ωστόσο, το μοντέλο ενός επιπέδου για προβλήματα μη γραμμικά διαχωρίσιμων στιγμιότυπων, στα οποία απαιτείται η εύρεση μια σύνθετης καμπύλης για το



Σχήμα 2.1: Βιολογικός και Τεχνητός Νευρώνας

διαχωρισμό τους, είναι ανεπαρκές. Για τα προβλήματα αυτά απεδείχθη ότι μοντέλα Perceptron δύο επιπέδων/στρωμάτων -ενός κρυφού (hidden) επιπέδου και της εξόδου βλ. σχήμα 2.2- (Multi Layer Perceptron) μπορούν να προσεγγίσουν την καμπύλη διαχωρισμού με το ελάχιστο επιθυμητό σφάλμα [7], αναδεικνύοντας τις δυνατότητες και την αξία των πολυεπίπεδων αρχιτεκτονικών, στις οποίες είναι στραμμένο ένα μεγάλο τμήμα του ερευνητικού ενδιαφέροντος μέχρι και σήμερα.



Σχήμα 2.2: Πολυστρωματικό Perceptron για το μη διαχωρίσιμο πρόβλημα XOR

Αξίζει να σημειωθεί ότι αν και τα νευρωνικά δίκτυα υφίστανται πάνω από μισό αιώνα ως ευφυή θεωρητικά υπολογιστικά μοντέλα, πρακτική εφαρμογή και τη μεγαλύτερη άνθιση ως επιστημονικός κλάδος γνώρισαν με την αύξηση της υπολογιστικής δύναμης των ηλεκτρονικών υπολογιστών και την ανάπτυξη ισχυρότερων επεξεργαστικών μονάδων, όπως οι μονάδες επεξεργασίας γραφικών (Graphics Processing Units - GPUs). Η δυνατότητα παροχής πληθώρας υπολογιστικών πόρων, κατέστησε δυνατή την υλοποίηση των νευρωνικών δικτύων σε προγραμματιστικό περιβάλλον, τη διε-

ρεύνηση της πληθώρας των δυνατοτήτων τους και την αξιοποίησή τους για τη μελέτη μίας τεράστιας γκάμας προβλημάτων.

Με την ανάπτυξη της έρευνας στο πεδίο των νευρωνικών δικτύων αναπτύχθηκαν διαφορετικοί τύποι νευρώνων και κατ' επέκταση επιπέδων, όπως και συναρτήσεων ενεργοποίησης. Οι κυριότεροι τύποι αυτών, οι οποίοι χρησιμοποιούνται κατά κόρον στα προβλήματα κατηγοριοποίησης εικόνων και όχι μόνο, παρουσιάζονται στην επόμενη ενότητα, ενώ στη συνέχεια αναλύεται η διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου.

### 2.2.1 Δομικά Στοιχεία

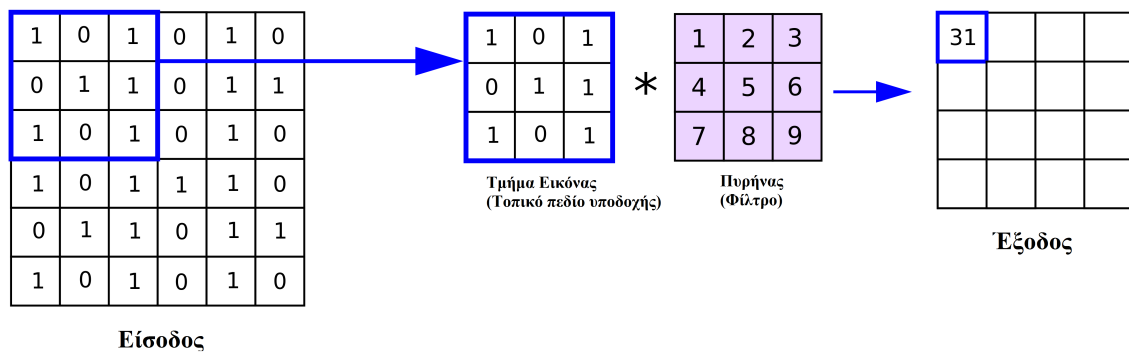
#### Επίπεδα

**Πλήρως Συνδεδεμένο** Πρόκειται για τον πιο απλό τύπο επιπέδου, ο οποίος μπορεί να αποτελείται από έναν ή περισσότερους νευρώνες, καθένας από τους οποίους λαμβάνει ως είσοδο όλες τις εξόδους του προηγούμενου επιπέδου, εξού και η ονομασία του. Στις εφαρμογές κατηγοριοποίησης εικόνας χρησιμοποιείται συνήθως ως τελευταίο επίπεδο, με πλήθος νευρώνων ίσο με το πλήθος των κλάσεων, δίνοντας την πρόβλεψη για την κατηγορία της εικόνας.

**Συνελικτικό** Τα συνελικτικά επίπεδα δημιούργησαν τη μεγάλη οικογένεια των συνελικτικών δικτύων (Convolutional Neural Networks - CNNs) τα οποία αποτελούν σταθμό για την εξέλιξη των νευρωνικών δικτύων στον τομέα της κατηγοριοποίησης εικόνας αλλά και σε πολλούς άλλους τομείς, όπως η ανάλυση βίντεο, η αναγνώριση αντικειμένων, η επεξεργασία φυσικής γλώσσας κ.ά. Η τεράστια αναγνώριση κι επιτυχία τους συνίσταται στην ικανότητα εξαγωγής ωφέλιμης πληροφορίας από δεδομένα υψηλής διαστατικότητας με τοπικά χαρακτηριστικά, όπως οι εικόνες.

Ως είσοδος σε ένα συνελικτικό επίπεδο θεωρείται ένα τρισδιάστατο διάνυσμα, με ορισμένο πλάτος, μήκος και βάθος, πρακτικά τον τρόπο αναπαράστασης μίας εικόνας. Στο διάνυσμα αυτό εφαρμόζεται ένα ή περισσότερα φίλτρα που ονομάζονται και πυρήνες (kernels), δηλαδή ένα επίσης τρισδιάστατο διάνυσμα βαρών ίδιου βάθους, με μήκος και πλάτος αρκετά μικρότερα από αυτά του διανύσματος εισόδου. Από τη διαδικασία αυτή προκύπτει ένα νέο διάνυσμα με βάθος ίσο με το πλήθος των φίλτρων που εφαρμόστηκαν, το οποίο αποτελεί το χάρτη ενεργοποίησης (activation map). Κάθε στοιχείο αυτού του χάρτη έχει προκύψει από το εσωτερικό γινόμενο κάθε φίλτρου με κάθε υποπεριοχή (τοπικό πεδίο υποδοχής - local receptive field) του διανύσματος εισόδου που εφαρμόστηκε. Σημειώνεται ότι τα φίλτρα διατρέχουν ολόκληρο το διάνυσμα εισόδου ως μετακινούμενα παράθυρα με τα βήματα της μετατόπισης (stride) ως προς τους άξονες μήκους και το πλάτους να αποτελούν υπερπαραμέτρο του συστήματος. Σε περίπτωση που το επιθυμητό stride οδηγεί το φίλτρο εκτός των ορίων του διανύσματος εισόδου, συνήθως πρακτική αποτελεί το παραγέμισμα του τελευταίου περιμετρικά με μηδενικά, προκειμένου πάντα να υπάρχει πλήρης επικάλυψη των φίλτρων με αυτό. Ακόμη, στο χάρτη ενεργοποίησης μπορεί να εφαρμοστεί και οποιαδήποτε συνάρτηση ενεργοποίησης. Σχηματικά, απεικονίζεται στο 2.3 ο υπολογισμός ενός στοιχείου του χάρτη ενεργοποίησης σε ένα ορισμένο βάθος. Από την παραπάνω διαδικασία και μέσω της εκπαίδευσης οι τιμές των βαρών κάθε πυρήνα με τέτοιο τρόπο ώστε ο χάρτης ενεργοποίησης να περιλαμβάνει την αναπαράσταση ορισμένων χαρακτηριστικών. Σε πολυεπίπεδες συνελικτικές αρχιτεκτονικές εκπαιδευμένες σε σύνολα εικόνων, επί παραδείγματι, στα κατώτερα επίπεδα εξάγονται κι αναγνωρίζονται χαρακτηριστικά χαμηλού επιπέδου, όπως ακμές, καμπύλες και γενικώς αδρότερα χαρακτηριστικά, ενώ σε υψηλότερα επίπεδα ανα-





Σχήμα 2.3: Συνέλιξη

γνωρίζονται πολυπλοκότερα οπτικά μοτίβα με περισσότερες λεπτομέρειες, όπως πρόσωπα ή τμήματα αυτών, αντικείμενα κ.τ.λ. .

**Επίπεδα δειγματοληψίας - Pooling** Εξαιρετικά σύνηθες είναι τα συνελκτικά επίπεδα να ακολουθούνται από επίπεδα δειγματοληψίας, τα οποία οδηγούν σε μείωση των διαστάσεων. Συγκεκριμένα τα επίπεδα αυτά λαμβάνουν ως είσοδο ένα σύνολο από χάρτες ενεργοποίησης, με καθέναν από αυτούς να διατρέχεται από ένα κυλιόμενο παράθυρο μικρότερου μεγέθους, το οποίο να δίνει ως έξοδο είτε τη μέγιστη, είτε τη μέση τιμή των στοιχείων κάθε υποτιμήματος στο οποίο εφαρμόζεται. Παράλληλα, αυτού του τύπου επιπέδου αποτελεί το επίπεδο καθολικής δειγματοληψίας (Global Pooling), στο οποίο εξάγεται για κάθε χάρτη ενεργοποίησης η μέγιστη ή η μέση τιμή όλων των στοιχείων. Η μείωση διαστατικότητας που επιτυγχάνεται με τη χρήση επιπέδων Pooling συνεπάγεται τη μείωση και τον αναγκαίων παραμέτρων, μειώνοντας αφενός το υπολογιστικό κόστος, αφετέρου τον κίνδυνο το μοντέλο να μάθει να αναγνωρίζει καλά μόνο τα δεδομένα με τα οποία εκπαιδεύτηκε και όχι άγνωστα δεδομένα, γνωστό και ως πρόβλημα υπερπροσαρμογής (overfitting).

**Επίπεδα απόρριψης - Dropout** Άλλο ένα είδος επιπέδου με σκοπό τη μείωση της υπερπροσαρμογής αποτελούν και τα επίπεδα απόρριψης. Τοποθετούνται μετά από οποιοδήποτε επίπεδο κι έχουν ως αποτέλεσμα τον τυχαίο μηδενισμό ενός ποσοστού των τιμών εισόδου που λαμβάνουν, το οποίο δίνεται ως παράμετρος. Όποια τιμή δε μηδενιστεί κανονικοποιείται, έτσι ώστε το άθροισμα των τιμών που "επιβιώνουν" να μη μεταβάλλεται, προκειμένου να αποφευχθούν υπερβολικά μεγάλες τιμές των βαρών κατά την εκπαίδευση.

**Επίπεδα κανονικοποίησης δέσμης - Batch Normalization** Ένα αρκετά συχνά χρησιμοποιούμενο είδος επιπέδου, το οποίο εισήχθη σχετικά πρόσφατα από τους Iofee και Szegedy [8], αποτελεί το επίπεδο κανονικοποίησης δέσμης, του οποίου η είσοδος μετασχηματίζεται (κανονικοποιείται) με τέτοιο τρόπο έτσι ώστε να έχει μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Στόχος της εφαρμογής του επιπέδου αυτού αποτελεί η αντιμετώπιση του προβλήματος της αλλαγής της κατανομής των εισόδων κάθε ενδιάμεσου επιπέδου λόγω της μεταβολής των παραμέτρων κατά την εκπαίδευση. Η συνεχής μεταβολή της κατανομής, μεταβάλλει ακολούθως την κατανομή στόχο στην οποία προσπαθούν να προσαρμοστούν οι παράμετροι, με αποτέλεσμα μεγάλη καθυστέρηση στη σύγκλιση άρα και

συνολική επιβράδυνση της διαδικασίας μάθησης. Με την εφαρμογή της κανονικοποίησης, το πρόβλημα αυτό αντιμετωπίζεται κι επιτυγχάνεται ταχύτερη σύγκλιση, ενώ επίσης η αποφυγή υπερβολικά μεγάλων τιμών εισόδου, λόγω κανονικοποίησης μπορεί να συνεισφέρει θετικά και για το πρόβλημα της υπεπροσαρμογής, όπως συμβαίνει και με τα επίπεδα απόρριψης.

## Συναρτήσεις ενεργοποίησης

Αν και υπάρχουν πολλές και διαφορετικές συναρτήσεις ενεργοποίησης με διαφορετικά χαρακτηριστικά η καθεμία, διακρίνονται σε δύο οικογένειες συναρτήσεων : τις γραμμικές και τις μη γραμμικές. Στην πρώτη ανήκουν μεταξύ άλλων οι συναρτήσεις κατωφλίου, οι συναρτήσεις προσήμου, οι συναρτήσεις ράμπας, ενώ στη δεύτερη συναρτήσεις όπως οι σιγμοειδείς. Μεγαλύτερη πρακτική εφαρμογή παρουσιάζουν οι μη-γραμμικές συναρτήσεις καθότι η μη γραμμικότητα επιτρέπει την εκμετάλλευση πολυεπίδων αρχιτεκτονικών, οι οποίες ενδείκνυνται για την κατηγοριοποίηση εικόνων. Στη συνέχεια παρουσιάζονται οι κυριότερες μη γραμμικές συναρτήσεις ενεργοποίησης, η πλειοψηφία των οποίων χρησιμοποιήθηκε και στα πειράματα:

- **Σιγμοειδής.** Η συνάρτηση αυτή παρουσιάζει το πλεονέκτημα ότι μπορεί να λαμβάνει συνεχείς τιμές στο διάστημα  $[0, 1]$ , επιθυμητό χαρακτηριστικό σε προβλήματα όπου απαιτείται πρόβλεψη πιθανοτήτων. Ακόμη, ωφέλιμο χαρακτηριστικό για την εκπαίδευση αποτελεί η παραγωγισμότητά της, ενώ τρωτό σημείο καθίσταται η σχεδόν μηδενική μεταβολή των τιμών εξόδου στις περιοχές "γωνάτου" για διαφορετικές τιμές εισόδου, η οποία οδηγεί σε πολύ μικρές τιμές παραγώγου, οδηγώντας σε τερματισμό της εκπαίδευσης (πρόβλημα εξαφανιζόμενης παραγώγου - vanishing gradient).

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

- **Υπερβολική Εφαπτομένη.** Όμοιες ιδιότητες με τις παραπάνω εμφανίζει και η συνάρτηση υπερβολικής εφαπτομένης, η οποία λαμβάνει τιμές στο διάστημα  $[-1, 1]$ . Η συμμετρία της ως προς την αρχή των αξόνων σε συνδυασμό με την πιο απότομη παράγωγη την κάνουν να πλεονεκτεί έναντι της απλής σιγμοειδούς, αναφορικά με την εκπαίδευση.

$$\text{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- **Rectified Linear Unit - ReLU.** Μία εξαιρετικά συχνά χρησιμοποιούμενη συνάρτηση ενεργοποίησης, η οποία έχει μικρότερη πολυπλοκότητα υπολογισμού από τις προαναφερθείσες, μη άνω φραγμένη έξοδο και δεν είναι παραγωγίσιμη στο 0.

$$\text{ReLU}(x) = \max(0, x)$$

- **Softmax.** Αποτελώντας γενίκευση της παραπάνω συνάρτησης, η συνάρτηση αυτή παρουσιάζει παρόμοια χαρακτηριστικά και χρησιμοποιείται ευρέως σε προβλήματα κατηγοριοποίησης πολλών κλάσεων στο τελευταίο επίπεδο για τον υπολογισμό της κατανομής πιθανότητας στις κλάσεις εξόδου.

$$\text{Softmax}(\vec{z})_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

## 2.2.2 Εκπαίδευση

Ο μηχανισμός εκπαίδευσης ενός νευρωνικού δικτύου συνίσταται στην προσαρμογή των βαρών των νευρώνων με βάση έναν κανόνα εκπαίδευσης, ή αλλιώς έναν αλγόριθμο εκπαίδευσης, ο οποίος διαφέρει ανάλογα με το είδος μάθησης. Αναφορικά με την επιβλεπόμενη μάθηση, το πρόβλημα της εκπαίδευσης μοντελοποιείται ως πρόβλημα μεταβολής των βαρών, έτσι ώστε να ελαχιστοποιείται μια συνάρτηση σφάλματος μεταξύ πραγματικής κι επιθυμητής εξόδου. Αντίστοιχη λογική ακολουθείται και στην ενισχυτική μάθηση, όπου τα βάρη προσαρμόζονται έτσι ώστε να μεγιστοποιείται η τιμή της επιβράβευσης. Στη μη επιβλεπόμενη μάθηση, η προσαρμογή των βαρών γίνεται με την εφαρμογή κάποιας συνάρτησης επάνω στα δεδομένα εισόδου χωρίς να υπάρχει κάποια αντικειμενική συνάρτηση, αφού η επιθυμητή έξοδος δεν είναι γνωστή και στόχος είναι να αναδειχθούν ομάδες μεταξύ των δεδομένων. [9]

Ειδικότερα, για το πρόβλημα της αναγνώρισης και κατηγοριοποίησης προτύπων, όπως εικόνες, επιστρατεύονται βαθιά, δηλαδή πολυεπίπεδα συνελκτικά δίκτυα, τα οποία εκπαιδεύονται με επιβλεπόμενη μάθηση. Κύρια συστατικά της διαδικασίας μάθησης αποτελούν η συνάρτηση σφάλματος ή απώλειας (loss function) και ο ίδιος ο αλγόριθμος μάθησης με βάση τον οποίο θα μεταβάλλονται τα βάρη προκειμένου να ελαχιστοποιηθεί η τιμή της συνάρτησης σφάλματος εφαρμοσμένη πάνω στην προβλεπόμενη και την επιθυμητή έξοδο. Στη συνέχεια παρουσιάζονται συνοπτικά ο κυριότερος αλγόριθμος μάθησης και η συνηθέστερη συνάρτηση σφάλματος, που επιλέχθηκαν και για τα πειράματα.

### Αλγόριθμος Οπίσθιας Διάδοσης - Backpropagation

Ο κύριος αλγόριθμος που εφαρμόζεται για την εκπαίδευση είναι ο αλγόριθμος οπίσθιας διάδοσης (backpropagation) που αν και διατυπώθηκε πρώτη φορά το 1974 από τον Werbos [10], δεν έλαβε αναγνώριση στην επιστημονική κοινότητα πριν από 1985, όταν επανεισήχθει ταυτόχρονα από τους LeCun [11], Parker [12], και το 1986 από τους Hinton et al. [13]. Βασική ιδέα του αλγορίθμου είναι τα βάρη σε κάθε βήμα εκπαίδευσης να μεταβάλλονται ελαφρά, ανάλογα με το πόσο συνεισφέρουν στη συνάρτηση σφάλματος, προς την κατεύθυνση που ελαχιστοποιεί την τιμή σφάλματος. Ως βήμα εκπαίδευσης θεωρείται η τροφοδότηση του μοντέλου με ένα δεδομένο ή μία μικρή ομάδα αυτών (batch) και η προσαρμογή των βαρών, δηλαδή η εκπαίδευση, με βάση τον κανόνα μάθησης.

Θεωρώντας τη συνάρτηση σφάλματος ως συνάρτηση των βαρών, είναι εμφανές πως για την ελαχιστοποίησή της, οι μεταβολές των βαρών πρέπει να γίνουν αντίθετα στην κατεύθυνση της παραγώγου της συνάρτησης σφάλματος ως προς αυτά. Για το λόγο αυτό ο γενικός κανόνας μάθησης με βάση τον αλγόριθμο backpropagation, ορίζει ότι η μεταβολή κάθε βάρους είναι ανάλογη της παραγώγου της συνάρτησης σφάλματος ως προς αυτό, πολλαπλασιασμένης επί μίας αριθμητικής σταθεράς, η οποία αποτελεί το ρυθμό μάθησης. Η μέθοδος αυτή, κατά την οποία τα βάρη μεταβάλλονται στην κατεύθυνση της παραγώγου, οδηγώντας την τιμή του σφάλματος σε όλο και χαμηλότερα επίπεδα είναι γνωστή και ως κατάβαση κλίσης (Gradient Descent). Εάν  $L$  είναι η συνάρτηση σφάλματος,  $w_{ij}^l(k)$  το βάρος μεταξύ των νευρώνων  $i$  και  $j$  του επιπέδου  $l$  για το βήμα εκπαίδευσης  $k$  και  $a$  ο ρυθμός μάθησης, τότε ο κανόνας μάθησης ορίζεται ως:

$$w_{ij}^l(k+1) = w_{ij}^l(k) - a \frac{\partial L}{\partial w_{ij}^l(k)}$$

Ο υπολογισμός της παραγώγου για τα βάρη του επιπέδου εξόδου είναι πιο απλός καθώς η τιμή της συνάρτησης σφάλματος άρα και της παραγώγου της υπολογίζεται με βάση την τιμή εξόδου και

την επιθυμητή έξοδο. Για τα βάρη των ενδιάμεσων επιπέδων για τον υπολογισμό απαιτείται η εφαρμογή του κανόνα της αλυσίδας και ως εκ τούτου, ο αναδρομικός υπολογισμός των παραγώγων των συναρτήσεων ενεργοποίησης (εξού και η σημασία της παραγωγισιμότητάς τους) από το επίπεδο εξόδου μέχρι και το ενδιάμεσο επίπεδο για το οποίο γίνεται η ενημέρωση των βαρών. Δεδομένου ότι η συνάρτηση σφάλματος υπολογίζεται απευθείας μόνο για το επίπεδο εξόδου, αναμενόμενο είναι να απαιτείται ένας υπολογισμός αναδρομικής φύσης, έτσι ώστε να "διαδοθεί" η παράγωγος του σφάλματος από την έξοδο στα ενδιάμεσα επίπεδα, δηλαδή προς τα πίσω.

Συνολικά η εκτέλεση του αλγορίθμου περιλαμβάνει δύο φάσεις υπολογισμών σε κάθε βήμα. Η πρώτη περιλαμβάνει τον υπολογισμό των εξόδων των ενδιάμεσων επιπέδων από το πρώτο μέχρι το επίπεδο εξόδου (προς τα εμπρός), ενώ η δεύτερη τον υπολογισμό των παραγόντων μεταβολής των βαρών με βάση τον κανόνα της αλυσίδας από το τελευταίο επίπεδο στο πρώτο (προς τα πίσω). Μετά από αυτές ανανεώνονται τα βάρη και ο αλγόριθμος συνεχίζει έως ότου συγκλίνει, δηλαδή η τιμή σφάλματος γίνει μικρότερη από μια μικρή θετική ποσότητα  $\epsilon$ , είτε μέχρι να διατρέξει όλα τα δεδομένα (εποχή) για προκαθορισμένο πλήθος φορών. Σημαντικό ρόλο σε ό,τι αφορά τη σύγκλιση παίζει ο ρυθμός μάθησης του οποίου, υπερβολικά μικρές τιμές οδηγούν σε πολύ αργή σύγκλιση ή και εγκλωβισμό των τιμών των βαρών σε τοπικά ακρότατα της συνάρτησης σφάλματος, ενώ υπερβολικά μεγάλες σε ταλάντωση γύρω από βέλτιστες τιμές [14].

Το πρόβλημα της προσαρμογής του αλγορίθμου μάθησης για ταχύτερη σύγκλιση και η ανάγκη αποφυγής εγκλωβισμού σε τοπικά ελάχιστα, οδήγησε στην ανάπτυξη παραλλαγών του βασικού αλγορίθμου κατάβασης κλίσης, με στόχο την καλύτερη προσαρμογή των παραμέτρων ανάλογα και με τη συνεισφορά τους στην εκπαίδευση. Οι πιο διαδεδομένοι τέτοιοι αλγόριθμοι βελτιστοποίησης που χρησιμοποιούνται είναι ο αλγόριθμος στοχαστικής κατάβασης κλίσης (Stochastic Gradient Descent-SGD), ο αλγόριθμος διάδοσης ρίζας μέσου τετραγώνου (Root Mean Square Propagation-RMSProp) και ο αλγόριθμος προσαρμοζόμενης εκτίμησης ροπής (Adaptive Moment Estimation-Adam), με τους δύο τελευταίους να χρησιμοποιούνται και στην πειραματική διαδικασία.

Εφαπτήριο για την ανάπτυξη του αλγορίθμου RMSProp αποτέλεσε το γεγονός ότι σε ένα πολυεπίπεδο δίκτυο, οι τιμές των παραγώγων μπορεί να διαφέρουν κατά πολύ, με αποτέλεσμα να απαιτούνται και διαφορετικοί ρυθμοί μάθησης σε κάθε περίπτωση. Σύμφωνα με τον RMSProp, ο οποίος εισήχθη από τον [Geoff Hinton](#), για την αντιμετώπιση του παραπάνω προβλήματος, στον παράγοντα που πολλαπλασιάζεται με την παράγωγο, εισάγεται πέραν του ρυθμού μάθησης και η ρίζα της εκθετικά μειούμενης μέσης τιμής του τετραγώνου των παραγώγων, με τον κανόνα μάθησης να δίνεται από τους τύπους:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g_t^2$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \cdot g_t$$

, όπου  $\theta_t$  η μεταβλητή των παραμέτρων για το βήμα  $t$ ,  $g_t$  η παράγωγος για το βήμα  $t$ ,  $\eta$  ο ρυθμός μάθησης,  $E[g^2]_t$  η μέση τιμή του τετραγώνου των παραγώγων για τη χρονική στιγμή  $t$  και  $\gamma$  ένας αριθμητικός παράγοντας με προτεινόμενη τιμή 0.9.

Σε παρόμοια λογική κινείται και ο αλγόριθμος Adam [15] με τη διαφορά ότι για να προσαρμοστούν οι παράγοντες μάθησης για κάθε παράμετρο, αξιοποιούνται οι εκτιμήσεις πρώτης και δεύτερης ροπής των παραγώγων, εξού και η ονομασία του. Με βάση τον αλγόριθμο οι εκτιμήσεις πρώτης και δεύτερης ροπής  $m_t$ ,  $v_t$  αντίστοιχα των παραγώγων  $g_t$  για το βήμα εκπαίδευσης  $t$ , αρχικοποιούνται

στο 0 και υπολογίζονται βάσει των τύπων:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$
$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$$

, όπου  $\beta_1$  και  $\beta_2$  αριθμητικές τιμές με προτεινόμενες τιμές 0.9 και 0.999 αντίστοιχα. Οι εκτιμήσεις αυτές, πριν χρησιμοποιηθούν στον κανόνα μάθησης, υφίστανται διόρθωση προκειμένου να μην τείνουν να μηδενίζονται ειδικά για τα πρώτα βήματα εκπαίδευσης, με τις διορθωμένες τιμές  $\hat{m}_t$  και  $\hat{v}_t$  να προκύπτουν ως εξής:

$$\hat{m}_t = \frac{m_t}{(1 - \beta_1^t)}$$
$$\hat{v}_t = \frac{v_t}{(1 - \beta_2^t)}$$

, όπου  $\beta_1^t$  και  $\beta_2^t$  οι παραπάνω αριθμητικές σταθερές υψωμένες στη δύναμη  $t$ . Με βάση τα παραπάνω ο κανόνας μάθησης για τον αλγόριθμο Adam δίνεται ως:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t$$

, όπου  $\theta_t$  η μεταβλητή των παραμέτρων για το βήμα  $t$  και  $\eta$  ο ρυθμός μάθησης και συνιστώμενη τιμή για το  $\epsilon$   $10^{-8}$ .

Ακόμη, ανεξάρτητα από τον αλγόριθμο που χρησιμοποιείται για την εκπαίδευση, στη διάθεση του προγραμματιστή είναι και αλγόριθμοι μεταβολής του ρυθμού μάθησης (schedulers) κατά τη διάρκεια της εκπαίδευσης. Τέλος, ρόλο παίζει και το πλήθος και η διάταξη των δεδομένων με τα οποία τροφοδοτείται το δίκτυο πριν από κάθε βήμα εκπαίδευσης, με την επικρατούσα πρακτική να περιλαμβάνει τη οργάνωση των δεδομένων σε τυχαία υποσύνολα ίδιου μεγέθους (mini-batches), με το βήμα εκπαίδευσης να εκτελείται μετά από την τροφοδότηση του δικτύου με ένα από αυτά.

## Διασταυρούμενη Εντροπία - Cross Entropy

Η συνηθέστερη συνάρτηση απώλειας που χρησιμοποιείται είναι η συνάρτηση απώλειας διασταυρούμενης εντροπίας (cross entropy loss) που στην γενική περίπτωση δίνεται από τον τύπο:

$$Loss = - \sum_{i=1}^n t_i \log(p_i)$$

, όπου το  $n$  είναι το πλήθος των κλάσεων, το  $t_i$  η τιμή αληθείας του  $i$ -οστού στοιχείου του διάνυσματος της ετικέτας και  $p_i$  η προβλεπόμενη πιθανότητα η έξοδος να ανήκει στην  $i$ -οστή κλάση. Σημειώνεται ότι το διάνυσμα της ετικέτας για ένα πρόβλημα  $n$  κατηγοριών αναπαρίσταται ως ένα διάνυσμα με  $n$  στοιχεία με μηδενικές τιμές παντού εκτός από τη θέση που αντιστοιχεί στην κλάση του διάνυσματος εισόδου, όπου και τίθεται ως τιμή η μονάδα. Από τον ορισμό, εύκολα προκύπτει ότι όσο μεγαλύτερη είναι η πιθανότητα να προβλεφθεί η σωστή κατηγορία τόσο μικρότερο είναι το σφάλμα.

## 2.3 Μετρικές Αξιολόγησης

Σε προβλήματα κατηγοριοποίησης, για την αξιολόγηση της απόδοσης ενός συστήματος μηχανικής μάθησης, πέραν από την τιμή της συνάρτησης απώλειας, υπάρχει κι ένα πλήθος από μετρικές

που εκφράζουν την ακρίβεια των προβλέψεων και χρησιμεύουν τόσο στην τελική αξιολόγηση του συστήματος όσο και για την επίβλεψη της πορείας του κατά την εκπαίδευσή του. Στη συνέχεια παρουσιάζονται οι βασικότερες μετρικές, που χρησιμοποιούνται γενικά σε προβλήματα κατηγοριοποίησης, όπως και μια ειδική μετρική που αφορά ειδικά τις εφαρμογές υποβοήθησης ιατρικής διάγνωσης. Οι μετρικές αυτές είναι κι εκείνες που χρησιμοποιήθηκαν στα πειράματα

**Ακρίβεια - Accuracy** Ορίζεται ως το ποσοστό των στιγμιότυπων εισόδου που ταξινομήθηκαν σωστά και δίνεται από τον τύπο:

$$Accuracy = \frac{N_{recognised}}{N}$$

, όπου  $N_{recognised}$  το πλήθος των ορθά αναγνωρισθέντων στιγμιότυπων και  $N$  το σύνολο των στιγμιότυπων που εξετάστηκαν.

**Precision** Η μετρική αυτή εκφράζει το ποσοστό των ορθών προβλέψεων στο σύνολο των στιγμιότυπων για τα οποία προβλέφθηκε μία ορισμένη κλάση. Μια χρήσιμη αναπαράσταση για τον υπολογισμό της παρούσας (και της επόμενης) μετρικής αποτελεί ο πίνακας σύγχυσης (Confusion Matrix), στο οποίο υπάρχει μία στήλη και μία γραμμή για κάθε κλάση, με τις στήλες να αναπαριστούν το πλήθος των στιγμιότυπων που προβλέφθηκε ότι ανήκουν στην αντίστοιχη κλάση και τις γραμμές τα πλήθη των στιγμιότυπων που πράγματι ανήκουν στην αντίστοιχη κλάση. Για το πρόβλημα της δυαδικής κατηγοριοποίησης συνήθεις είναι οι χαρακτηρισμοί true positive/negative ( $t_p/t_n$ ) και false positive/negative ( $f_p/f_n$ ) για τις τιμές του confusion matrix, με τους όρους true/false να αφορούν την ορθότητα ή μη της πρόβλεψης και τους positive/negative να αναπαριστούν την κλάση πρόβλεψης. Σύμφωνα με τα παραπάνω, η τιμή της μετρικής για το δυαδικό πρόβλημα δίνεται ως:

$$Precision = \frac{t_p}{t_p + f_p}$$

Για το πρόβλημα πολλών κλάσεων, η μετρική υπολογίζεται για κάθε κλάση, θεωρώντας αυτή ως "θετική" και τις υπόλοιπες ως "αρνητικές", οπότε ο παρανομαστής προκύπτει ως το άθροισμα των τιμών της αντίστοιχης στήλης του πίνακα σύγχυσης.

**Ανάκληση - Recall** Σε παρόμοια λογική με την παραπάνω μετρική, η παρούσα εκφράζει το για ορισμένη κλάση ποσοστό των ορθών προβλέψεων, στο σύνολο των στιγμιότυπων που πραγματικά ανήκουν σε αυτήν. Για το δυαδικό πρόβλημα και με βάση τη σημειολογία για τις τιμές του πίνακα συμβόλων που έχει ήδη δοθεί ορίζεται ως :

$$Recall = \frac{t_p}{t_p + f_n}$$

Αντίστοιχα για το πρόβλημα πολλών κλάσεων, η μετρική υπολογίζεται για κάθε κλάση με τον παρανομαστή να προκύπτει αυτή τη φορά ως το άθροισμα των στοιχείων της γραμμής του πίνακα σύγχυσης που αντιστοιχεί στην εκάστοτε κλάση.

**F1 Score** Αποτελεί τον αρμονικό μέσο των δύο προηγούμενων μετρικών και ορίζεται ως :

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Για το πρόβλημα πολλών κλάσεων οι μετρικές precision και recall προκύπτουν είτε ως ο μέσος όρος των μετρικών κάθε κλάσης (macro-average), είτε ως κλάσμα με αριθμητή και παρανομαστή το άθροισμα των αριθμητών και των παρονομαστών αντίστοιχα των μετρικών κάθε κλάσης (micro-average).

**Ακρίβεια σε επίπεδο ασθενούς** Μία εξειδικευμένη μετρική ακρίβειας για την αξιολόγηση ευφών συστημάτων υποβοήθησης ιατρικής διάγνωσης είναι και η ακρίβεια σε επίπεδο ασθενούς (Patient Level Accuracy), που ορίζεται ως ο μέσος όρος των ποσοστών αναγνώρισης των ιστοπαθολογικών εικόνων κάθε ασθενή. Εάν θεωρήσουμε ως  $N$  το πλήθος των σωστά αναγνωρισθέντων εικόνων για ορισμένο ασθενή στον οποίον αντιστοιχούν συνολικά  $N_p$  εικόνες, τότε το ποσοστό αναγνώρισης (patient score) ορίζεται ως:

$$Patient\ Score = \frac{N}{N_p}$$

και η ακρίβεια σε επίπεδο ασθενούς, σε σύνολο  $K$  ασθενών ως:

$$Patient\ Level\ Accuracy = \frac{\sum_{i=1}^K patient\_score_i}{K}$$

## 2.4 Μεταφορά μάθησης

Τρωτό σημείο των συνελκτικών δικτύων εκπαιδευόμενων με επιβλεπόμενη μάθηση, βαθιών και μη, αποτελεί το γεγονός ότι για να αναπτύξουν μία ικανότητα γενίκευσης - υπό την έννοια της επιτυχούς αναγνώρισης άγνωστων προτύπων - σε ικανοποιητικό βαθμό, απαιτείται η εκπαίδευσή τους με ένα μεγάλο αριθμό δεδομένων εισόδου. Εύκολα είναι σαφές ότι μια τέτοια πρακτική αφενός είναι κοστοβόρα, αφού μεγάλο πλήθος δεδομένων συνεπάγεται και υψηλές απαιτήσεις σε μνήμη, άρα και υλικό, αφετέρου μη εφαρμόσιμη αφού δεν είναι εφικτό για όλες τις εφαρμογές να συλλέγεται τεράστιο πλήθος δεδομένων. Τη θεώρηση αυτή ήρθε να ανατρέψει η μεταφορά μάθησης, της οποίας βασική αρχή αποτελεί η εκμετάλλευση της ικανότητας εξαγωγής χαρακτηριστικών αρχιτεκτονικών, προεκπαιδευμένων σε ένα υπάρχον πολύ μεγάλο σύνολο δεδομένων, στην αντιμετώπιση προβλημάτων με πολύ μικρότερα σύνολα δεδομένων, είτε όμοιων είτε και αρκετά διαφορετικών από αυτά που εκπαιδεύτηκαν αρχικά.

Ειδικά στο πεδίο της κατηγοριοποίησης εικόνων, η εκπαίδευση ενός βαθιού συνελκτικού δικτύου σε ένα πολύ μεγάλο σύνολο από εικόνες γενικού χαρακτήρα (π.χ αντικειμένων, ανθρώπων, ζώων κ.τ.λ.), έχει ως αποτέλεσμα την ανάπτυξη ενός μοντέλου εξαιρετικά ικανού να διακρίνει αδρά χαρακτηριστικά και σχήματα στα πρώτα επίπεδα και μοτίβα μεγαλύτερης λεπτομέρειας στα ανώτερα. Αυτή η γνώση αποδεικνύεται εξαιρετικά ωφέλιμη στην αναγνώριση εικόνων, για τις οποίες δεν έχει εκπαιδευτεί αρχικά το μοντέλο, αφού ακόμα και αν το περιεχόμενο των είναι πολύ διαφορετικό, βασικά χαρακτηριστικά, όπως η ύπαρξη ακμών και βασικών σχημάτων συνεχίζουν να υφίστανται.

Ένα τέτοιο πολυπληθές σύνολο δεδομένων, το ImageNet, δημιουργήθηκε από τους Deng et al. [16], προκειμένου αρχικά να χρησιμοποιηθεί για την αξιολόγηση της επίδοσης διαφόρων αρχιτεκτονικών ως προς την κατηγοριοποίηση εικόνων και την αναγνώριση αντικειμένων σε μεγάλη κλίμακα μέσω του διαγωνισμού Imagenet Large Scale Visual Recognition Challenge (ILSVRC) [17]. Διεξήχθη για πρώτη χρονιά το 2010 και συνέχισε για αρκετά χρόνια, με κάθε χρονιά να δοκιμάζονται και να κατατάσσονται πρώτες αρχιτεκτονικές με διαφορετικές τεχνολογίες. Οι καλύτερες αρχιτεκτονικές εξ αυτών, μαζί με τα βάρη τους όπως αυτά προέκυψαν από την εκπαίδευση στο ImageNet



[16], ενσωματώθηκαν σε πολλές βιβλιοθήκες μηχανικής μάθησης, ως μοντέλα μεταφοράς μάθησης αφού διαπιστώθηκε η αποτελεσματικότητά τους ως προς την εξαγωγή χαρακτηριστικών. Έτσι, τέθηκε στη διάθεση των ερευνητών ένα μεγάλο πλήθος ισχυρών μοντέλων, τα οποία μπορούσαν να προσαρμοστούν σε διαφορετικές εφαρμογές, επιδεικνύοντας μεγάλες δυνατότητες γενίκευσης, χωρίς να απαιτείται ένας τεράστιος όγκος εξειδικευμένων δεδομένων για την εκπαίδευσή τους. Στη συνέχεια παρουσιάζονται οι βασικότερες στρατηγικές και αρχιτεκτονικές μεταφορές μάθησης.

### 2.4.1 Στρατηγικές

Ανεξαρτήτως της στρατηγικής μεταφοράς μάθησης που θα εφαρμοστεί, πρώτο βήμα αποτελεί η εκπαίδευση του δικτύου σε ένα πολύ μεγάλο σύνολο δεδομένων όπως το ImageNet. Από εκεί και πέραν είναι στη διακριτική ευχέρεια των ερευνητών το πόσο και αν θα προσαρμόσει (fine-tune) το προεκπαιδευμένο μοντέλο για το εκάστοτε πρόβλημα που εξετάζει. Ανάλογα με το πλήθος και το είδος των δεδομένων του ειδικού προβλήματος, το προεκπαιδευμένο μοντέλο μπορεί να χρησιμοποιηθεί είτε στατικά για *εξαγωγή χαρακτηριστικών*, είτε να *προσαρμοστεί (fine-tune)* στο ειδικό σύνολο δεδομένων.

**Εξαγωγή χαρακτηριστικών** Η στρατηγική αυτή εφαρμόζεται συνήθως όταν το σύνολο των ειδικών δεδομένων του προβλήματος είναι αρκετά μικρό αλλά παρόμοιο ως προς το περιεχόμενο με το σύνολο προεκπαίδευσης. Το προεκπαιδευμένο μοντέλο χρησιμοποιείται ως έχει με μόνη παρέμβαση, την αφαίρεση του τελευταίου επιπέδου και την αντικατάστασή του με έναν ταξινομητή προσαρμοσμένο στα δεδομένα του προβλήματος -προσαρμογή του πλήθους των κλάσεων και του είδους του ταξινομητή-. Από το συνολικό μοντέλο εκπαιδεύεται μόνο το επίπεδο εξόδου με όλα τα βάρη των προηγούμενων επιπέδων να διατηρούνται σταθερά στις τιμές που προέκυψαν από την προεκπαίδευση. Για προβλήματα με λίγα και ανόμοια δεδομένα με αυτά της προεκπαίδευσης, υπάρχει και η δυνατότητα εξαγωγής χαρακτηριστικών μόνο από τα κατώτερα επίπεδα του μοντέλου και της χρήσης αυτών για την εκπαίδευση του ταξινομητή.

**Προσαρμογή - Fine-tuning** Μία άλλη στρατηγική που ενδείκνυται όταν το πλήθος των ειδικών δεδομένων είναι σχετικά μεγάλο, περιλαμβάνει εκτός από την αντικατάσταση του επιπέδου εξόδου, όπως αναφέρθηκε παραπάνω, με τον κατάλληλο ταξινομητή, την εκπαίδευση και των υπόλοιπων επιπέδων (όλων ή μερικών), αφού αυτά αρχικοποιηθούν με τις τιμές που προέκυψαν από την αρχική εκπαίδευση του μοντέλου στο μεγάλο σύνολο δεδομένων. Ανάλογα με την ομοιότητα των ειδικών δεδομένων με αυτά της προεκπαίδευσης συνιστάται η συνέχιση της εκπαίδευσης (fine-tuning) περισσότερων ή λιγότερων επιπέδων. Για δεδομένα με όμοια χαμηλού και μέσου επιπέδου χαρακτηριστικά απαιτείται η συνέχιση της εκπαίδευσης λιγότερων επιπέδων ενώ για ανόμοια δεδομένα περισσότερων ή και όλων. Η επιλογή του αριθμού των επιπέδων που θα συνεχίσουν να εκπαιδεύονται καθορίζει και το συνολικό αριθμό των παραμέτρων εκπαίδευσης και για το λόγο αυτό πρέπει να γίνεται με προσοχή καθότι ένα μεγάλο πλήθος παραμέτρων εκπαίδευσης σε συνδυασμό με ένα μικρό σχετικά σύνολο δεδομένων μπορεί να οδηγήσει σε υπερπροσαρμογή (overfitting) του μοντέλου.

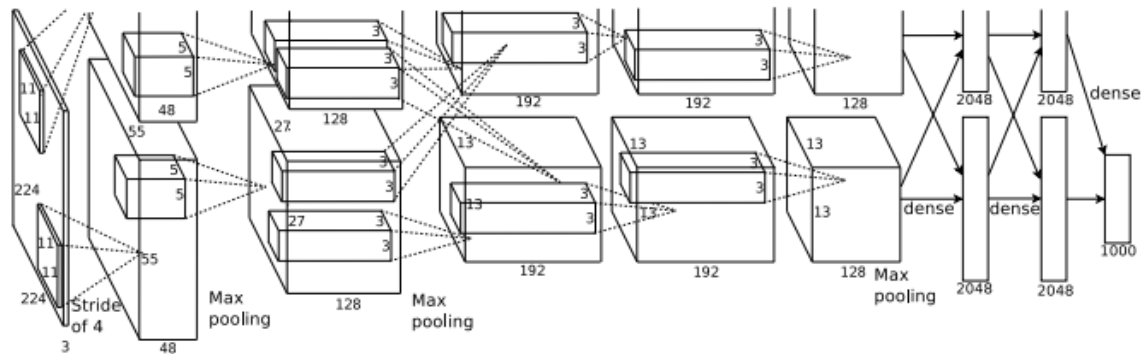


## 2.4.2 Δείγματα Αρχιτεκτονικών

Δεδομένου του πολύ μεγάλου όγκου δεδομένων του ImageNet άρα και του χρόνου που απαιτείται για την εκπαίδευση μιας αρχιτεκτονικής πάνω σε αυτό, η πιο συνηθισμένη πρακτική για την εφαρμογή μεταφοράς μάθησης, είναι να χρησιμοποιούνται μοντέλα που διακρίθηκαν στο ILSVRC, τα οποία είναι διαθέσιμα σε δημόσια αποθετήρια ή βιβλιοθήκες. Τα βασικότερα από αυτά παρουσιάζονται παρακάτω.

**LeNet** Η παλαιότερη αρχιτεκτονική εξέχουσας ιστορικής σημασίας, εισήχθη το 1998 από τους LeCun et al. [18] και ήταν εξειδικευμένη στην αναγνώριση χειρόγραφων ψηφίων. Αν κι εξαιρετικά απλή, αποτελείται από συνελκτικά, πλήρως συνδεδεμένα και pooling επίπεδα, κρατήθηκε για αρκετά χρόνια στο προσκήνιο ως πρωτοπόρα στο πεδίο της βαθιάς μηχανικής μάθησης.

**AlexNet** Το 2012 οι Krizhevsky et al. [19] ανέπτυξαν την παρούσα αρχιτεκτονική, νικώντας με μεγάλη απόσταση το ILSVRC της ίδιας χρονιάς. Το μοντέλο τους αν και παρόμοιο με το LeNet, ήταν βαθύτερο, με περισσότερα φίλτρα ανά συνελκτικό επίπεδο, ενώ καινοτομία σε σχέση με το LeNet αποτέλεσε αφενός η τοποθέτηση διαδοχικών συνελκτικών επιπέδων -χωρίς να μεσολαβεί pooling επίπεδο όπως συνηθιζόταν μέχρι τότε-, αφετέρου η προσθήκη επιπέδων dropout. Για την εκπαίδευσή του επιστρατεύτηκαν μονάδες επεξεργασίας γραφικών, οδηγώντας σε δραματική μείωση του χρόνου εκπαίδευσης, ενώ οι υψηλές του επιδόσεις το έκαναν να αποτελέσει σταθμό στη βαθιά μάθηση.

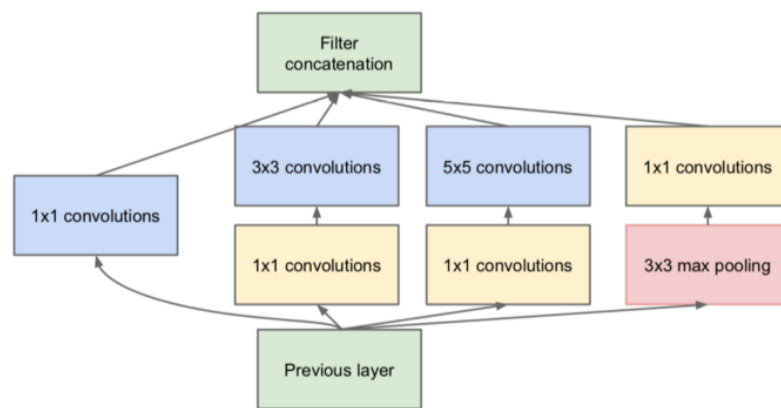


Σχήμα 2.4: Η αρχιτεκτονική AlexNet

**VGG** Ένα από τα πιο δημοφιλή μοντέλα, ως προς την εξαγωγή χαρακτηριστικών λόγω της ομοιόμορφης αρχιτεκτονικής του, το VGG εισήχθη από τους Simonyan και Zisserman [20] κι έλαβε τη δεύτερη θέση στο ILSVRC το 2014. Ακολουθώντας, παρόμοια λογική με αυτή του AlexNet, διατήρησαν μεγάλο πλήθος φίλτρων, μειώνοντας ωστόσο τις διαστάσεις μήκους και πλάτους του πυρήνα, με στόχο την αναπαράσταση της επίδρασης πυρήνων μεγαλύτερων διαστάσεων, με περισσότερα φίλτρα μικρότερων με στόχο τη μείωση των παραμέτρων. Υπήρξαν δύο παραλλαγές του μία με 16 και μία με 19 επίπεδα, με την πρώτη να είναι η πιο διαδεδομένη, έχοντας ωστόσο μεγάλο πλήθος παραμέτρων.

**Inception** Μεγάλος νικητής του ILSVRC το 2014 ήταν η αρχιτεκτονική GoogLeNet, γνωστή και ως Inception V1 [21], η οποία παρουσίασε εξαιρετικές επιδόσεις, με σημαντικά μικρότερο πλήθος

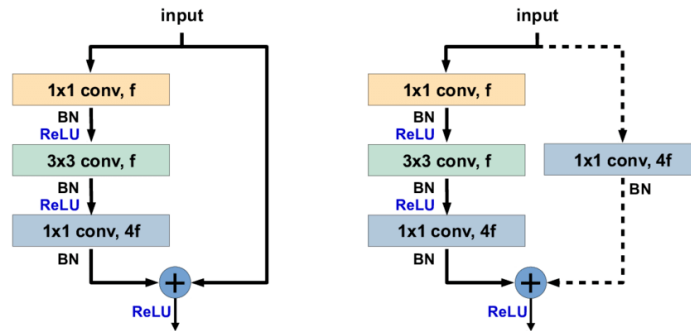
παραμέτρων σε σχέση με προκατόχους του, όπως το AlexNet. Πηγή έμπνευσης για το μοντέλο αποτέλεσε η αρχιτεκτονική LeNet με την προσθήκη μίας καινοτόμας δομικής μονάδας, που έλαβε το όνομα Inception. Βασικό χαρακτηριστικό του σύνθετου αυτού στρώματος, αποτελεί η δημιουργία παράλληλων κλάδων με συνελκτικά επίπεδα, με πυρήνες διαφορετικού μεγέθους, με σκοπό την ταυτόχρονη εξαγωγή χαρακτηριστικών τοπικού και καθολικού χαρακτήρα μέσω πυρήνων μικρότερων και μεγαλύτερων διαστάσεων αντίστοιχα. Ακόμη, εφαρμόστηκε και η τεχνική κανονικοποίησης του υποσυνόλου των δεδομένων κάθε βήματος - δέσμης (batch normalization). Το μικρό πλήθος παραμέτρων του, παρόλο το πλήθος των επιπέδων του (22), μείωσε δραστικά το πλήθος των υπολογισμών, τραβώντας την προσοχή για την ανάπτυξη εφαρμογών πολύ μεγάλης κλίμακας.



Σχήμα 2.5: Το στρώμα Inception

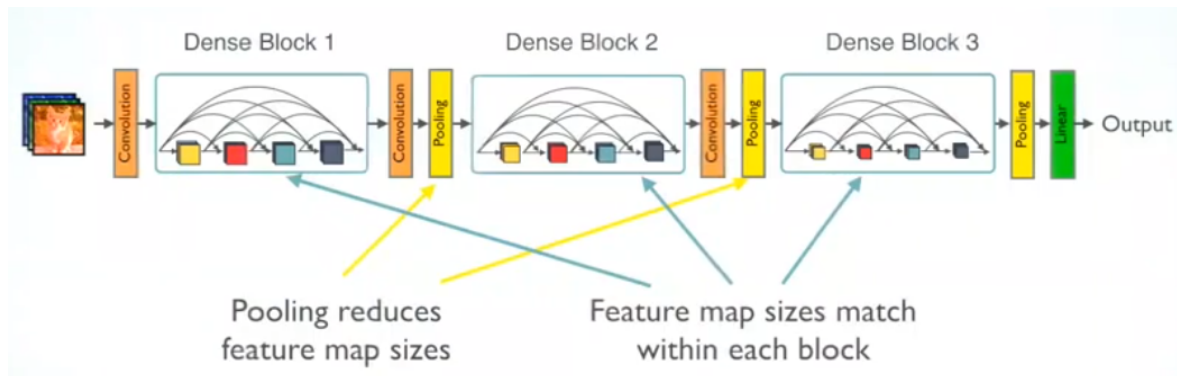
**ResNet** Νικητής του ILSVRC για το έτος 2015 ανεδείχθη η αρχιτεκτονική ResNet των He et al. [22] με 152 επίπεδα. Καινοτομία του εν λόγω μοντέλου αποτελεί η εισαγωγή ενός τύπου σύνθετου στρώματος, με τον όρο residual block, αποτελούμενο από ένα μικρό πλήθος συνελκτικών επιπέδων, με την είσοδο του στρώματος αυτού να προστίθεται απευθείας στην έξοδο, παραλείποντας ενδιάμεσες συνδέσεις. Η λογική αυτή οδήγησε σε δύο είδη τέτοιων δομικών μονάδων ανάλογα με το αν η απευθείας σύνδεση περιείχε ή όχι κι ένα παράλληλο επίπεδο, συνιστώντας τις δομικές μονάδες ταυτότητας-identity και προβολής-projection αντίστοιχα. Το πρώτο είδος επιλεγόταν σε περιπτώσεις όπου η είσοδος στο στρώμα είχε ίδιες διαστάσεις με την έξοδο, ενώ το δεύτερο διαφορετικά, με το παρεμβαλλόμενο επίπεδο να μπορεί να είναι είτε ένα απλό επίπεδο παραγεμίσματος με μηδενικά για τη διόρθωση των διαστάσεων, είτε ένα συνελκτικό επίπεδο. Στόχος της συγκεκριμένης συνδεσμολογίας είναι η αντιμετώπιση του προβλήματος εξαφανιζόμενης παραγωγού (vanishing gradient) που μπορεί να προκληθεί και από την τοποθέτηση πολλών διαδοχικών επιπέδων ακολουθιακά. Παρόλο το μεγάλο βάθος της αρχιτεκτονικής, το πρόβλημα αυτό δε προκύπτει, ενώ αρκετά μικρό, σε σχέση και με τον αριθμό των επιπέδων, είναι και το πλήθος των παραμέτρων. Έχουν αναπτυχθεί διαφορετικά μοντέλα ίδιας τεχνοτροπίας με μικρότερο πλήθος επιπέδων ενώ εξαιρετικό ενδιαφέρον παρουσιάζει κι ένα υβριδικό μοντέλο που συνδυάζει δομικά στοιχεία της εν λόγω αλλά και της Inception αρχιτεκτονικής [23].

**DenseNet** Μία αρκετά διαδεδομένη προεκπαιδευμένη αρχιτεκτονική εισήχθη το 2017 από τους Huang et al. [24], παρουσιάζοντας αντίστοιχες επιδόσεις με εκείνες των αρχιτεκτονικών ResNet, με



Σχήμα 2.6: Οι residual δομικές μονάδες ταυτότητας-identity (αριστερά) και προβολής-projection (δεξιά)

όμως σημαντικά μικρότερο πλήθος παραμέτρων. Πρόκειται για μία βαθιά συνελκτική αρχιτεκτονική αποτελούμενη στο κύριο μέρος από πυκνά σύνθετα επίπεδα (dense blocks) και μεταβατικά επίπεδα (transition blocks). Ιδιαιτερότητα των πρώτων, η οποία και τα διακρίνει από τις άλλες αρχιτεκτονικές, αποτελεί το γεγονός ότι η έξοδος κάθε επιπέδου εντός ενός dense block, τροφοδοτεί την είσοδο και κάθε επόμενου -εντός του dense block-, με την τελευταία να προκύπτει από τη συνένωση των διανυσμάτων χαρακτηριστικών της εξόδου των προηγούμενων επιπέδων. Τα μεταβατικά επίπεδα περιέχουν κυρίως συνελκτικά και pooling επίπεδα με μια ενδεικτική σκιαγράφιση της αρχιτεκτονικής να παρουσιάζεται στο σχήμα 2.7

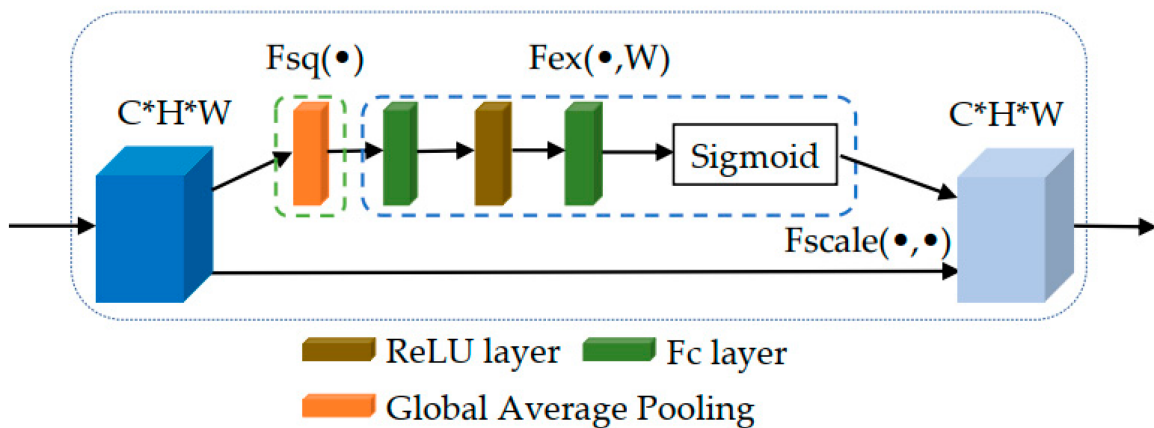


Σχήμα 2.7: Δείγμα αρχιτεκτονικής DenseNet

Η πυκνή συνδεσιμότητα του παρόντος μοντέλου προσδίδει το πλεονέκτημα της ευκολότερης διάδοσης του σφάλματος μεταξύ των επιπέδων, αντιμετωπίζοντας το πρόβλημα της εξαφανιζόμενης παραγώγου. Στα θετικά συγκαταλέγονται επίσης και η διάδοση κι επαναχρησιμοποίηση ποικιλόμορφων χαρακτηριστικών από διαφορετικά επίπεδα, οδηγώντας σε πλουσιότερες αναπαραστάσεις των μοτίβων πληροφορίας. Ο συνδυασμός δε των παραπάνω με το μικρό αριθμό παραμέτρων και το μικρότερο υπολογιστικό κόστος, καθιστούν το παρόν μοντέλο μια δυνατή επιλογή στη διάθεση των ερευνητών, η οποία και χρησιμοποιείται στα πειράματα.

**SENet** Μεγάλος νικητής του ILSVRC 2017 αποτέλεσε μοντέλο που ενσωμάτωσε την τεχνοτροπία συμπίεσης-διέγερσης (Squeeze-and-Excitation - SE), η οποία προτάθηκε από τους Hu et al. [25] και αφορά την εισαγωγή μιας καινοτόμας δομικής μονάδας (Squeeze-and-Excitation block), η οποία αναπροσαρμόζει τις τιμές ενός τρισδιάστατου διανύσματος χαρακτηριστικών ανά επίπεδο βάθους,

ή αλλιώς ανά κανάλι, μέσω της αναλυτικής μοντελοποίησης των αλληλεξαρτήσεων μεταξύ των καναλιών. Κύρια συστατικά της δομικής αυτής μονάδας αποτελούν ένα επίπεδο εξαγωγής καθολικής μέσης τιμής ανά κανάλι (Global Average Pooling), ακολουθούμενο από δύο διαδοχικά πλήρως συνδεδεμένα επίπεδα με συναρτήσεις ενεργοποίησης ReLU και Sigmoid αντίστοιχα. Το πρώτο υλοποιεί τη λειτουργία συμπίεσης (squeeze), εξάγοντας ανά κανάλι στατιστική πληροφορία καθολικού χαρακτήρα, ενώ τα δύο επόμενα τη λειτουργία διέγερσης (excitation), μέσα από τα οποία αναδεικνύονται οι εσωτερικές σχέσεις κι εξαρτήσεις μεταξύ των καναλιών. Η έξοδος του τελευταίου πλήρους επιπέδου πολλαπλασιάζεται με το διάνυσμα εισόδου της δομικής μονάδας, προκειμένου αυτό να αναπροσαρμοστεί, με βάση την πληροφορία που εξήχθη.



Σχήμα 2.8: Δομική Μονάδα Squeeze-and-Excitation

Μια ενδεικτική απεικόνιση των παραπάνω φαίνεται στο σχήμα 2.8, με  $F_{sq}$ ,  $F_{ex}$ ,  $F_{scale}$  να συμβολίζουν τις λειτουργίες συμπίεσης, διέγερσης και πολλαπλασιασμού αντίστοιχα και με  $H$ ,  $W$ ,  $C$ , οι διαστάσεις μήκους, πλάτους και βάθους αντίστοιχα. Συνήθης πρακτική αποτελεί ο συνδυασμός της παρούσας δομικής μονάδας με άλλες μονάδες π.χ. τύπου Inception, ResNet κ.ά., με το διάνυσμα εισόδου να προκύπτει από κάποια από αυτές δημιουργώντας μία σύνθετη δομική μονάδα, αποτελώντας βάση για διάφορα νέα μοντέλα. Τα υβριδικά αυτά μοντέλα συνδυάζουν τις ιδιότητες αμφοτέρων των δομικών στοιχείων που τις αποτελούν, επιδεικνύοντας υψηλές ικανότητες γενίκευσης. Χαρακτηριστικό παράδειγμα η επικρατούσα αρχιτεκτονική του ILSVRC 2017, η οποία συνδύασε τις αρχιτεκτονικές ResNet και SE, επιτυγχάνοντας ακόμη καλύτερη επίδοση. Η πρακτική αυτή έχει ακολουθηθεί και σε ένα μέρος της πειραματικής διαδικασίας, συνδυάζοντας τις αρχιτεκτονικές DenseNet και SENet.

## Κεφάλαιο 3

### Ανασκόπηση Ερευνητικού Έργου

Αναφορικά με την ανάπτυξη ενός ευφυούς συστήματος υποβοήθησης ιατρικής διάγνωσης καρκίνου του μαστού με βάση ιστοπαθολογικές εικόνες, η ερευνητική κοινότητα έχει επιστρατεύσει αρκετές διαφορετικές τεχνικές και προσεγγίσεις, οι οποίες εμπίπτουν στην πλειοψηφία τους στο πεδίο της επιβλεπόμενης μάθησης, ενώ σε ορισμένες εργασίες έχουν εφαρμοστεί επίσης τεχνικές ημι- και μη επιβλεπόμενης αλλά κι ενισχυτικής μάθησης. Οι τεχνικές αυτές αποσκοπούν στην επίλυση δύο εκφάνσεων του προβλήματος της κατηγοριοποίησης ιστοπαθολογικών εικόνων μικροσκοπίας, της δυαδικής κατηγοριοποίησης, κατά την οποία το σύστημα αποφαινεται εάν το είδος της αλλοίωσης που εντοπίζεται στις ιστοπαθολογικές εικόνες του ασθενούς είναι καλοήθης ή κακοήθης, και της κατηγοριοποίησης σε υποτύπους της αλλοίωσης. Ένας ακόμη παράγοντας διάκρισης του ευρύτερου προβλήματος αποτελεί και η χρήση ή μη εικόνων μικροσκοπίας διαφορετικής μεγέθυνσης κατά την εκπαίδευση και τον έλεγχο του συστήματος, δηλαδή το κατά πόσο εικόνες διαφορετικής μεγέθυνσης θα αποτελέσουν ένα ενιαίο σύνολο δεδομένων ή διακριτά ανεξάρτητα υποσύνολα. Για κάθε μία από αυτές τις πτυχές του αρχικού προβλήματος, παρατίθενται στη συνέχεια οι προσεγγίσεις διαφόρων ερευνητικών ομάδων που συναντώνται στη βιβλιογραφία, αναφορικά με το σύνολο δεδομένων BreakHis [3].

#### 3.1 Δυαδική Κατηγοριοποίηση Κοινής Μεγέθυνσης

Στις πρώτες προσπάθειες επίλυσης ακολουθήθηκε η μέθοδος εξαγωγής χαρακτηριστικών με κλασικές τεχνικές επεξεργασίας ψηφιακής εικόνας μέσω διαφόρων αναπαραστάσεων υφής και περιγραφητών [3]. Τα χαρακτηριστικά αυτά στη συνέχεια αποτέλεσαν την είσοδο σε απλούς ταξινομητές μηχανικής μάθησης και συγκεκριμένα κοντινότερου γείτονα (1-NN), τετραγωνικής γραμμικής ανάλυσης (QDA), διανυσμάτων υποστήριξης (SVM) και τυχαίων δασών (RF) ενώ δοκιμάστηκε και η δυναμική επιλογή ταξινομητών μέσω μαντείου που προβλέπει τον καταλληλότερο ταξινομητή για κάθε δείγμα ελέγχου. Η ακρίβεια για τους καλύτερους συνδυασμούς περιγραφητών και ταξινομητών ήταν της τάξης 70-80% για όλες τις μεγεθύνσεις ενώ η ακρίβεια με τη χρήση του μαντείου κυμάνθηκε από 97.7% έως και 100% ανάλογα με τη μεγέθυνση.

Η εξαιρετική ικανότητα αυτόματης εξαγωγής χαρακτηριστικών των βαθιών συνελκτικών δικτύων, χωρίς να απαιτείται η ύπαρξη ειδικών γνώσεων επεξεργασίας εικόνας σαφώς αποτέλεσε πόλο έλξης των ερευνητών, οι πλειοψηφία των οποίων στράφηκε σε βαθιά νευρωνικά δίκτυα στις μετέπειτα προσπάθειες. Οι Spanhol et al. [26] ανέπτυξαν ένα βαθύ δίκτυο πέντε επιπέδων αποτελούμενο από τρία συνελκτικά στρώματα και δύο πλήρως συνδεδεμένα, βασισμένο στην αρχιτεκτονική AlexNet [19] καταφέροντας υψηλότερη ακρίβεια από εκείνη που είχαν πετύχει με τους κορυφαίους περιγραφη-

τές υφής [3]. Πέραν από τη χρήση συνελκτικών επιπέδων, στη συνεισφορά του [26] συγκαταλέγεται ακόμη η υιοθέτηση και μελέτη της στρατηγικής εξαγωγής σημαντικά μικρότερων επιμέρους, επικαλυπτόμενων ή μη, τμημάτων της αρχικής εικόνας, τα οποία αποτέλεσαν την είσοδο στην προαναφερθείσα αρχιτεκτονική. Αντίστοιχη στρατηγική εφαρμόστηκε και σε μετέπειτα μελέτη των Spanhol et al. [27] πάνω στην χρήση και το συνδυασμό DeCAF χαρακτηριστικών [28] ως είσοδο ταξινομητή, τα οποία εξήχθησαν από μοντέλο εκπαιδευμένο στο σύνολο δεδομένων ImageNet [16], πετυχαίνοντας συγκρίσιμα αποτελέσματα με το [26], ακολουθώντας μια απλούστερη και ταχύτερη διαδικασία εκπαίδευσης.

Εκτενέστερη μελέτη σχετικά με τα οφέλη της μεταφοράς μάθησης παρουσίασαν οι Matos et al. [29], οι οποίοι αξιοποίησαν τη γνώση πάνω σε δύο σύνολα δεδομένων, το ImageNet [16] και το CRC [30], με το πρώτο να συνεισφέρει στην αναπαράσταση χαρακτηριστικών και το δεύτερο να προσφέρει πληροφορίες ως προς τη δομή των ιστοπαθολογικών ιστών. Συγκεκριμένα, εκπαιδεύσαν σε πρώτη φάση έναν SVM ταξινομητή πάνω στο CRC σύνολο δεδομένων, έτσι ώστε να μπορεί να διακρίνει ανάμεσα σε σχετικά ή μη τμήματα εικόνων ανάλογα με τον τύπο τους -δοκιμάστηκαν διαφορετικοί συνδυασμοί κριτηρίων σχετικότητας-. Το διάλυμα χαρακτηριστικών που αποτελούσε την είσοδο στον ταξινομητή αυτό, προέκυπτε από το προ-εκπαιδευμένο στο ImageNet Inception-v3 μοντέλο [31]. Στη συνέχεια, ο ταξινομητής χρησιμοποιήθηκε για να απορρίψει τα μη σχετικά τμήματα των εικόνων του κυρίου συνόλου δεδομένων του προβλήματος και μόνον όσα έκρινε σχετικά χρησιμοποιήθηκαν για την εκπαίδευση του τελικού ταξινομητή -επίσης SVM-. Δεδομένης της μεγάλης διάστασης του διανύσματος χαρακτηριστικών του Inception-v3, μελετήθηκε επίσης η εφαρμογή της τεχνικής μείωσης διαστατικότητας PCA. Η βέλτιστη ακρίβεια για τις παραπάνω μεθόδους άγγιξε το 91% για μεγέθυνση  $100\times$ , ενώ για  $40\times$  και  $200\times$  ήταν επίσης της τάξης του 89%.

Αναφορικά με τον τομέα της ημι-επιβλεπόμενης μάθησης οι Sudharshan et al. [32] μελέτησαν διαφορετικές μεθόδους μάθησης πολλαπλών στιγμιότυπων, η οποία συνίσταται στην οργάνωση των στιγμιότυπων σε ομάδες για τον χαρακτηρισμό των οποίων δεν απαιτείται πάντα η γνώση των ετικετών όλων των στιγμιότυπων τους. Για το συγκεκριμένο πρόβλημα ως στιγμιότυπα θεωρήθηκαν μικρά υποτμήματα των αρχικών εικόνων, ενώ ως ομάδα θεωρήθηκε τόσο το σύνολο των υποτμημάτων των εικόνων που ανήκαν στον ίδιο ασθενή, όσο και το σύνολο των υποτμημάτων που ανήκαν στην ίδια εικόνα. Μία ομάδα χαρακτηριζόταν ως θετική εφόσον περιείχε τουλάχιστον ένα στιγμιότυπο κακοήθους αλλοιώσεως και αρνητική διαφορετικά. Από τη μελέτη διαφορετικών μεθόδων [32] που εστίαζαν σε επίπεδο στιγμιότυπων αλλά και μεθόδων που βασίζονταν στη διαφοροποίηση σε επίπεδο ομάδων, καλύτερα αποτελέσματα προέκυψαν για τις τελευταίες σε κάθε μεγέθυνση και για τα δύο επίπεδα ομαδοποίησης. Συγκεκριμένα, υψηλότερη ακρίβεια σε όλα τα διαφορετικά επίπεδα μεγέθυνσης που άγγιξε το 92% για μεγέθυνση  $40\times$  επιτεύχθηκε με τη μη παραμετρική μέθοδο μάθησης πολλαπλών στιγμιότυπων [33], η οποία αποτελεί μια πιο εύρωστη παραλλαγή του ταξινομητή κοντινότερου γείτονα (k-NN), ενώ εξίσου ικανοποιητική ήταν και η βασισμένη σε συνελκτικά νευρωνικά δίκτυα μέθοδος πολλαπλών στιγμιότυπων [34].

Επιπρόσθετα, έρευνα έχει γίνει και για μεθόδους που συνδυάζουν τεχνικές μη επιβλεπόμενης κι επιβλεπόμενης μάθησης. Οι Nahid et al. [35] πειραματίστηκαν με βαθιά συνελκτικά και αναδρομικά μοντέλα καθοδηγούμενα από μεθόδους ομαδοποίησης μη επιβλεπόμενης μάθησης, μετασχηματίζοντας τα αρχικά δεδομένα μέσω των αλγορίθμων ομαδοποίησης K-μέσου (K-means clustering) και μετατόπισης μέσου (Mean shift clustering). Σκοπός τους ήταν μέσω της μη επιβλεπόμενης μάθησης να αναδειχθούν τα κοινά μοτίβα πληροφορίας που χαρακτηρίζουν εγγενώς τη δομή των αλλοιώσεων



που ανήκουν στην ίδια κλάση. Στα πειράματά ο βέλτιστος συνδυασμοί αλγόριθμοι ομαδοποίησης και μοντέλου φάνηκε να είναι το βαθύ συνελικτικό και ο Mean shift με ακρίβεια γύρω στο 90% για την πλειοψηφία των μεγεθύνσεων.

Εξαιρετικά αποτελέσματα με ακρίβεια γύρω στο 98% για όλα τα επίπεδα μεγέθυνσης, πέτυχαν οι Pratiher et al. [36] εκμεταλλευόμενοι τις δυνατότητες μη επιβλεπόμενης μάθησης χαρακτηριστικών του αραιού κωδικοποιητή στοίβας (Stacked Sparse Autoencoder-SSAE). Καινοτομία τους ακόμα αποτελεί η εφαρμογή της τεχνικής μείωσης διαστατικότητας Landmark-Isomap [37] σε συνδυασμό με τον παραπάνω κωδικοποιητή.

### 3.2 Δυαδική Κατηγοριοποίηση Ανεξαρτήτου Μεγέθυνσης

Η διάκριση των εικόνων μικροσκοπίας με βάση τη μεγέθυνσή τους δημιουργεί την ανάγκη ανάπτυξης ξεχωριστών μοντέλων, αυξάνοντας την πολυπλοκότητα του σχεδιασμού και περιορίζοντας τις δυνατότητες γενίκευσης. Για το λόγο αυτό υπήρξε στροφή προς την ανάπτυξη μοντέλων που θα μπορούσαν να αναγνωρίζουν το είδος των αλλοιώσεων χωρίς να λαμβάνεται υπόψη το επίπεδο μεγέθυνσης. Αρχικά, οι Bayramoglu et al. [38] με ένα απλό σχετικά συνελικτικό δίκτυο και εφαρμόζοντας διάφορες τεχνικές επαύξησης δεδομένων πέτυχαν αξιολογικά αποτελέσματα ακρίβειας γύρω στο 83%, ενώ δοκιμάστηκε και παραλλαγή του μοντέλου για ταυτόχρονη πρόβλεψη του είδους και της μεγέθυνσης της εικόνας. Σε μετέπειτα μελέτη [39] δοκιμάστηκε η εξαγωγή χαρακτηριστικών μέσω περιγραφητών χρώματος-δομής και η ταξινόμησή τους από ένα σύνολο (Ensemble) [40] τυπικών ταξινομητών επιβλεπόμενης μάθησης (όπως SVM, kNN κ.ά.) με σκοπό να συγκριθεί η ικανότητα γενίκευσης σε επίπεδο μεγέθυνσης μοντέλων εκπαιδευμένων σε ορισμένη μεγέθυνση και μοντέλων εκπαιδευμένων σε όλες μαζί. Από τα αποτελέσματα των πειραμάτων προέκυψε αφενός ότι το εκπαιδευμένο σε σύνολο δεδομένων ανεξαρτήτου μεγέθυνσης μοντέλο να μην είχε υψηλότερη ακρίβεια που άγγιξε και το 90%, αφετέρου ότι μοντέλο εκπαιδευμένο σε ορισμένη μεγέθυνση έδωσε προβλέψεις για εικόνες διαφορετικής μεγέθυνσης με ακρίβεια να μην μικρότερη αλλά αρκετά κοντά στο πρώτου.

Πειράματα [41] έχουν πραγματοποιηθεί και όσον αφορά τη μεταφορά μάθησης και συγκεκριμένα με τα προεκπαιδευμένα στο ImageNet [16] μοντέλα VGG16 [20], VGG19 [20], ResNet50 [22], τα οποία μελετήθηκαν και ως αρχιτεκτονικές, δηλαδή εκπαιδεύτηκαν με τυχαία αρχικοποιημένα βάρη. Εξετάστηκε ακόμη και η επίδραση των ποσοστών διαχωρισμού των δεδομένων σε σύνολο εκπαίδευσης κι ελέγχου. Από τα παραπάνω βέλτιστη ακρίβεια 92.6% επιτεύχθηκε για το VGG16 με μεταφορά μάθησης, διαχωρισμό 90% για εκπαίδευση, 10% για έλεγχο κι εφαρμογή τεχνικών επαύξησης των αρχικών δεδομένων. Τη μεταφορά μάθησης εκμεταλλεύτηκαν και οι Gupta και Bhavsar [42], αναπτύσσοντας ένα μοντέλο βασισμένο στην προεκπαιδευμένη αρχιτεκτονική ResNet [22], με την καινοτομία ότι η τελική πρόβλεψη στηριζόταν στις αποφάσεις με βάση ορισμένα χαρακτηριστικά κάποιων επιλεγμένων ενδιάμεσων επιπέδων. Για την επιλογή των χαρακτηριστικών χρησιμοποιήθηκε η τεχνική μείωσης διαστατικότητας XGBoost [43], για την επιλογή των επιπέδων η μετρική ITS [44], ενώ κύριος άξονας της μελέτης αποτέλεσε η επίδραση της συνένωσης χαρακτηριστικών πολλαπλών επιπέδων, δίνοντας πολύ αξιολογικά αποτελέσματα που έφτασαν το 97%.

Τις δυνατότητες μεταφοράς μάθησης, εκμεταλλεύτηκαν και οι Li et al. [45], με τη διαφορά ότι αξιοποίησαν ένα προεκπαιδευμένο στο ImageNet μοντέλο, το DenseNet121 [24], συνδυάζοντας την αρχιτεκτονική του με την αρχιτεκτονική SENet (Squeeze-and-Excitation Network). Η υβριδική αρχι-

τεκτονική που προέκυψε, συνδυάζοντας τις ιδιότητες αμφοτέρων των συστατικών δομικών μονάδων της παρουσίασε, σε επαυξημένα δεδομένα, αξιόλογες επιδόσεις που άγγιξαν το 89%, διατηρώντας ταυτόχρονα σημαντικά χαμηλό το πλήθος των παραμέτρων.

Η αρχιτεκτονική ResNet [22] χρησιμοποιήθηκε και για την εφαρμογή ενός πρωτοκόλλου εκπαίδευσης [46] κατά το οποίο, από τα αρχικά δεδομένα προέκυψαν ακολουθίες (cases), έτσι ώστε σε κάθε ακολουθία να υπάρχει μία εικόνα από κάθε μεγέθυνση, περιέχοντας μόνο εικόνες του ίδιου είδους αλλοίωσης. Το μοντέλο εκπαιδεύτηκε έχοντας ως είσοδο ένα ορισμένο πλήθος τέτοιων ακολουθιών, η βέλτιστη τιμή του οποίου προέκυψε πειραματικά, αποδίδοντας μέση ακρίβεια αναγνώρισης ακολουθιών 91%.

Ενδιαφέρον παρουσιάζει, ακόμη, η προσέγγιση των Jonnalagedda et al. [47], οι οποίοι εφάρμοσαν τεχνικές μη-επιβλεπόμενης μάθησης για την επαύξηση των αρχικών δεδομένων, αναπτύσσοντας επίσης μια δική τους αρχιτεκτονική με παράλληλους συνελκτικούς κλάδους, προκειμένου να λαμβάνονται υπόψη χαρακτηριστικά τόσο τοπικού όσο και καθολικού χαρακτήρα. Αναφορικά με την επαύξηση των δεδομένων κάθε εικόνα ομαδοποιήθηκε με μείγμα Γκαουσιανών μοντέλων (Gaussian Model Mixture) μέσω του αλγορίθμου μεγιστοποίησης μέσης τιμής (Expectation Maximization) έτσι ώστε να απομονωθούν οι πυρήνες των κυττάρων των ιστών και να εξαχθεί το τμήμα της εικόνας με τη μεγαλύτερη συγκέντρωση ωφέλιμης πληροφορίας, το οποίο χρησιμοποιήθηκε για να εμπλουτίσει το αρχικό σύνολο δεδομένων. Συνδυαστικά, εφαρμόστηκαν και κλασικές τεχνικές επαύξησης με την ακρίβεια του προτεινόμενου μοντέλου στα επαυξημένα δεδομένα να είναι της τάξης του 92%.

### 3.3 Κατηγοριοποίηση Πολλών Κλάσεων Κοινής Μεγέθυνσης

Η πρώτη προσπάθεια αντιμετώπισης του πιο απαιτητικού και πολύπλοκου προβλήματος της κατηγοριοποίησης ιστοπαθολογικών εικόνων σε υποτύπους, έγινε από τους Han et al. [48], οι οποίοι βασίστηκαν στην αρχιτεκτονική GoogLeNet [21]. Στα πειράματά τους με μεταφορά μάθησης από το ImageNet σε επαυξημένα δεδομένα, προσαρμόζοντας τη στρατηγική εκπαίδευσης, έτσι ώστε να ελαχιστοποιείται η Ευκλείδεια απόσταση μεταξύ χαρακτηριστικών εικόνων ίδιου τύπου και να μεγιστοποιείται μεταξύ διαφορετικού, κατάφεραν ακρίβεια γύρω στο 93% σε όλα τα επίπεδα μεγέθυνσης, με πολύ αξιόλογα αποτελέσματα και σχετικά με το δυαδικό πρόβλημα. Σε παρόμοια φιλοσοφία σχετικά με την στρατηγική εκπαίδευσης κινήθηκαν και οι Pan et al. [49], χρησιμοποιώντας μία σύνθετη συνάρτηση απώλειας συνιστάμενη από την απώλεια διασταυρούμενης εντροπίας και μίας συνάρτησης απώλειας βασισμένης στην Ευκλείδεια απόσταση μεταξύ χαρακτηριστικών. Η εκπαίδευση έγινε με διπλό στόχο, αφενός τη βελτιστοποίηση της αναγνωριστικής ικανότητας αφετέρου τη διατήρηση της διασποράς των τιμών των χαρακτηριστικών μεταξύ εικόνων διαφορετικών υποτύπων ή τη μείωσή της μεταξύ όμοιων υποτύπων. Ως μοντέλο χρησιμοποιήθηκε μία παραλλαγή της αρχιτεκτονικής Xception [50] προεκπαιδευμένης στο ImageNet, με την ακρίβεια να είναι της τάξης του 94% για την πλειοψηφία των μεγεθύνσεων.

Την αξία της μεταφοράς μάθησης ανέδειξαν επίσης οι Sharma et al. [51] στη συγκριτική τους μελέτη αναφορικά με την εξαγωγή χαρακτηριστικών μεταξύ βαθιών προεκπαιδευμένων συνελκτικών μοντέλων και συμβατικών τεχνικών επεξεργασίας εικόνας, με την πρώτη να αποδίδει καλύτερα και για την παρούσα πτυχή του προβλήματος. Ειδικότερα, μελετήθηκαν οι αρχιτεκτονικές VGG16, VGG19 [20] και ResNet50 [22] ως εξαγωγείς χαρακτηριστικών τόσο με τυχαία αρχικοποίηση, όσο και με τα βάρη μετά από εκπαίδευση στο ImageNet, ενώ δοκιμάστηκαν και διαφορετικοί ταξινομητές (SVM,



RF, kNN κ.ά.), με τον συνδυασμό προεκπαιδευμένου στο ImageNet VGG16, με ταξινομητή SVM γραμμικού πυρήνα σε ισορροπημένα κι επαυξημένα δεδομένα, να επιτυγχάνει ακρίβεια γύρω στο 93% για όλα τα επίπεδα μεγέθυνσης.

Εξαιρετικά αποτελέσματα με μέγιστη ακρίβεια που άγγιξε το 97% για μεγέθυνση  $100\times$  και μέση ακρίβεια για όλες τις μεγεθύνσεις γύρω στο 95% πέτυχαν οι Nawaz et al. [52], ομοίως με μεταφορά μάθησης, εφαρμόζοντας δύο στάδια εκπαίδευσης σε αρχιτεκτονική βασισμένη στο DenseNet [24] προεκπαιδευμένο στο ImageNet. Στο πρώτο στάδιο τα βάρη όλων των επιπέδων πλην του τελευταίου αρχικοποιούνται με τιμές που έχουν προκύψει από την προεκπαίδευση και παραμένουν σταθερά όσο εκπαιδεύεται το τελευταίο επίπεδο. Στο δεύτερο στάδιο όλα τα βάρη αρχικοποιούνται με τις τιμές που προέκυψαν από το πρώτο στάδιο και μένουν όλα σταθερά πλην εκείνων του πρώτου επιπέδου, το οποίο επίπεδο είναι και το μόνο που εκπαιδεύεται. Στόχος της στρατηγικής αυτής αποτελεί η μείωση της πολυπλοκότητας της εκπαίδευσης αλλά και η μείωση των παραμέτρων εκπαίδευσης προκειμένου να αποφευχθεί η υπερπροσαρμογή.

Αξιοσημείωτη είναι ακόμη η προσέγγιση των Xu et al. [53], οι οποίοι ανέπτυξαν ένα μοντέλο ενισχυτικής μάθησης ικανό να εξάγει τα πλέον σημαντικά υποτιμήματα των αρχικών εικόνων έτσι ώστε να μεγιστοποιείται η ακρίβεια της κατηγοριοποίησης. Ειδικότερα, η επιλογή των κατάλληλων υποτιμημάτων μοντελοποιήθηκε ως μία μερικώς παρατηρήσιμη μαρκοβιανή διαδικασία απόφασης (Partially observed Markov decision process - POMDP) και πραγματοποιούνταν από ένα δίκτυο ανάδρασης μακράς βραχυπρόθεσμης μνήμης (Long short-term memory - LSTM). Στη συνέχεια οι επιλεγθέντες περιοχές τροφοδοτούσαν ένα βαθύ συνελκτικό δίκτυο προσοχής (attention), δύο κλάδων, το οποίο εξήγαγε τα πιο σημαντικά για τη ταξινόμηση χαρακτηριστικά, τα ταξινομούσε και παρήγαγε τα σήματα επιβράβευσης βάσει των οποίων εκπαιδευόταν το LSTM. Το υβριδικό αυτό συνεργατικό μοντέλο απέδωσε υψηλότερη ακρίβεια της τάξης 98% για όλα τα επίπεδα μεγέθυνσης.

Τα τελευταία χρόνια ορισμένες ερευνητικές μελέτες στράφηκαν στην ανάπτυξη μοντέλων ικανών να αντιμετωπίσουν τόσο το πρόβλημα της κατηγοριοποίησης πολλών κλάσεων όσο και της δυαδικής κατηγοριοποίησης. Ένα μέρος αυτών [54, 55] επικεντρώθηκε σε μοντέλα μεταφοράς μάθησης, ενώ άλλες στην ανάπτυξη εξειδικευμένων για το πρόβλημα αρχιτεκτονικών [56, 57, 58]. Κοινό χαρακτηριστικό όλων αποτελεί ότι όσα μοντέλα πέτυχαν υψηλότερη ακρίβεια για το δυαδικό πρόβλημα, εξαιρετική επίδοση παρουσίασαν και για το πρόβλημα πολλών κλάσεων.

Την προεκπαιδευμένη στο ImageNet αρχιτεκτονική ResNet152 [22] αξιοποίησαν οι Gandomkar et al. [54]. Συγκεκριμένα ανέπτυξαν ένα μοντέλο δύο σταδίων, όπου στο πρώτο προβλεπόταν το είδος της αλλοίωσης, καλοήθης ή κακοήθης και στο δεύτερο ανάλογα με το είδος προβλεπόταν και ο υπότυπος. Οι προβλέψεις σε επίπεδο ασθενούς προέκυπταν από ένα δέντρο μέτα-απόφασης (Meta-Decision Tree) [59] με εισόδους από τα μοντέλα όλων των μεγεθύνσεων. Εισόδους για το συνολικό μοντέλο αποτέλεσαν υποτιμήματα των αρχικών εικόνων μετά από την εφαρμογή διαφορετικών τεχνικών επαύξησης και κανονικοποίησης, με τη συνολική μέση ακρίβεια να κυμαίνεται για όλες τις μεγεθύνσεις περί του 95%.

Ακόμη, οι Xie et al. [55] μελέτησαν τις αρχιτεκτονικές μεταφοράς μάθησης InceptionV3 [31] και InceptionResNetV2 [23], με αρχικοποιημένα βάρη από την προεκπαίδευση στο ImageNet και με εκπαίδευση μόνο του τελευταίου επιπέδου. Σε επαυξημένα δεδομένα το InceptionResNetV2 παρουσίασε κορυφαία ακρίβεια πάνω από 99% για όλα τα επίπεδα μεγέθυνσης στο δυαδικό πρόβλημα και ακρίβεια περί του 97% για την πολυ-κατηγοριοποίηση. Παράλληλα, πειραματίστηκαν και με τεχνικές μη επιβλεπόμενης μάθησης κάνοντας ομαδοποίηση με τον αλγόριθμο K Μέσων (KMeans) πάνω

στο δiάνυσμα χαρακτηριστικών που εξήγαγε το μοντέλο InceptionResNetV2. Επίσης, εξετάστηκε η μείωση της διάστασης του εν λόγω διανύσματος με κωδικοποιητή, πριν αυτό αποτελέσει είσοδο του αλγορίθμου ομαδοποίησης, με την ακρίβεια να κυμαίνεται μεν γενικά σε χαμηλά επίπεδα αλλά να είναι σχετικά υψηλότερη στην περίπτωση εφαρμογής του κωδικοποιητή.

Αναφορικά με τις ειδικές αρχιτεκτονικές οι Jiang et al. [57] ανέπτυξαν ένα βαθύ συνελκτικό μοντέλο βασισμένο σε μια παραλλαγή με λιγότερες παραμέτρους της δομικής μονάδας SE-ResNet [60], η οποία συνδυάζει τα βασικά χαρακτηριστικά των ResNet [22] και Squeeze-and-Excitation [25] αρχιτεκτονικών. Επιπρόσθετα εισήγαγαν ένα νέο αλγόριθμο προσαρμογής του ρυθμού μάθησης κατά τη διαδικασία εκπαίδευσης, η χρήση του οποίου οδήγησε σε ελαφρά καλύτερα αποτελέσματα σε σχέση με κάποιες κλασικές μεθόδους προσαρμογής. Τα μοντέλα τους σε συνδυασμό με τον αλγόριθμο προσαρμογής που ανέπτυξαν απέδωσαν κορυφαία ακρίβεια περί του 99% για το δυαδικό πρόβλημα για όλα τα επίπεδα μεγέθυνσης και της τάξης του 94% για μεγεθύνσεις 40×, 100× κι ελαφρά χαμηλότερη για τις υπόλοιπες για το πρόβλημα κατηγοριοποίησης σε υποτύπους. Οι Alom et al. [56] ανέπτυξαν ένα βαθύ συνελκτικό μοντέλο βασισμένο στην αρχιτεκτονική IRRCNN (Inception Recurrent Residual Convolutional Neural Network) [61, 62], της οποίας η βασική δομική μονάδα αποτελεί συνδυασμό των κύριων δομικών στοιχείων των αρχιτεκτονικών τύπου Inception [31], τύπου ResNet [22] προσθέτοντας και βρόχους ανάδρασης σε ορισμένα συνελκτικά επίπεδα. Το μοντέλο τους και για τα δύο προβλήματα κατηγοριοποίησης εφαρμοσμένο σε επαυξημένα δεδομένα απέδωσε ακρίβεια τάξης 97% σε όλα τα επίπεδα μεγέθυνσης.

### 3.4 Κατηγοριοποίηση Πολλών Κλάσεων Ανεξαρτήτου Μεγέθυνσης

Δεδομένης της αυξημένης πολυπλοκότητας του προβλήματος, οι ερευνητικές μελέτες που παρουσιάζουν ικανοποιητικά αποτελέσματα είναι ακόμη περιορισμένες σε αριθμό. Κοινό σημείο τους αποτελεί η εφαρμογή τεχνικών μεταφοράς μάθησης κι επαύξησης δεδομένων.

Οι Murtaza et al. [63] ανέπτυξαν ένα ιεραρχικό μοντέλο αποτελούμενο από τρία επιμέρους μοντέλα, ένα υπεύθυνο για τη διάκριση του είδους της αλλοίωσης-καλοήθους ή κακοήθους-, ένα για τη διάκριση των καλοήθων υποτύπων κι ένα για τη διάκριση των κακοήθων, τα οποία επιλέγονταν να χρησιμοποιηθούν ανάλογα με την διάκριση στο πρώτο επίπεδο. Για την εξαγωγή χαρακτηριστικών χρησιμοποίησαν την προεκπαιδευμένη στο ImageNet αρχιτεκτονική AlexNet [19], συνεχίζοντας την εκπαίδευση μόνο στα ανώτερα επίπεδα, ενώ δοκιμάστηκαν κλασικοί ταξινομητές μηχανικής μάθησης, όπως SVM, kNN, DT κ.ά., όπως και επίσης και οι τεχνικές μείωσης διαστατικότητας PCA [64] και IG (Information Gain) [65] πάνω στο δiάνυσμα χαρακτηριστικών. Τη βέλτιστη ακρίβεια απέδωσε η εφαρμογή IG και ο kNN ταξινομητής σε επαυξημένα δεδομένα, με ακρίβεια 95.48% για το πρώτο επίπεδο διάκρισης που αντιστοιχεί στο δυαδικό πρόβλημα, 94.62% και 92.45% στο δεύτερο επίπεδο για τα μοντέλα διάκρισης καλοήθων και κακοήθων υποτύπων αντίστοιχα.

Οι Jannesari et al. [66] συνέκριναν τις επιδόσεις διαφορετικών μοντέλων μεταφοράς μάθησης, συγκεκριμένα προεκπαιδευμένων στο ImageNet αρχιτεκτονικών Inception [31] και ResNet [22], συνεχίζοντας να εκπαιδεύουν είτε το τελευταίο είτε όλα τα επίπεδά τους. Η τελευταία πρακτική απέδωσε για το ResNet152 ακρίβεια 98.7% για το δυαδικό πρόβλημα, 94.5% και 96.4% για την κατηγοριοποίηση καλοήθων και κακοήθων αλλοιώσεων αντίστοιχα. Παρόμοια μελέτη πραγματοποίησαν και οι Nawaz et al. [67], μελετώντας μόνο το πρόβλημα της κατηγοριοποίησης υποτύπων, με τη βέλτιστη ακρίβεια των πειραμάτων τους να είναι 95%, επιλέγοντας το μοντέλο ResNet50V1 [22], συνεχίζοντας

την εκπαίδευση όλων των επιπέδων πάνω σε επαυξημένα δεδομένα.



## Κεφάλαιο 4

### Σύνολο Δεδομένων

Το σύνολο δεδομένων BreakHis (Breast Cancer Histopathological Database) που μελετάται στην παρούσα εργασία δημιουργήθηκε από τους Spanhol et al. [3] το 2015 σε συνεργασία με το εργαστήριο πρόγνωσης και διάγνωσης Βραζιλίας (P&D Laboratory-Brazil) και διατίθεται δημοσίως μέσω του [παρόντος](#) ιστοτόπου. Δημιουργήθηκε προκειμένου να αποτελέσει ένα κοινό πλαίσιο αναφοράς ως προς την αξιολόγηση και ανάπτυξη εφαρμογών υποβοήθησης ιατρικής διάγνωσης και επεξεργασίας ιατρικής εικόνας. Δεδομένης της πολυπλοκότητας και των υψηλών απαιτήσεων που συνεπάγεται για τη δημιουργία μίας μεγάλης βάσης ιατρικών δεδομένων, ειδικά όσον αφορά τον τομέα της ιστοπαθολογίας, το εγχείρημα της δημιουργίας και δημόσιας διάθεσης ενός τέτοιου συνόλου δεδομένων συνιστά μία πολύ σημαντική προσφορά στην επιστημονική κοινότητα, καθώς παρέχεται η δυνατότητα πρόσβασης σε ένα εκτενές σύνολο δεδομένων σε ερευνητικές ομάδες που διαφορετικά δε θα μπορούσαν να διαθέτουν. Ακολουθεί η συνοπτική παρουσίαση του συνόλου δεδομένων, η ανάλυση των τεχνικών αποδοτικής αποθήκευσης καθώς και οι τεχνικές προεπεξεργασίας που ακολουθήθηκαν.

#### 4.1 Περιγραφή

Το παρόν σύνολο δεδομένων αποτελείται από 7909 εικόνες μικροσκοπίας από βιοψίες καλοήθων και κακοήθων όγκων μαστού. Τα δείγματα λήφθηκαν μέσω χειρουργικής ανοιχτής βιοψίας (Surgical open biopsy - SOB) από 82 ασθενείς, προετοιμάστηκαν με την τυπική διαδικασία για ιστολογική εξέταση σε μικροσκόπιο και υπέστησαν χρώση αιματοξυλίνης-ηωσίνης. Οι τελικές εικόνες λήφθηκαν με κάμερα προσαρμοσμένη στο φακό του μικροσκοπίου με μέγεθος εικονοστοιχείου  $6.5\mu\text{m} \times 6.25\mu\text{m}$ , είχαν συνολικό μέγεθος μετά από αφαίρεση περιθωρίων  $700 \times 460$  εικονοστοιχεία και τρία κανάλια χρώματος (κόκκινο, πράσινο, μπλε-RGB), με 8 bits το καθένα. Αποθηκεύτηκαν σε μορφή PNG με μηδενική συμπίεση ενώ σημειώνεται ότι δεν εφαρμόστηκε καμία τεχνική κανονικοποίησης χρώματος. Για κάθε περίπτωση, οι ετικέτες των δειγμάτων προήλθαν από τη γνωμάτευση πεπειραμένων παθολόγων, η οποία επιβεβαιώθηκε συμπληρωματικά και από ανοσοϊστοχημικές εξετάσεις.

Για κάθε ασθενή λήφθηκαν κατά μέσο όρο 24 εικόνες αρχικά, σε μεγέθυνση  $40 \times$  και στη συνέχεια άλλες τόσες περίπου για μεγεθύνσεις  $100 \times$ ,  $200 \times$  και  $400 \times$  εστιάζοντας όλο και περισσότερο στην περιοχή ενδιαφέροντος, όπως αυτή κρίθηκε από τον παθολόγο. Έτσι, το τελικό σύνολο δεδομένων περιέχει εικόνες από 4 διαφορετικά επίπεδα μεγέθυνσης, με 2480 να χαρακτηρίζονται ως καλοήθειες και 5429 ως κακοήθειες. Πέραν από αυτή τη διάκριση, για κάθε δείγμα διακρίθηκε και ο ειδικός τύπος της αλλοίωσης, οποίος ήταν ένας από τους κάτωθι: αδένωση (Adenosis-A), ινοαδένωμα (Fibroadenoma-F), φυλλοειδής όγκος (Phyllodes Tumor-PT), σωληνώδες αδένωμα (Tubular Adenoma-TA), πορογενές καρκίνωμα (Ductal Carcinoma-DC), λοβιακό καρκίνωμα (Lobular Carcinoma-LC), βλενωδές

καρκίνωμα (Mucinous Carcinoma-MC) και θυλώδες καρκίνωμα (Papillary Carcinoma-PC). Από αυτούς οι τέσσερις πρώτοι ανήκουν στους καλοήθεις όγκους, ενώ οι τέσσερις τελευταίοι στους κακοήθεις. Η κατανομή των εικόνων για τις διάφορες κατηγορίες, για όλα τα επίπεδα μεγέθυνσης δίνεται από τους παρακάτω πίνακες.

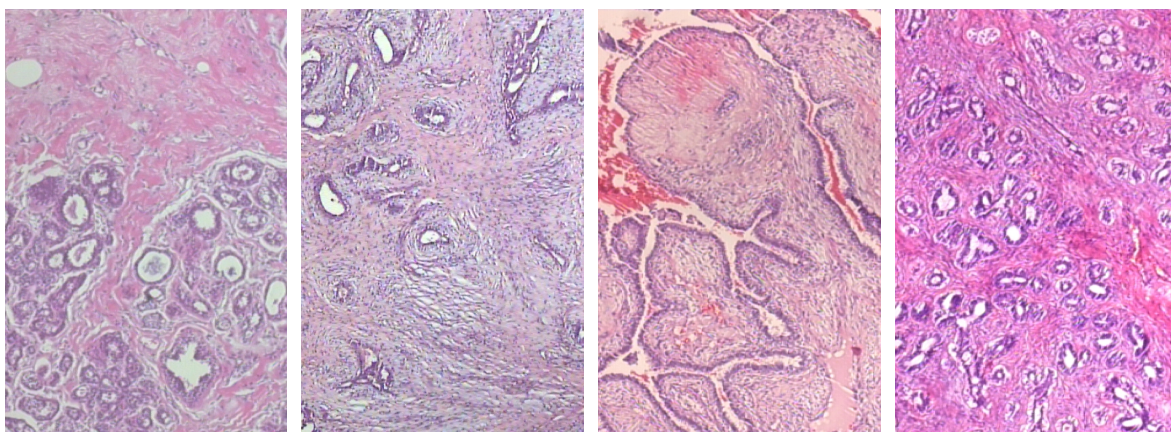
Μεγέθυνση	A	F	TA	PT	Σύνολο
40×	114	253	109	149	625
100×	113	260	121	150	644
200×	111	264	108	140	623
400×	106	237	115	130	588
Σύνολο	444	1014	453	569	2480
# Ασθενών	4	10	3	7	24

Πίνακας 4.1: Κατανομή καλοήθων υπότυπων

Μεγέθυνση	DC	LC	MC	PC	Σύνολο
40×	864	156	205	145	1370
100×	903	170	222	142	1437
200×	896	163	196	135	1390
400×	788	137	169	138	1232
Σύνολο	3451	626	792	560	5429
# Ασθενών	38	5	9	6	58

Πίνακας 4.2: Κατανομή κακοήθων υπότυπων

Από την κατανομή είναι εμφανής η έντονη ανισοκατανομή των δεδομένων ως προς τα είδη και τους υπότυπους των αλλοιώσεων. Οι κακοήθεις αλλοιώσεις είναι υπερδιπλάσιες των καλοήθων στο σύνολό τους, ενώ το πλήθος των δεδομένων που ανήκουν στις κλάσεις F και DC είναι τάξεις μεγέθους μεγαλύτερο από τα υπόλοιπα. Το γεγονός αυτό, καθιστά τόσο το πρόβλημα δυαδικής κατηγοριοποίησης, όσο και την κατηγοριοποίηση υπότυπων πιο δύσκολο λόγω της ανισοκατανομής των δεδομένων, αφού μεγάλος είναι ο κίνδυνος οι προβλέψεις οποιουδήποτε συστήματος μηχανικής μάθησης να είναι βεβιασμένες. Πέραν της ανισοκατανομής, η εξαιρετικά μεγάλη ποικιλομορφία των δεδομένων ακόμα και για δείγματα εντός της ίδιας κατηγορίας, αποτελεί ένα ακόμα στοιχείο που αυξάνει την πολυπλοκότητα του προβλήματος. Ενδεικτικά στα σχήματα 4.1 και 4.2 παρατίθεται ένα δείγμα για κάθε καλοήθη και κακοήθη κλάση αντίστοιχα.



(1) Αδένωση

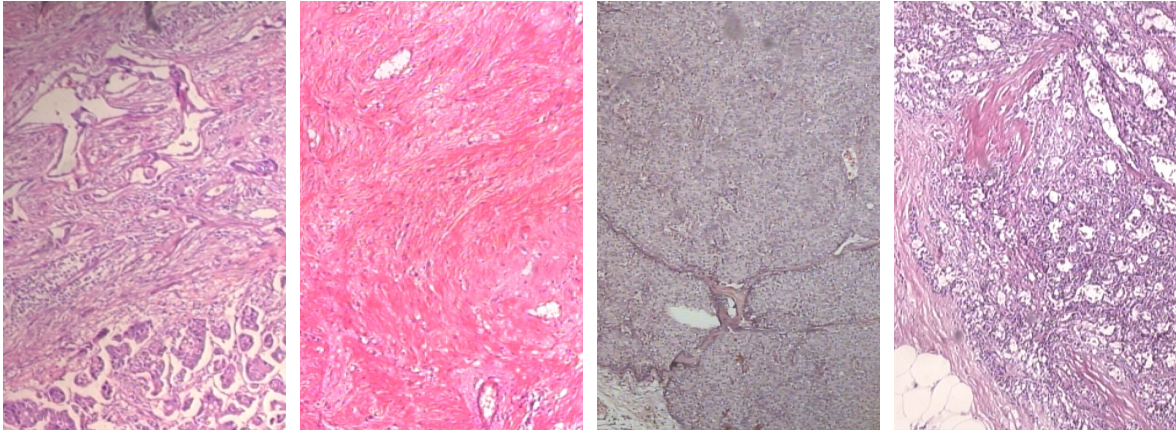
(2) Ινοαδένωμα

(3) Φυλλοειδής όγκος

(4) Σωληνώδες αδένωμα

Σχήμα 4.1: Καλοήθη δείγματα σε μεγέθυνση 40×





(1) Πορογενές καρκίνωμα (2) Λοβιακό καρκίνωμα (3) Βλεννώδες καρκίνωμα (4) Θυλώδες καρκίνωμα

Σχήμα 4.2: Κακοήθη δείγματα σε μεγέθυνση 40×

## 4.2 Αποθήκευση

Το σχετικά μεγάλο πλήθος των δεδομένων του υπό εξέταση συνόλου έχει ως συνέπεια η οποιαδήποτε επεξεργασία του, άρα και η χρήση του κατά τη διαδικασία εκπαίδευσης κι ελέγχου, να παρουσιάζει αυξημένες απαιτήσεις σε μνήμη. Μνήμη όμως είναι αναγκαία και για τις παραμέτρους του εκάστοτε μοντέλου που εξετάζεται, οι οποίες ανάλογα με το μοντέλο μπορεί να είναι ακόμα και μερικά εκατομμύρια. Σε κάθε περίπτωση, η τακτική ανάγνωσης των δεδομένων από την αρχική τους μορφή και η τοποθέτηση όλων τους στην κύρια μνήμη, προσθέτει πέραν από χρονική επιβάρυνση, αυξημένο φορτίο σε επίπεδο μνήμης, σπαταλώντας πολύτιμους πόρους.

Για την αντιμετώπιση των παραπάνω, συνιστάται η αποθήκευση των δεδομένων ως μία σειρά από δυαδικές εγγραφές, καθιστώντας δυνατή την ταχεία και τμηματική τους ανάγνωση και τοποθέτηση στη μνήμη, μειώνοντας στο ελάχιστο την απαίτηση πόρων κύριας μνήμης. Ο τύπος αρχείου που είναι πλέον διαδεδομένος για τη διαδικασία αυτή και υποστηρίζεται από τις κύριες βιβλιοθήκες μηχανικής μάθησης είναι ο τύπος TFRecord. Για τη μετατροπή των δεδομένων σε αρχεία τέτοιου τύπου απαιτείται αρχικά η μετατροπή κάθε δείγματος σε μορφή `tf.train.Example` και στη συνέχεια η σειριοποίησή τους και αποθήκευσή τους σε TFRecord αρχεία. Η μορφή `tf.train.Example` αποτελεί ένα είδος ενδιάμεσης αναπαράστασης, κατά την οποία κάθε δείγμα μπορεί να αναπαρασταθεί ως ένα σύνολο πεδίων, με τύπο δομής προκαθορισμένο από τον προγραμματιστή, με τελική μορφή προσφιλή για σειριοποίηση. Ο ενδιάμεσος αυτός τρόπος αναπαράστασης εισήχθη από την Google, με σκοπό να αποτελέσει έναν αποδοτικό κι επεκτάσιμο μηχανισμό σειριοποίησης δομημένων δεδομένων, ανεξαρτήτου περιβάλλοντος ανάπτυξης και γλώσσας προγραμματισμού, γνωστό και ως [protocol buffer](#).

Η πρακτική αυτή ακολουθήθηκε και στην παρούσα εργασία, με το σύνολο των δεδομένων να μετατρέπεται σε ένα σύνολο αρχείων TFRecord. Για κάθε δείγμα στην ενδιάμεση αναπαράσταση `tf.train.Example` αποθηκεύτηκε πέραν από την εικόνα και την ετικέτα της κατηγορίας της, το επίπεδο μεγέθυνσης και ο κωδικός του ασθενούς στον οποίο αντιστοιχεί (προφανώς το σύνολο δεδομένων παρέχεται ανωνυμοποιημένο, με τους ασθενείς να διακρίνονται μέσω αλφαριθμητικών αναγνωριστικών). Λόγω του σχετικά μεγάλου μεγέθους του συνόλου δεδομένων (περί τα 4GB) και προκειμένου να γίνει βέλτιστη εκμετάλλευση των λειτουργιών παραλληλοποίησης και προφόρτωσης κατά

την ανάγνωση και φόρτωση, επιλέχθηκε η δημιουργία πολλών TFRecord αρχείων, μεγέθους περίπου 150MB το καθένα.

## 4.3 Προεπεξεργασία

### 4.3.1 Βασική Προεπεξεργασία

Τα δεδομένα προτού αποτελέσουν είσοδο για τα μοντέλα που μελετήθηκαν στα πειράματα υπέστησαν ορισμένη επεξεργασία προκειμένου να μετατραπούν στον κατάλληλο τύπο και να λάβουν τις απαραίτητες διαστάσεις ανάλογα με την εκάστοτε αρχιτεκτονική. Συγκεκριμένα, αρχικά από τη σειριοποιημένη δυαδική μορφή κωδικοποιήθηκαν ως διανύσματα εικόνας τριών καναλιών - ένα για κάθε χρώμα (κόκκινο, πράσινο, κυανό) -, με τις τιμές κάθε στοιχείου του δισδιάστατου πλέγματος κάθε καναλιού να αποτελούν την ένταση του χρώματος αυτού, στο αντίστοιχο εικονοστοιχείο, λαμβάνοντας ακέραιες τιμές στο διάστημα  $[0,255]$ . Στη συνέχεια, οι τιμές αυτές μετατράπηκαν σε πραγματικές, προκειμένου να επιτρέπονται οι μετέπειτα υπολογισμοί, είτε για περαιτέρω επεξεργασία, είτε για εκπαίδευση, ενώ σε επόμενο στάδιο οι διαστάσεις μήκους και πλάτους αναπροσαρμόστηκαν, ανάλογα με τις απαιτήσεις εισόδου κάθε μοντέλου. Για τη προσαρμογή των διαστάσεων χρησιμοποιήθηκε ο αλγόριθμος γεωμετρικού μετασχηματισμού διγραμμικής παρεμβολής. Επίσης, δεδομένης της έντονης ποικιλομορφίας των εικόνων, η οποία εκφράζεται από μεγάλες διακυμάνσεις στις τιμές κάθε χρώματος, προτιμήθηκε κανονικοποίηση με βάση τη μέση τιμή και διασπορά κάθε διανύσματος, έτσι ώστε όλα τα δείγματα να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση, αντί απλής κανονικοποίησης στο  $[0,1]$ .

Στην παραπάνω επεξεργασία, υποβλήθηκαν τόσο τα σύνολα εκπαίδευσης κι επαλήθευσης, όσο και το σύνολο ελέγχου. Τα μη επικαλυπτόμενα αυτά σύνολα δεν προέκυψαν με τυχαία δειγματοληψία, αλλά με βάση το πρωτόκολλο της ποσοστιαίας διάκρισης με βάση κάθε ασθενή, όπως ακολουθήθηκε πρώτα από τους Spanhol et al. [3] και στη συνέχεια και από άλλους, έτσι ώστε να μην περιέχονται εικόνες του ίδιου ασθενούς και στο σύνολο ελέγχου και στο σύνολο εκπαίδευσης, (αντίστοιχα και στα σύνολα επαλήθευσης κι εκπαίδευσης). Στο σύνολο ελέγχου (κι επαλήθευσης), πρέπει να περιλαμβάνονται εντελώς άγνωστα δείγματα, προκειμένου να μετρηθεί αντικειμενικά η αξιοπιστία και η απόδοση του εκάστοτε μοντέλου. Δείγματα προερχόμενα από την ίδια πηγή και άρα εικόνες του ίδιου ασθενούς, είναι αναμενόμενο να παρουσιάζουν έντονες ομοιότητες και για το λόγο αυτό η ύπαρξή τους και στα δύο σύνολα, οδηγεί στην αξιολόγηση ενός ευκολότερου προβλήματος, αφού το σύστημα έτσι καλείται να αναγνωρίσει δείγματα πανομοιότυπα με αυτά που εκπαιδεύτηκε. Δεν είναι τυχαίο άλλωστε ότι αρκετές από τις κορυφαίες επιδόσεις μοντέλων που παρατηρήθηκαν στη βιβλιογραφία, αφορούσαν τυχαίο διαμοιρασμό δεδομένων, μελετώντας ουσιαστικά ένα ευκολότερο πρόβλημα. Επειδή, όμως πρόκειται για σύστημα υποβοήθησης ιατρικής διάγνωσης απαιτούνται οι αυστηρότεροι περιορισμοί απόδοσης και αξιοπιστίας, ακολουθήθηκε η ανά ασθενή διάκριση, με τα δείγματα του 80% των ασθενών να χρησιμοποιούνται για εκπαίδευση και για επαλήθευση (αναλογία 9:1) και το υπόλοιπο 20% σε συνδυασμό με το σύνολο επαλήθευσης για έλεγχο.

### 4.3.2 Επαύξηση Δεδομένων

Δεδομένου ότι για το πρόβλημα της κατηγοριοποίησης εικόνων με επιβλεπόμενη μάθηση, απαιτείται μεγάλο πλήθος δεδομένων εκπαίδευσης, προκειμένου να προκύψει μια εύλογη ικανότητα γε-

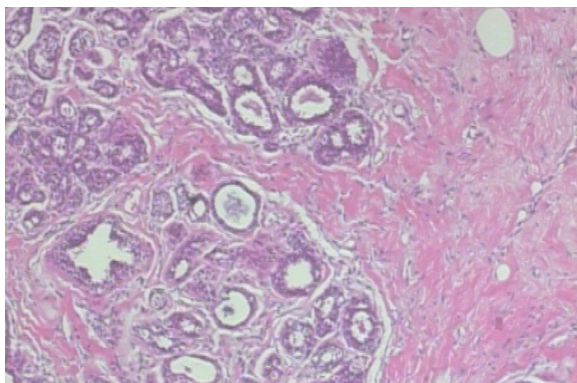


νίκευσης από το εκάστοτε μοντέλο, η επαύξηση του συνόλου εκπαίδευσης αποτελεί μια συνήθη τεχνική για σχετικά μικρά σύνολα δεδομένων. Με τον όρο επαύξηση εννοείται η εφαρμογή διαφόρων μετασχηματισμών σε κάθε δείγμα, έτσι ώστε να προκύπτει μία ελαφρά διαφοροποιημένη εκδοχή της αρχικής εικόνας για κάθε μετασχηματισμό και η ενσωμάτωσή της στο σύνολο εκπαίδευσης. Οι μετασχηματισμοί αυτοί μπορεί να εφαρμόζονται είτε με κάποια τυχαιότητα για κάθε δείγμα, είτε για όλα τα δείγματα. Πέραν από την αύξηση του πλήθους των δεδομένων εκπαίδευσης στόχος της επαύξησης αποτελεί και ο εμπλουτισμός του αρχικού συνόλου με διαφοροποιημένα δεδομένα για να ενισχυθεί περαιτέρω η ικανότητα γενίκευσης του τελικού συστήματος.

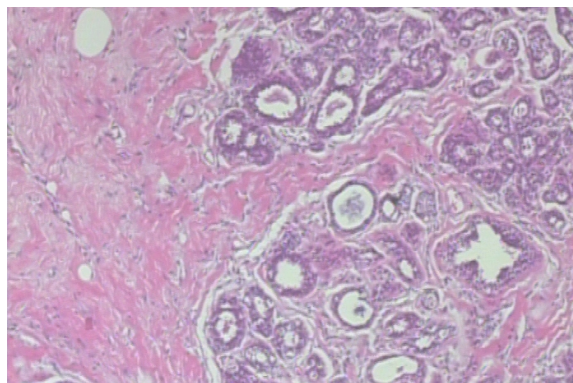
Με δεδομένα τα παραπάνω και λαμβάνοντας υπόψη ότι στην πλειοψηφία των περιπτώσεων των πειραμάτων της βιβλιογραφίας οι υψηλότερες επιδόσεις προέκυψαν από εκπαίδευση σε επαυξημένα δεδομένα, τεχνικές επαύξησης εφαρμόστηκαν και σε μέρος των πειραμάτων της παρούσας μελέτης. Για τις κατηγορίες προβλημάτων με δεδομένα κοινής μεγέθυνσης -όπως αυτές διακρίνονται στη βιβλιογραφική ανασκόπηση-, οι ίδιες πειραματικές διατάξεις εφαρμόστηκαν σε επαυξημένα και μη δεδομένα εκπαίδευσης, προκειμένου να μελετηθεί η επίδρασή τους στη απόδοση των μοντέλων. Ειδικότερα, για τις κατηγορίες αυτές, το σύνολο εκπαίδευσης επαυξήθηκε πέντε φορές με τους ακόλουθους μετασχηματισμούς:

- Αναστροφή στον οριζόντιο άξονα
- Αναστροφή στον κάθετο άξονα
- Περιστροφή  $90^\circ$  προς τα αριστερά
- Περιστροφή  $180^\circ$  προς τα αριστερά
- Περιστροφή  $270^\circ$  προς τα αριστερά

Για τις κατηγορίες πειραμάτων με δεδομένα από όλες τις μεγεθύνσεις λόγω περιορισμού πόρων μνήμης, αλλά και υπολογιστικής ισχύος, δεν εφαρμόστηκαν τεχνικές επαύξησης, μια και το μέγεθος του αρχικού συνόλου εκπαίδευσης προέκυπτε περίπου τριπλάσιο σε σχέση με την κατηγορία πειραμάτων με δεδομένα κοινής μεγέθυνσης. Σημειώνεται ότι η προσαρμογή των διαστάσεων και η κανονικοποίηση των τιμών των διανυσμάτων πραγματοποιήθηκε μετά από την επαύξηση των δεδομένων, καθώς για την ελαχιστοποίηση απώλειας πληροφορίας των αρχικών δεδομένων λόγω των μετασχηματισμών είθισται οι μετασχηματισμοί αυτοί να εφαρμόζονται ως τελευταία βήματα της προεπεξεργασίας.



(1) Αρχικό δείγμα



(2) Αναστραμμένο δείγμα στον κάθετο άξονα



## Κεφάλαιο 5

### Πειραματικές Διατάξεις

Αντικείμενο μελέτης των πειραμάτων αποτέλεσε η υλοποίηση και αξιολόγηση διαφορετικών μοντέλων επιβλεπόμενης μάθησης, για της διαφορετικές εκφάνσεις του προβλήματος της κατηγοριοποίησης ιστοπαθολογικών εικόνων καρκίνου του μαστού, όπως αυτές προκύπτουν από το εξεταζόμενο σύνολο δεδομένων BreakHis[3]. Συγκεκριμένα, υλοποιήθηκαν τρία βαθιά συνελκτικά μοντέλα, καθένα από τα οποία είχε ως κύριο συστατικό κάποια προεκπαιδευμένη στο ImageNet αρχιτεκτονική. Το πρώτο εξ αυτών, το οποίο αναπτύχθηκε αποκλειστικά για τους σκοπούς της παρούσας εργασίας, στηρίχθηκε στην αρχιτεκτονική VGG16 [20], το δεύτερο αποτελεί υλοποίηση του μοντέλου των Nawaz et al. [52] και στηρίζεται στην προεκπαιδευμένη αρχιτεκτονική DenseNet121 [24] ενώ το τρίτο συνδυάζει την αρχιτεκτονική αυτή με την αρχιτεκτονική SENet [25], σύμφωνα με την εργασία των Li et al. [45], οι οποίοι του έδωσαν την ονομασία IDNet. Καθένα από τα παραπάνω μοντέλα εκπαιδεύτηκε και αξιολογήθηκε για καθένα από τα τέσσερα είδη υποπροβλημάτων, δηλαδή:

- Δυαδική κατηγοριοποίηση σε δεδομένα κοινής μεγέθυνσης
- Δυαδική κατηγοριοποίηση σε δεδομένα ανεξαρτήτου μεγέθυνσης
- Κατηγοριοποίηση πολλών κλάσεων σε δεδομένα κοινής μεγέθυνσης
- Κατηγοριοποίηση πολλών κλάσεων σε δεδομένα ανεξαρτήτου μεγέθυνσης

Στην πρώτη και τρίτη περίπτωση, υπενθυμίζεται ότι τα μοντέλα εκπαιδεύτηκαν κι αξιολογήθηκαν με επαυξημένα και μη δεδομένα. Σε όλες τις περιπτώσεις τα δεδομένα μετά τη βασική προεπεξεργασία (και την επαύξησή όπου αυτή εφαρμόστηκε) αναδιατάχθηκαν συνολικά με τυχαίο τρόπο και οργανώθηκαν σε δέσμες (batches), τα δεδομένα των οποίων σε κάθε επανάληψη επίσης αναδιατάσσονταν με τυχαίο τρόπο. Η αξία της τυχαία αναδιάταξης έγκειται στο γεγονός ότι σε κάθε επανάληψη η κάθε δέσμη θα έχει μία μοναδική διάταξη, οπότε και κάθε μεταβολή που θα προκύπτει στην εκπαίδευση θα είναι μοναδική και ανεξάρτητη από τις προηγούμενες. Η ιδιότητα αυτή, της ανεξαρτησίας των μεταβολών, προκύπτουσα από την τυχειότητα της διάταξης των δεδομένων, συμβάλλει στην ενίσχυση της ικανότητας γενίκευσης και στην αποφυγή μεροληπτικών προβλέψεων, στις οποίες είναι πολύ εύκολο να οδηγηθεί ένα μοντέλο ειδικά για ένα έντονα ανισοκαταμεμημένο σύνολο δεδομένων, όπως αυτό που εξετάζεται. Ακόμη, για λόγους υψηλότερης ταχύτητας υπολογισμών, αξιοποιήθηκαν οι λειτουργίες προ-αποθήκευσης (caching) και προ-φόρτωσης (prefetching), που παρείχε η βιβλιοθήκη που χρησιμοποιήθηκε, για την οποία γίνεται λόγος παρακάτω.

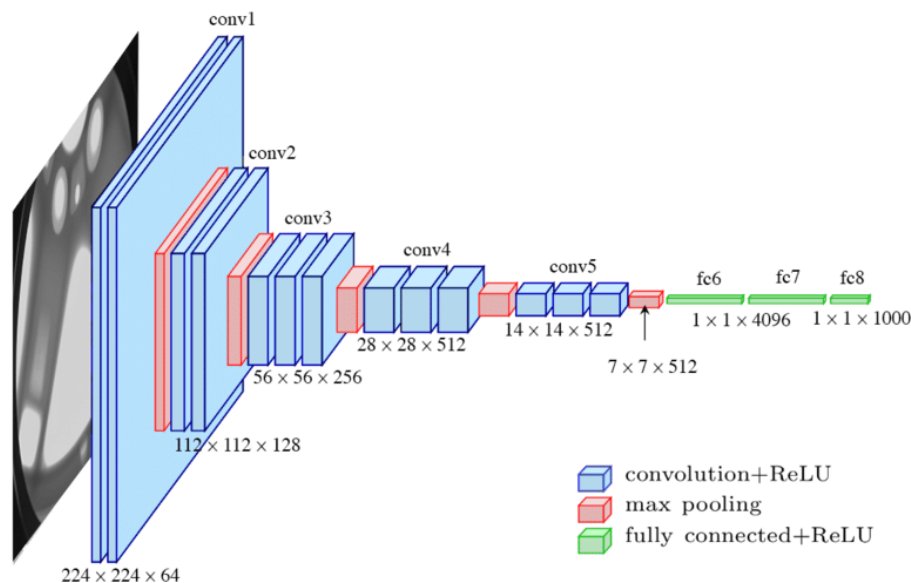
Για όλα τα προβλήματα και για όλα τα μοντέλα ως συνάρτηση σφάλματος επιλέχθηκε η συνάρτηση διασταυρούμενης εντροπίας, ενώ ως μετρικές αξιολόγησης επίσης για όλα μοντέλα, τέθηκαν η ακρίβεια σε επίπεδο εικόνας και σε επίπεδο ασθενούς, η μετρική Precision, η ανάκληση καθώς και

το F1 score. Ως περιβάλλον υλοποίησης κι εκτέλεσης των πειραμάτων, επιλέχθηκε η διαδικτυακή πλατφόρμα νέφους [Google Colab](#), ως γλώσσα υλοποίησης η Python 3.6 και ειδικά οι βιβλιοθήκες Tensorflow και Keras, με το περιβάλλον υλοποίησης κι εκτέλεσης να παρέχει μονάδα επεξεργασίας γραφικών (GPU) 12GB τύπου NVIDIA Tesla K80. Παρακάτω παρουσιάζεται αναλυτικά κάθε αρχιτεκτονική καθώς και οι αλγόριθμοι μάθησης και τα πρωτόκολλα εκπαίδευσης που χρησιμοποιήθηκαν σε κάθε περίπτωση.

## 5.1 VGG16

### 5.1.1 Αρχιτεκτονική

Πρόκειται για μια βαθιά συνελκτική αρχιτεκτονική, όπως φαίνεται και από το σχήμα 5.1, ο βασικός κορμός της οποίας αποτελείται από πέντε σύνθετα στρώματα δύο ή τριών συνεχόμενων συνελκτικών επιπέδων, μεταξύ των οποίων μεσολαβούν επίπεδα δειγματοληψίας μεγίστου (max pooling), ενώ ειδικά η αρχιτεκτονική που εκπαιδεύτηκε στο ImageNet περιέχει επιπρόσθετα την "κεφαλή" κατηγοριοποίησης (classification head) αποτελούμενη από τρία πλήρως συνδεδεμένα επίπεδα. Τα μεγέθη πυρήνα όλων των συνελκτικών επιπέδων είναι  $3 \times 3$ , το πλήθος των φίλτρων που εφαρμόζονται για όλα τα συνελκτικά επίπεδα εντός του ίδιου στρώματος είναι 64, 128, 256, 512 και 512 αντίστοιχα, ενώ το μέγεθος του παραθύρου δειγματοληψίας είναι  $2 \times 2$ , με αποτέλεσμα κάθε διάσταση να μειώνεται στο μισό, όπως φαίνεται και στα μεγέθη εξόδου του κάθε στρώματος σχηματικά.



Σχήμα 5.1: Η αρχιτεκτονική VGG16

Για την ανάπτυξη του μοντέλου με το οποίο πραγματοποιήθηκαν τα πειράματα, χρησιμοποιήθηκε ο βασικός κορμός της παραπάνω αρχιτεκτονικής με αρχικοποιημένα βάρη στο ImageNet κι ερευνήθηκαν σε πρώτη φάση διάφορες παραλλαγές της κεφαλής κατηγοριοποίησης με προσθήκη πλήρως συνδεδεμένων επιπέδων διαφορετικών διαστάσεων, επιπέδων απόρριψης (dropout), καθολικής δειγματοληψίας (global pooling) και κανονικοποίησης δέσμης (batch normalization). Η προσθήκη των τελευταίων έγινε με στόχο τον περιορισμό του προβλήματος της υπερπροσαρμογής (overfitting), το

οποίο ήταν αρκετά έντονο λόγω του σχετικά μεγάλου πλήθους παραμέτρων του βασικού κορμού της αρχιτεκτονικής VGG16. Παράλληλα, για τον ίδιο λόγο, προκειμένου να μειωθεί το πλήθος των εκπαιδευόμενων παραμέτρων, μελετήθηκε και το ποσοστό των κατώτερων επιπέδων των οποίων τα βάρη παρέμειναν σταθερά, στις αρχικοποιημένες τιμές τους, όπως αυτές προέκυψαν από την εκπαίδευση στο ImageNet, με τα επίπεδα αυτά έτσι να μη συμμετέχουν στην εκπαίδευση. Η επιλογή μη εκπαίδευσης ποσοστού των κατώτερων επιπέδων βασίστηκε στο γεγονός ότι τα κατώτερα επίπεδα, είναι εκπαιδευμένα στην αναγνώριση γενικότερων χαρακτηριστικών χαμηλού επιπέδου, όπως π.χ. ακμές, τα οποία συναντώνται στις περισσότερες εικόνες, οπότε δεν είναι απαραίτητη περαιτέρω εξειδίκευση σε αυτό το στάδιο. Αντίθετα, στα ανώτερα επίπεδα, τα οποία ειδικεύονται στην εξαγωγή ειδικότερων χαρακτηριστικών, απαιτείται περαιτέρω εκπαίδευση και ειδικά για ένα σύνολο δεδομένων τόσο διαφορετικό από το ImageNet.

Μετά από στοχευμένες πειραματικές δοκιμές, το τελικό μοντέλο που επιλέχθηκε με βάση την καλύτερη απόδοση, περιείχε το βασικό κορμό της αρχιτεκτονικής VGG16 με το 50% των επιπέδων να μην εκπαιδεύονται περαιτέρω, και classification head που περιείχε τα ακόλουθα επίπεδα με τη σειρά που αναφέρονται: global average pooling, dropout με ποσοστό απόρριψης 0.5, batch-normalization, πλήρως συνδεδεμένο επίπεδο 256 νευρώνων με συνάρτηση ενεργοποίησης ReLU, batch-normalization, dropout με ποσοστό απόρριψης 0.5, πλήρως συνδεδεμένο επίπεδο 128 νευρώνων με συνάρτηση ενεργοποίησης ReLU και τέλος πλήρως συνδεδεμένο επίπεδο 2 ή 8 νευρώνων, ανάλογα με το πρόβλημα, με συνάρτηση ενεργοποίησης Softmax.

### 5.1.2 Εκπαίδευση

Για όλες τις κατηγορίες προβλημάτων, χρησιμοποιήθηκε μέγεθος δέσμης 32, το μέγεθος εισόδου για κάθε δείγμα τέθηκε ίσο με  $224 \times 224$  όσον αφορά το μήκος και το πλάτος, ενώ ως αλγόριθμος βελτιστοποίησης επιλέχθηκε ο RMSProp με αρχικό ρυθμό μάθησης 0.00001. Ο τελευταίος επιλέχθηκε μετά από δοκιμές διαφόρων, με βάση την καλύτερη απόδοση. Για το δυαδικό πρόβλημα κατηγοριοποίησης κοινής μεγέθυνσης το μοντέλο εκπαιδεύτηκε για 20 εποχές, για το δυαδικό πρόβλημα κατηγοριοποίησης ανεξαρτήτου μεγέθυνσης για 30, ενώ για τα δύο προβλήματα πολυ-κατηγοριοποίησης για 60, με μη επαυξημένα δεδομένα, και για το πρόβλημα πολυ-κατηγοριοποίησης σε δεδομένα κοινής μεγέθυνσης με επαυξημένα δεδομένα 25 εποχές.

## 5.2 DenseNet121

### 5.2.1 Αρχιτεκτονική

Άλλη μία βαθιά συνελκτική αρχιτεκτονική, με την ιδιαιτερότητα των δομικών μονάδων με "πυκνές" (dense) συνδέσεις, όπου η είσοδος σε κάθε επίπεδο εντός της μονάδας προκύπτει από τη συνένωση των διανυσμάτων εξόδου όλων των προηγούμενων επιπέδων της μονάδας, όπως έχει ήδη αναφερθεί στη θεωρία. Η συγκεκριμένη εκδοχή DenseNet περιλαμβάνει τέσσερις τέτοιες δομικές μονάδες (dense blocks), τρία μεταβατικά επίπεδα (transition layers), καθώς και ένα επιπλέον συνελκτικό επίπεδο στην είσοδο ακολουθούμενο από ένα επίπεδο pooling. Τα παραπάνω αποτελούν το βασικό κορμό του μοντέλου, ενώ ειδικά για την προεκπαιδευμένη στο ImageNet αρχιτεκτονική περιλαμβάνεται κι ένα τελευταίο σύνθετο επίπεδο κατηγοριοποίησης, αποτελούμενο από ένα pooling επίπεδο κι ένα πλήρως συνδεδεμένο επίπεδο.

Αναφορικά με τις μονάδες dense, κύριο συστατικό αποτελεί η ακολουθία δύο συνελκτικών επιπέδων με μεγέθη πυρήνα  $1 \times 1$  και  $3 \times 3$  αντίστοιχα το οποίο επαναλαμβάνεται για ορισμένο πλήθος φορών, συγκεκριμένα 6,12,24 και 16 φορές για το πρώτο, δεύτερο, τρίτο και τέταρτο dense block αντίστοιχα. Σημειώνεται ότι η είσοδος κάθε συνελκτικού επιπέδου διέρχεται από ένα επίπεδο κανονικοποίησης δέσμης και από τη συνάρτηση ενεργοποίησης ReLU. Το πλήθος των χαρτών ενεργοποίησης σε κάθε επίπεδο καθορίζεται από έναν παράγοντα  $k$  (στη βιβλιογραφία [24] αναφερόμενο ως growth factor) και το πλήθος των χαρτών ενεργοποίησης της εισόδου του επιπέδου. Τα επίπεδα μετάβασης αποτελούνται από ένα συνελκτικό επίπεδο πυρήνα  $1 \times 1$  κι ένα επίπεδο pooling μέσης τιμής (average pooling) με μέγεθος φίλτρου  $2 \times 2$ . Το συνελκτικό επίπεδο της εισόδου, όπως και το επίπεδο pooling μέσης τιμής του επιπέδου κατηγοριοποίησης έχουν μέγεθος φίλτρου  $7 \times 7$ .

Η αρχιτεκτονική που δοκιμάστηκε διατηρήθηκε αυτούσια, όπως παρέχεται προεκπαιδευμένη από τη βιβλιοθήκη του προγραμματιστικού περιβάλλοντος, με μόνη διαφορά ότι το τελευταίο πλήρες επίπεδο ανάλογα με το πρόβλημα περιέλαβε 2 ή 8 νευρώνες αντίστοιχα και συνάρτηση ενεργοποίησης Softmax, με τυχαία αρχικοποιημένα βάρη. Οι παράμετροι όλων των επιπέδων πλην του τελευταίου αρχικοποιήθηκαν με τις τιμές από την εκπαίδευσης στο ImageNet, ενώ τα μόνα επίπεδα που εκπαιδεύτηκαν ήταν μόνο το πρώτο ή/και το τελευταίο ανάλογα με τα προβλήματα στα οποία εφαρμόστηκαν, όπως αναλύεται στη συνέχεια.

## 5.2.2 Εκπαίδευση

Για τα προβλήματα δυαδικής κατηγοριοποίησης, δεδομένης της μικρότερης πολυπλοκότητας και προκειμένου να διερευνηθεί η απόδοση του μοντέλου ως εξαγωγέα χαρακτηριστικών με μεταφορά μάθησης χωρίς περαιτέρω εκπαίδευση, εκπαιδεύτηκε μόνο το τελευταίο πλήρες επίπεδο για 40 εποχές. Για τα προβλήματα πολυ-κατηγοριοποίησης ακολουθήθηκε το πρωτόκολλο εκπαίδευσης δύο σταδίων της βιβλιογραφίας [52], με βάση το οποίο κατά το πρώτο στάδιο εκπαιδεύεται μόνο το τελευταίο επίπεδο και το στο δεύτερο συνεχίζει να εκπαιδεύεται και το πρώτο συνελκτικό επίπεδο. Ανάλογα με το πρόβλημα, το πλήθος των εποχών κάθε σταδίου επιλέχθηκε πειραματικά, μια και η συγκεκριμένη πληροφορία δεν συμπεριλήφθηκε στη δημοσιευμένη εργασία [52], στην οποία στηρίχθηκε το μοντέλο. Για τα προβλήματα πολυ-κατηγοριοποίησης με μη επαυξημένα δεδομένα η πρώτη φάση εκπαίδευσης ολοκληρώθηκε σε 300 εποχές και η δεύτερη σε 150, ενώ για το πρόβλημα πολυ-κατηγοριοποίησης με επαυξημένα δεδομένα η πρώτη σε 45 και η δεύτερη σε 35, καθώς παρατηρήθηκε πολύ ταχύτερη σύγκλιση δεδομένου του σημαντικά αυξημένου μεγέθους του συνόλου εκπαίδευσης. Για όλα τα προβλήματα, το μέγεθος δέσμης τέθηκε ίσο με 32, οι διαστάσεις μήκους και πλάτους εισόδου προσαρμόστηκαν όπως ενδείκνυται για την εν λόγω αρχιτεκτονική ( $224 \times 224$ ), ενώ εφαρμόστηκε ο αλγόριθμος βελτιστοποίησης Adam, όπως και στη βιβλιογραφία [52], με αρχικό ρυθμό μάθησης 0.0001 ο οποίος επιλέχθηκε πειραματικά.

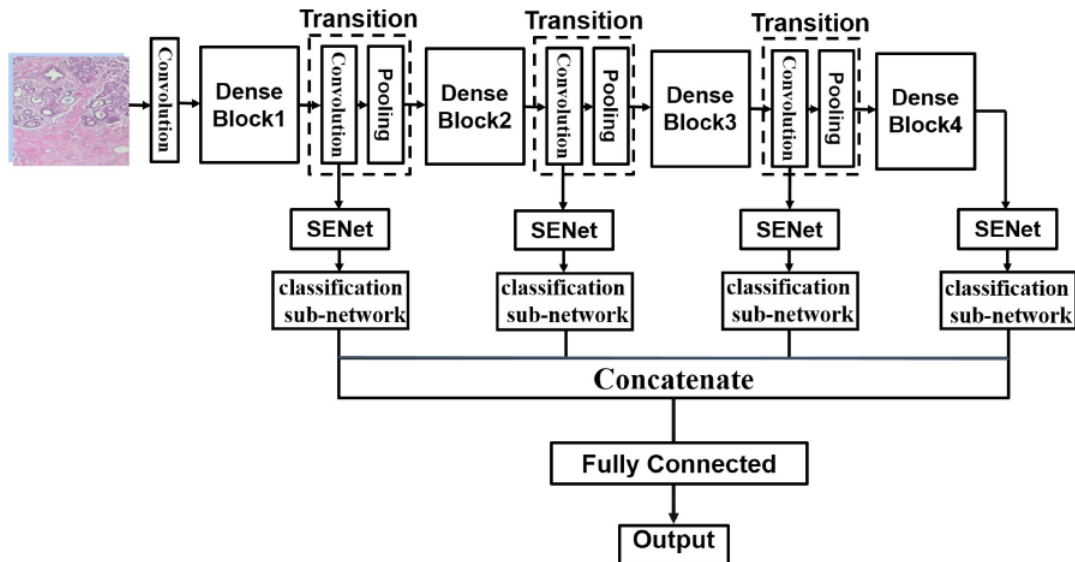
## 5.3 IDSNet

### 5.3.1 Αρχιτεκτονική

Το παρόν μοντέλο υλοποιήθηκε από τους Li et al. [45] ειδικά για το πρόβλημα δυαδικής κατηγοριοποίησης σε δεδομένα ανεξαρτήτου μεγέθυνσης, του συνόλου δεδομένων που εξετάζεται και συνδυάζει τις αρχιτεκτονικές DenseNet121 και SENet, με στόχο την αξιοποίηση των πλεονεκτημά-



των αμφότερων των κύριων δομικών τους μονάδων. Το υβριδικό μοντέλο που προτείνουν, ενέχει την καινοτομία της αξιοποίησης χαρακτηριστικών από επίπεδα διαφορετικού βάθους και τη χρησιμοποίηση αυτών για περαιτέρω εξαγωγή χρήσιμης πληροφορίας σε επίπεδο καναλιών μέσω της δομικής μονάδας Squeeze-and-Excitation. Η πληροφορία αυτή στη συνέχεια διοχετεύεται σε υποδίκτυα κατηγοριοποίησης, των οποίων οι έξοδοι συνενώνονται και οδηγούνται στο συνολικό επίπεδο κατηγοριοποίησης.



Σχήμα 5.2: Η αρχιτεκτονική IDSNet

Αναλυτικότερα, όπως φαίνεται και από τη σχηματική απεικόνιση 5.2 [45] του μοντέλου, η έξοδος από κάθε επίπεδο μετάβασης, όπως και η έξοδος του τελευταίου dense block διέρχονται από μία δομική μονάδα Squeeze-and-Excitation με τα μεγέθη των πλήρων επιπέδων της, όσον αφορά τη λειτουργία διέγερσης (excitation) να είναι 32 και 256, 64 και 512, 128 και 1024 για τις μονάδες με είσοδο από το πρώτο, δεύτερο και τρίτο επίπεδο μετάβασης αντίστοιχα, και ομοίως 128 και 1024 για τη μονάδα με είσοδο από το τελευταίο dense block. Σημειώνεται ακόμη ότι η έξοδος κάθε SENet προστίθεται στην είσοδό του, όπως συμβαίνει και στις κύριες δομικές μονάδες τύπου ResNet. Αναφορικά με τα υποδίκτυα κατηγοριοποίησης αποτελούνται από ένα επίπεδο εξαγωγής καθολικού μέσου ανά κανάλι (global average pooling), ένα επίπεδο κανονικοποίησης δέσμης κι ένα πλήρως συνδεδεμένο επίπεδο με συνάρτηση ενεργοποίησης Softmax και με πλήθος νευρώνων ίσο με το πλήθος των κλάσεων ανάλογα το πρόβλημα -στην βιβλιογραφία [45] το πλήθος των νευρώνων τέθηκε ίσο με δύο αφού εξετάστηκε μόνο το πρόβλημα δυαδικής κατηγοριοποίησης -. Το τελικό στάδιο κατηγοριοποίησης υλοποιείται από ένα επίπεδο απόρριψης (Dropout) ακολουθούμενο από ένα πλήρες επίπεδο με πλήθος νευρώνων ίσο με το πλήθος των κλάσεων του προβλήματος, επίσης με συνάρτηση ενεργοποίησης Softmax.

### 5.3.2 Εκπαίδευση

Για όλα τα προβλήματα επιλέχθηκε, όπως και στις προηγούμενες περιπτώσεις μέγεθος δέσμης 32 και οι διαστάσεις μήκους και πλάτους της εισόδου, ομοίως προσαρμόστηκαν σε  $224 \times 224$ , το προτεινόμενο μέγεθος εισόδου για αρχιτεκτονική DenseNet121. Ως αλγόριθμος βελτιστοποίησης επίσης

για όλα τα προβλήματα επιλέχθηκε ο Adam, όπως και στη βιβλιογραφία [45], ενώ ο ρυθμός μάθησης για τα προβλήματα πολυ-κατηγοριοποίησης τέθηκε ίσος με 0.00001, ενώ για τα δυαδικής κατηγοριοποίησης ακολουθήθηκε το πρωτόκολλο των Li et al. [45], σύμφωνα με το οποίο λάμβανε τιμές ανάλογα με την εποχή και συγκεκριμένα τιμές 0,0003,00015, 0.000075, 0,00003 και 0,00001 μέχρι την 25<sup>η</sup>, 30<sup>η</sup>, 35<sup>η</sup>,40<sup>η</sup> και 60<sup>η</sup> εποχή αντίστοιχα (εφόσον αυτές υπήρξαν). Για το πρόβλημα δυαδικής κατηγοριοποίησης με δεδομένα κοινής μεγέθυνσης το μοντέλο εκπαιδεύτηκε για 20 εποχές, ενώ για το αντίστοιχο πρόβλημα με δεδομένα ανεξαρτήτου μεγέθυνσης για 60, όπως και για το πρόβλημα πολυ-κατηγοριοποίησης κοινής μεγέθυνσης με επαυξημένα δεδομένα. Για τα υπόλοιπα προβλήματα η εκπαίδευση διήρκεσε 80 εποχές.



## Κεφάλαιο 6

### Πειραματικά αποτελέσματα

Για κάθε έκφανση του του προβλήματος της κατηγοριοποίησης ιστοπαθολογικών εικόνων καρκίνου του μαστού, όπως αυτές έχουν οριστεί για το εξεταζόμενο σύνολο δεδομένων, κάθε μοντέλο εκπαιδεύτηκε και αξιολογήθηκε σύμφωνα με την παραμετροποίηση και τις μετρικές που αναφέρονται στην προηγούμενη ενότητα. Στη συνέχεια παρατίθενται και σχολιάζονται συνοπτικά για κάθε πρόβλημα που εξετάστηκε, πίνακες με τις τιμές κάθε μετρικής για κάθε μοντέλο, όπως προέκυψαν από την αξιολόγηση στο σύνολο ελέγχου. Για τα προβλήματα πολυ-κατηγοριοποίησης, όπου απαιτείται οι τιμές ορισμένων μετρικών να υπολογιστούν ως μέση τιμή των επιμέρους τιμών κάθε κλάσης, χρησιμοποιήθηκε η τιμή macro average. Σημειώνεται ότι στους πίνακες αποτελεσμάτων το επίθεμα "\_aug" υποδηλώνει ότι προηγήθηκε εκπαίδευση του μοντέλου σε επαυξημένα δεδομένα, σύμφωνα με τις τεχνικές που περιγράφηκαν παραπάνω, ενώ για λόγους συντόμευσης οι μετρικές ακρίβειας σε επίπεδο εικόνας και ασθενούς αναφέρονται ως IL Acc (Image Level Accuracy) και PL Acc (Patient Level Accuracy) αντίστοιχα.

#### 6.1 Δυαδική Κατηγοριοποίηση Κοινής Μεγέθυνσης

##### 6.1.1 Μεγέθυνση 40X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.939	0.945	0.925	0.988	0.956
DenseNet121	0.891	0.876	0.951	0.879	0.914
IDSNet	0.948	0.936	0.954	0.970	0.962
VGG16_aug	<b>0.962</b>	<b>0.955</b>	0.946	<b>1.000</b>	<b>0.972</b>
DenseNet121_aug	0.920	0.912	<b>0.982</b>	0.895	0.936
IDSNet_aug	0.955	0.946	0.975	0.954	0.965

Πίνακας 6.1: Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης 40×

Εξαιρετικές επιδόσεις όσον αφορά την ακρίβεια τόσο σε επίπεδο εικόνας όσο και ασθενούς παρουσίασαν όλα τα μοντέλα, με το VGG16 να προηγείται συνολικά όλων και το IDSNet να ακολουθεί με μικρή διαφορά. Από τη σύγκριση των τιμών κάθε μετρικής, όπως προέκυψε από τη χρήση επαυξημένων ή μη δεδομένων για εκπαίδευση, διαφαίνεται ξεκάθαρα ο ευεργετικός ρόλος της εφαρμογής των τεχνικών επαύξησης μια και οι μέγιστες τιμές παρατηρούνται όπου αυτές εφαρμόστηκαν. Εξαιρετικής σημασίας και ισχυρή ένδειξη για την πολύ υψηλή απόδοση του VGG16 αποτελεί η τιμή της

μετρικής Recall, η οποία στην περίπτωση εκπαίδευσης σε επαυξημένα δεδομένα αγγίζει το απόλυτο. Η φυσική σημασία του αποτελέσματος αυτού είναι ότι δεν υπήρξαν ψευδώς αρνητικές προβλέψεις, δηλαδή κακοήθη δείγματα που κατηγοριοποιήθηκαν ως καλοήθη, γεγονός που αποτελεί έναν από τους βασικούς στόχους ενός συστήματος υποβοήθησης ιατρικής διάγνωσης, αλλά και κάθε τελικής ιατρικής γνωμάτευσης καθώς ειδικά για μία τόσο σοβαρή ασθένεια, οι ψευδώς αρνητικές προβλέψεις δεν είναι αποδεκτές καθώς οδηγούν σε ανθρώπινες απώλειες.

### 6.1.2 Μεγέθυνση 100X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.898	0.894	0.897	<b>0.951</b>	0.923
DenseNet121	0.894	0.898	0.910	0.928	0.919
IDSNet	0.900	0.861	<b>0.981</b>	0.865	0.920
VGG16_aug	0.909	0.906	0.947	0.911	0.929
DenseNet121_aug	<b>0.927</b>	<b>0.922</b>	0.967	0.918	<b>0.942</b>
IDSNet_aug	0.880	0.847	0.977	0.833	0.899

Πίνακας 6.2: Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης 100×

Διαφορετικός νικητής για τα δεδομένα της εν λόγω μεγέθυνσης, σε σχέση με την προηγούμενη, αναδεικνύεται, με βάση την ακρίβεια σε επίπεδο ασθενούς αλλά και εικόνας, αλλά και με βάση το F1-score, η αρχιτεκτονική DenseNet121. Αν και στην παρούσα υποκατηγορία προβλήματος εκ των αποτελεσμάτων η εφαρμογή των τεχνικών επαύξης στην εκπαίδευση δεν ευνόησαν όλες τις μετρικές σε όλα τα μοντέλα, η καλύτερη επίδοση συνολικά σημειώθηκε και πάλι σε μοντέλο στου οποίου την εκπαίδευση εφαρμόστηκαν. Ακόμη, είναι αξιοσημείωτο ότι η αρχιτεκτονική VGG16 χωρίς να κατέχει τα πρωτεία σε ό,τι αφορά την ακρίβεια σε οποιοδήποτε επίπεδο, επιτυγχάνει με διαφορά την υψηλότερη επίδοση με κριτήριο την εξέχουσας σημασίας, όπως τονίστηκε ανωτέρω, μετρική Recall.

### 6.1.3 Μεγέθυνση 200X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.900	0.893	0.880	<b>0.975</b>	0.925
DenseNet121	0.905	0.902	0.890	0.968	0.927
IDSNet	<b>0.967</b>	<b>0.970</b>	<b>0.975</b>	0.973	<b>0.974</b>
VGG16_aug	0.900	0.884	0.952	0.886	0.918
DenseNet121_aug	0.903	0.887	0.903	0.948	0.925
IDSNet_aug	0.959	0.951	0.967	0.967	0.967

Πίνακας 6.3: Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης 200×

Υψηλότερη επίδοση σε αυτήν την υποκατηγορία του προβλήματος, σημείωσε το μοντέλο IDSNet για την πλειοψηφία των μετρικών, με την τιμή ακρίβειας σε επίπεδο εικόνας να προηγείται σχεδόν 7% από τις τιμές των άλλων μοντέλων, αφήνοντάς τα αρκετά πίσω. Η συγκεκριμένη αρχιτεκτονική

προηγείται στην περίπτωση εκπαίδευσης των μοντέλων τόσο με μη όσο επαυξημένο και με επαυξημένο σύνολο εκπαίδευσης. Σε αντίθεση με τις προηγούμενες περιπτώσεις οι βέλτιστες τιμές να μην παρατηρούνται για εκπαίδευση σε μη επαυξημένο σύνολο δεδομένων, ωστόσο οι διαφορές ειδικά για την περίπτωση της αρχιτεκτονικής που σημείωσε την υψηλότερη επίδοση, είναι στην πλειοψηφία των περιπτώσεων μικρότερες του 1%. Αναφορικά με τη μετρική ανάκλησης, για μία ακόμα φορά μέγιστη τιμή επιτυγχάνεται με την αρχιτεκτονική VGG16, με την αρχιτεκτονική IDNet ωστόσο να ακολουθεί με διαφορά μόλις 0.2%, με αποτέλεσμα συνολικά η τελευταία, λαμβάνοντας υπόψη και τις υψηλότερες τιμές των άλλων μετρικών, να φαίνεται ως καταλληλότερη για τα δεδομένα της παρούσας μελέτης.

### 6.1.4 Μεγέθυνση 400X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.826	0.847	0.788	<b>0.983</b>	0.875
DenseNet121	0.870	0.865	0.886	0.908	0.897
IDNet	0.896	0.862	0.907	0.928	0.917
VGG16_aug	0.913	0.906	0.901	0.963	0.931
DenseNet121_aug	0.891	0.871	0.920	0.901	0.910
IDNet_aug	<b>0.941</b>	<b>0.924</b>	<b>0.976</b>	0.926	<b>0.950</b>

Πίνακας 6.4: Αποτελέσματα δυαδικού προβλήματος για δεδομένα μεγέθυνσης 400×

Νικητής με διαφορά για την πλειοψηφία των μετρικών αναδεικνύεται όπως και στην προηγούμενη περίπτωση η αρχιτεκτονική IDNet, αυτή η φορά εκπαιδευμένη σε επαυξημένα δεδομένα. Εξαιρεση αποτελεί για ακόμη μία φορά η μετρική ανάκλησης με βάση την οποία βέλτιστη τιμή παρουσιάζει το μοντέλο VGG16, παρόλη τη σχετικά χαμηλή επίδοση με βάση τις υπόλοιπες μετρικές.

## 6.2 Κατηγοριοποίηση Πολλών Κλάσεων Κοινής Μεγέθυνσης

### 6.2.1 Μεγέθυνση 40X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.619	0.651	0.418	0.502	0.456
DenseNet121	0.541	0.567	0.370	0.461	0.411
IDNet	0.567	0.590	0.369	0.471	0.414
VGG16_aug	<b>0.639</b>	<b>0.676</b>	<b>0.585</b>	<b>0.633</b>	<b>0.608</b>
DenseNet121_aug	0.534	0.570	0.341	0.442	0.385
IDNet_aug	0.558	0.579	0.339	0.451	0.387

Πίνακας 6.5: Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης 40×

Με βάση τα πειραματικά αποτελέσματα του πίνακα 6.5, αδιαφιλονίκητος νικητής αναδεικνύεται η αρχιτεκτονική VGG16 και μάλιστα για εκπαίδευση σε επαυξημένα δεδομένα, αναδεικνύοντας και

για την παρούσα υποκατηγορία του προβλήματος, την αξία της εφαρμογής των τεχνικών επαύξησης στα δεδομένα εκπαίδευσης. Ακόμα, όμως και στις περιπτώσεις όπου τα μοντέλα δεν εκπαιδεύτηκαν με επαυξημένα δεδομένα, παρατηρείται επίσης η υψηλότερη επίδοση για τη συγκεκριμένη αρχιτεκτονική και μάλιστα σε όλες τις μετρικές, γεγονός που οδηγεί στο συμπέρασμα ότι το συγκεκριμένο μοντέλο είναι καταλληλότερο για το εν λόγω υποπρόβλημα, σε σχέση με τις άλλες αρχιτεκτονικές. Σε αντίθεση, με το δυαδικό πρόβλημα, όλα τα μοντέλα παρουσιάζουν εξαιρετικά χαμηλότερη επίδοση, προφανώς μη ικανοποιητική, γεγονός που κάνει αντιληπτή την αυξημένη δυσκολία του προβλήματος, όταν το πλήθος των κλάσεων είναι μεγαλύτερο του δύο. Στο σημείο αυτό, αξίζει να σημειωθεί ότι παρόλο που στη βιβλιογραφία [52] αναφέρονται πολύ ικανοποιητικά αποτελέσματα για την αρχιτεκτονική DenseNet121, η έλλειψη λεπτομερούς περιγραφής του πρωτοκόλλου εκπαίδευσης και λόγω αυτής η αντίστοιχη πειραματική επιλογή ορισμένων παραμέτρων, δεν κατέστησαν δυνατή την αναπαραγωγή ή την προσέγγισή τους, για τα δεδομένα τόσο αυτής, όσο και των υπολοίπων μεγεθύνσεων.

### 6.2.2 Μεγέθυνση 100X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.644	<b>0.702</b>	<b>0.571</b>	<b>0.575</b>	<b>0.573</b>
DenseNet121	0.556	0.584	0.429	0.467	0.447
IDSNet	0.562	0.614	0.483	0.502	0.492
VGG16_aug	<b>0.653</b>	0.690	0.491	0.570	0.527
DenseNet121_aug	0.564	0.599	0.401	0.461	0.429
IDSNet_aug	0.577	0.612	0.398	0.478	0.434

Πίνακας 6.6: Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθυνσης 100×

Για το εν λόγω υποπρόβλημα, η αρχιτεκτονική VGG16 συνεχίζει να κατέχει τα πρωτεία σε όλες τις μετρικές, με τις υψηλότερες τιμές να σημειώνονται για εκπαίδευση σε μη επαυξημένα δεδομένα με εξαίρεση την ακρίβεια σε επίπεδο εικόνας, η οποία παρουσίασε μέγιστη τιμή για την ίδια αρχιτεκτονική εκπαιδευμένη σε επαυξημένα δεδομένα. Η διαφορά ωστόσο των δύο τιμών είναι κάτω από 0.1% και μπορεί να θεωρηθεί αμελητέα. Ομοίως, παρατηρείται και σε αυτή την περίπτωση επικράτηση του μοντέλου VGG16 σε σχέση με τα υπόλοιπα για εκπαίδευση στο ίδιο σύνολο δεδομένων, ενώ οι απόλυτες τιμές των μετρικών συνεχίζουν να μην καθίστανται ικανοποιητικές, με θετική ωστόσο εξέλιξη της τιμής της ακρίβειας σε επίπεδο ασθενούς που άγγιξε το 70%.

### 6.2.3 Μεγέθυνση 200X

Με κριτήρια τις μετρικές απόδοσης ακρίβειας και ανάκλησης, σε αντίθεση με τις προηγούμενες δύο περιπτώσεις, προηγείται η αρχιτεκτονική IDSNet εκπαιδευμένη στα αρχικά δεδομένα, ενώ καλύτερη απόδοση με βάση τις υπόλοιπες μετρικές παρουσιάζει το μοντέλο DenseNet121 με εκπαίδευση στα επαυξημένα. Ωστόσο μεταξύ των δύο μοντέλων, λαμβάνοντας υπόψη και τη σημαντική διαφορά στην τιμή της ανάκλησης και τις σχετικά μικρές διαφορές στις τιμές των υπολοίπων μετρικών, καλύτερο φαίνεται να είναι το IDSNet μοντέλο. Οι ακόμη πιο μειωμένες τιμές των μετρικών σε σχέση με τις αντίστοιχες στα υποπροβλήματα με δεδομένα μικρότερης μεγέθυνσης, αναδεικνύει ότι η αύ-

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.539	0.559	0.254	0.331	0.288
DenseNet121	0.539	0.547	0.360	0.345	0.352
IDSNet	<b>0.567</b>	<b>0.600</b>	0.344	<b>0.462</b>	0.395
VGG16_aug	0.564	0.588	0.294	0.391	0.336
DenseNet121_aug	0.561	0.578	<b>0.399</b>	0.395	<b>0.397</b>
IDSNet_aug	0.539	0.556	0.291	0.378	0.329

Πίνακας 6.7: Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθους 200×

ξηση της μεγέθυνσης, με αποτέλεσμα όλο και μεγαλύτερο τμήμα κάθε δείγματος να περιλαμβάνει τμήμα της περιοχής ενδιαφέροντος, εντείνει την πολυπλοκότητα του προβλήματος, δυσχεραίνοντας έτι περισσότερο την ανάπτυξη της ικανότητας γενίκευσης των μοντέλων.

#### 6.2.4 Μεγέθυνση 400X

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.521	0.521	0.303	0.382	0.338
DenseNet121	0.519	0.528	0.362	0.340	0.351
IDSNet	0.556	0.555	0.324	0.362	0.342
VGG16_aug	0.484	0.517	0.289	0.278	0.283
DenseNet121_aug	0.536	0.540	0.339	0.349	0.344
IDSNet_aug	<b>0.578</b>	<b>0.591</b>	<b>0.391</b>	<b>0.421</b>	<b>0.405</b>

Πίνακας 6.8: Αποτελέσματα κατηγοριοποίησης πολλών κλάσεων για δεδομένα μεγέθους 400×

Η αρχιτεκτονική που προηγείται με διαφορά για τις τιμές όλων των μετρικών, είναι και πάλι η IDSNet με εκπαίδευση αυτή τη φορά στο επαυξημένο σύνολο δεδομένων, ενώ μεταξύ των μοντέλων με εκπαίδευση και στο αρχικό σύνολο δεδομένων προηγείται και πάλι όσον αφορά τις μετρικές ακρίβειας. Αντίστοιχα μειωμένες παρουσιάζονται σε γενικό επίπεδο οι τιμές των μετρικών σε σχέση με εκείνες στις μικρότερες μεγεθύνσεις, επιβεβαιώνοντας την αύξηση της δυσκολίας του προβλήματος κατηγοριοποίησης, για αυξημένα επίπεδα μεγέθυνσης.

### 6.3 Δυαδική Κατηγοριοποίηση Ανεξαρτήτου Μεγέθυνσης

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	0.906	0.889	0.936	0.917	0.926
DenseNet121	0.914	<b>0.892</b>	<b>0.956</b>	0.909	0.932
IDSNet	<b>0.920</b>	0.890	0.942	<b>0.931</b>	<b>0.936</b>

Πίνακας 6.9: Αποτελέσματα δυαδικής κατηγοριοποίησης ανεξαρτήτου μεγέθυνσης

Το παρόν πρόβλημα καθίσταται ναί μεν απλούστερο σε σχέση με το πρόβλημα πολυ-κατηγοριοποίησης, πολυπλοκότερο δε από τη δυαδική κατηγοριοποίηση με δεδομένα κοινής μεγέθυνσης, δεδομένης της ακόμη πιο έντονης ποικιλομορφίας των αρχικών δεδομένων λόγω της συνύπαρξης εικόνων διαφορετικής μεγέθυνσης. Παρόλη την αυξημένη πολυπλοκότητα, όλα τα μοντέλα επιδεικνύουν αρκετά ικανοποιητική επίδοση για όλες τις μετρικές, με το IDNet να προηγείται στην πλειοψηφία των μετρικών. Στη δεύτερη θέση με πολύ μικρή διαφορά έρχεται η απλή αρχιτεκτονική DenseNet121, με εξαίρεση την καίριας σημασίας μετρική ανάκλησης, όπου και το μοντέλο IDNet προηγείται με ελαφρά μεγαλύτερη. Λαμβάνοντας υπόψη ότι η αρχιτεκτονική αυτή αναπτύχθηκε από τους Li et al.[45] για τη συγκεκριμένη έκφανση του προβλήματος, από τα αποτελέσματα διαφαίνεται η αξία της ενσωμάτωσης της δομικής μονάδας Squeeze-and-Excitation στην αρχιτεκτονική DenseNet121, αφού το υβριδικό αυτό μοντέλο σημείωσε καλύτερες επιδόσεις σε σχέση με το απλό.

## 6.4 Κατηγοριοποίηση Πολλών Κλάσεων Ανεξαρτήτου Μεγέθυνσης

Μοντέλο	IL Acc	PL Acc	Precision	Recall	F1-score
VGG16	<b>0.617</b>	<b>0.644</b>	<b>0.431</b>	<b>0.476</b>	<b>0.453</b>
DenseNet121	0.536	0.562	0.339	0.441	0.383
IDNet	0.554	0.569	0.306	0.422	0.355

Πίνακας 6.10: Αποτελέσματα πολυ-κατηγοριοποίησης ανεξαρτήτου μεγέθυνσης

Για το τελευταίο αυτό πολυσύνθετο πρόβλημα, το οποίο παρουσιάζει τη μέγιστη πολυπλοκότητα αφενός λόγω των διαφορετικών μεγεθύνσεων των δεδομένων, όπως και στην προηγούμενη περίπτωση, αφετέρου λόγω του αυξημένου αριθμού των κλάσεων, ξεκάθαρα νικήτη για όλες τις μετρικές αποτελεί το μοντέλο VGG16. Οι φαινομενικά μη ικανοποιητικές απόλυτες τιμές των μετρικών, που επ' ουδενί δεν καθιστούν το σύστημα αξιόπιστο, σε αντίθεση με τις εξαιρετικά ικανοποιητικές τιμές για το αντίστοιχο δυαδικό πρόβλημα αναδεικνύουν το πόσο αυξημένη είναι η δυσκολία επίλυσης ενός προβλήματος με δεδομένα που ανήκουν σε περισσότερες από δύο κλάσεις.

## Κεφάλαιο 7

# Συμπεράσματα - Μελλοντικές επεκτάσεις

### 7.1 Συμπεράσματα

Αξιολογώντας τη συνολική εικόνα των αποτελεσμάτων για όλες τις πτυχές του προβλήματος, για κάθε αρχιτεκτονική, παρατηρήθηκε ότι ανάλογα με το πρόβλημα αναδεικνυόταν διαφορετική αρχιτεκτονική ως καταλληλότερη, με βάση τις τιμές των μετρικών. Παρόλο που δεν προέκυψε μοναδικός νικητής για όλα τα προβλήματα, στη συντριπτική πλειοψηφία των περιπτώσεων, τα πρωτεία κατείχαν τα μοντέλα VGG16 και IDSNet.

Ειδικά για τα υποπροβλήματα με δεδομένα κοινής μεγέθυνσης, παρουσιάζει εξαιρετικό ενδιαφέρον το γεγονός ότι ανεξάρτητα από το πλήθος των κλάσεων, καλύτερη επίδοση στα χαμηλότερα επίπεδα μεγέθυνσης παρουσιάζει η αρχιτεκτονική VGG16, ενώ στα υψηλότερα το μοντέλο IDSNet. Η συμπεριφορά αυτή αναδεικνύει την καταλληλότητα της αντίστοιχης αρχιτεκτονικής αφενός για την αναγνώριση εικόνων με πιο "συμπιεσμένη" και ανωτέρου επιπέδου πληροφορία, όταν η μεγέθυνση είναι χαμηλή, αφετέρου για την αναγνώριση εικόνων με αναλυτικότερη πληροφορία και περισσότερη λεπτομέρεια, στην περίπτωση μεγαλύτερης μεγέθυνσης. Η πειραματικά παρατηρούμενη συμπεριφορά αυτή, αιτιολογείται απόλυτα λαμβάνοντας υπόψη την αρχιτεκτονική των δύο μοντέλων. Το VGG16 ως "συμβατική" βαθιά συνελκτική αρχιτεκτονική, ακολουθεί τη λογική της κατηγοριοποίησης με βάση τα υψηλού επιπέδου χαρακτηριστικά που εξήχθησαν από τα ανώτερα επίπεδα, οπότε αναμενόμενο είναι να αποδίδει καλύτερα για εικόνες με πλούσια υψηλού επιπέδου πληροφορία όπως οι εικόνες υψηλής μεγέθυνσης. Αντίθετα στην υβριδική αρχιτεκτονική IDSNet, για την τελική κατηγοριοποίηση λαμβάνονται υπόψη χαρακτηριστικά από επίπεδα διαφορετικού βάθους, τα οποία επίσης λόγω της ιδιαίτερης αρχιτεκτονικής κάθε dense block, έχουν εξαχθεί λαμβάνοντας υπόψη την έξοδο επιπέδων διαφορετικού βάθους. Τα παραπάνω σε συνδυασμό με τη προσθήκη των δομικών μονάδων Squeeze-and-Excitation, συνιστούν ένα μοντέλο που αναμένεται να είναι σε θέση να αξιοποιεί ταυτόχρονα πληροφορίες διαφορετικών επιπέδων, άρα και να είναι καταλληλότερο για εικόνες με μεγαλύτερη λεπτομέρεια όπως οι εικόνες μεγαλύτερης μεγέθυνσης.

Όσον αφορά τα πιο σύνθετα προβλήματα με δεδομένα από όλα τα επίπεδα μεγέθυνσης, για το σχετικά ευκολότερο δυαδικό πρόβλημα επέδειξε καλύτερη επίδοση το μοντέλο IDSNet, ενώ για το πρόβλημα πολλών κλάσεων καταλληλότερο αναδείχθηκε το VGG16. Η μεγάλη αύξηση της πολυπλοκότητας του προβλήματος με την αύξηση του πλήθους των κλάσεων διαφαίνεται ξεκάθαρα από όλα τα πειράματα για το πρόβλημα της κατηγοριοποίησης πολλών κλάσεων, από τις πολύ χαμηλότερες τιμές όλων των μετρικών σε σχέση με τις τιμές για τα αντίστοιχα δυαδικά προβλήματα. Ωστόσο, οι εξαιρετικά καλές επιδόσεις που σημειώθηκαν για όλες τις υποκατηγορίες του δυαδικού προβλήματος, αναδεικνύουν τη μεγάλη αξία της μεταφοράς μάθησης, αφού χάρη σε αυτή επετεύχθη η δημιουργία

μοντέλων με πολύ ικανοποιητικές ικανότητες γενίκευσης και υψηλή αξιοπιστία με βάση την πειραματική αξιολόγηση. Ακόμη, η εφαρμογή τεχνικών επαύξησης στα δεδομένα εκπαίδευσης επέδρασε αρκετά θετικά σε γενικές γραμμές, όπως άλλωστε αναμενόταν.

## 7.2 Μελλοντικές επεκτάσεις

Μια πολύ ενδιαφέρουσα κατεύθυνση μελέτης των δυνατοτήτων της μεταφοράς μάθησης, ειδικά για το πρόβλημα της κατηγοριοποίησης σε πολλές κλάσεις, στον τομέα της επιβλεπόμενης μάθησης, θα ήταν η δημιουργία προεκπαιδευμένων μοντέλων σε δεδομένα παραπλήσιας φύσης με τα υπό εξέταση. Δεδομένης της πολύ πρόσφατης προσθήκης ενός πολύ μεγάλου συνόλου δεδομένων, με υποτιμήματα ιστοπαθολογικών εικόνων καρκίνου των λεμφαδένων, του Patch Camelyon [68], στην εξαιρετικά δημοφιλή συλλογή συνόλων δεδομένων [Tensorflow Datasets](#), είναι πρακτικά ζήτημα χρόνου, να μελετηθεί από την ερευνητική κοινότητα, η επίδραση της μεταφοράς μάθησης από αρκετά πιο όμοια δεδομένα σε σχέση με αυτά του ImageNet. Κρίνοντας από τα εξαιρετικά αποτελέσματα που επέδειξε η εφαρμογή της μεταφοράς μάθησης από τόσο διαφορετικά δεδομένα, για το δυαδικό πρόβλημα κατηγοριοποίησης ιστοπαθολογικών εικόνων καρκίνου του μαστού, η ανάπτυξη μοντέλων με μεταφορά μάθησης από δεδομένα πολύ πιο όμοια με τα υπό εξέταση, μόνο προσδοφόρα αναμένεται να είναι.

Δεδομένου του μειονεκτήματος της επιβλεπόμενης μάθησης, που πηγάζει από την απαίτηση πολύ μεγάλου όγκου δεδομένων προκειμένου να εξασφαλιστεί μεγάλη ικανότητα γενίκευσης, η στροφή στην βαθύτερη κι εκτενέστερη μελέτη μη επιβλεπόμενων μεθόδων μάθησης, σε συνδυασμό με μεταφορά μάθησης, καθίσταται επίσης μία πολύ ενδιαφέρουσα εναλλακτική προσέγγιση. Στην ίδια λογική, η αξιοποίηση προεκπαιδευμένων μοντέλων για εξαγωγή χαρακτηριστικών, σε συνδυασμό με ευφυείς πράκτορες (agents) ενισχυτικής μάθησης αποτελεί μία ακόμα ανεξερεύνητη και πολλά υποσχόμενη κατεύθυνση για το παρόν πρόβλημα, δεδομένων των ολοένα και μεγαλύτερων δυνατοτήτων που παρουσιάζουν τα μοντέλα ενισχυτικής μάθησης σε όλο και περισσότερες εφαρμογές.



## Βιβλιογραφία

- [1] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan και et al. *BACH: Grand challenge on breast cancer histology images. Medical Image Analysis*, 56:122–139, 2019.
- [2] Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi, Oscar Geessink, Maschenka Balkenhol, Peter Bult, Altuna Halilovic, Meyke Hermsen, Robvan de Loo, Rob Vogels, Quirine F Manson, Nikolas Stathonikos, Alexi Baidoshvili, Paulvan Diest, Carla Wauters, Marcoryvan Dijk και Jeroenvan der Laak. *1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience*, 7(6), 2018. giy065.
- [3] F. A. Spanhol, L. S. Oliveira, C. Petitjean και L. Heutte. *A Dataset for Breast Cancer Histopathological Image Classification. IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016.
- [4] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.
- [5] F. Rosenblatt. *The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review*, 65 6:386–408, 1958.
- [6] W. McCulloch και W. Pitts. *A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [7] G Cybenkot. *Approximation by Superpositions of a Sigmoidal Function \**. 2006.
- [8] Sergey Ioffe και Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015.
- [9] Howard B. Demuth, Mark H. Beale, Orlando De Jess και Martin T. Hagan. *Neural Network Design*. Martin Hagan, Stillwater, OK, USA, 2ndη έκδοση, 2014.
- [10] Paul Werbos και Paul John. *Beyond regression : new tools for prediction and analysis in the behavioral sciences /*. Διδακτορική Διατριβή, Harvard University, 1974.
- [11] Yann Lecun. *Une procedure d'apprentissage pour reseau a seuil asymmetrique (A learning scheme for asymmetric threshold networks). Proceedings of Cognitiva 85, Paris, France*, σελίδες 599–604, 1985.
- [12] D.B. Parker. *Learning-logic: Casting the Cortex of the Human Brain in Silicon*. Technical report: Center for Computational Research in Economics and Management Science.

Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science, 1985.

- [13] David E. Rumelhart, Geoffrey E. Hinton και Ronald J. Williams. *Learning representations by back-propagating errors*. *Nature*, 323(6088):533–536, 1986.
- [14] Γεωργούλη Α. *Τεχνητή νοημοσύνη*. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, 2015.
- [15] Diederik P. Kingma και Jimmy Ba. *Adam: A Method for Stochastic Optimization*, 2017.
- [16] J. Deng, W. Dong, R. Socher, L. Li, Kai Li και Li Fei-Fei. *ImageNet: A large-scale hierarchical image database*. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, σελίδες 248–255, 2009.
- [17] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg και Li Fei-Fei. *ImageNet Large Scale Visual Recognition Challenge*. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [18] Y. LeCun, L. Bottou, Y. Bengio και P. Haffner. *Gradient-Based Learning Applied to Document Recognition*. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [19] Alex Krizhevsky, Ilya Sutskever και Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. *Advances in Neural Information Processing Systems 25*. Pereira, C. J. C. Burges, L. Bottou και K. Q. Weinberger, επιμελητές, σελίδες 1097–1105. Curran Associates, Inc., 2012.
- [20] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2015.
- [21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke και Andrew Rabinovich. *Going Deeper with Convolutions*, 2014.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.
- [23] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke και Alex Alemi. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*, 2016.
- [24] Gao Huang, Zhuang Liu, Laurens van der Maaten και Kilian Q. Weinberger. *Densely Connected Convolutional Networks*, 2018.
- [25] J. Hu, L. Shen και G. Sun. *Squeeze-and-Excitation Networks*. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 7132–7141, 2018.
- [26] F. A. Spanhol, L. S. Oliveira, C. Petitjean και L. Heutte. *Breast cancer histopathological image classification using Convolutional Neural Networks*. *2016 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 2560–2567, 2016.

- [27] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean και L. Heutte. *Deep features for breast cancer histopathological image classification*. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, σελίδες 1868–1873, 2017.
- [28] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng και Trevor Darrell. *DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition*, 2013.
- [29] J.d. Matos, A.d. S. Britto, L. E. S. Oliveira και A. L. Koerich. *Double Transfer Learning for Breast Cancer Histopathologic Image Classification*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8, 2019.
- [30] J. N. Kather, C. A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx και F. G. Zöllner. *Multi-class texture analysis in colorectal cancer histology*. *Scientific Reports*, 6(1):27988, 2016.
- [31] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens και Zbigniew Wojna. *Rethinking the Inception Architecture for Computer Vision*, 2015.
- [32] P.J. Sudharshan, Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte και Paul Honeine. *Multiple instance learning for histopathological breast cancer image classification*. *Expert Systems with Applications*, 117:103 – 111, 2019.
- [33] R. Venkatesan, P. S. Chandakkar και B. Li. *Simpler Non-Parametric Methods Provide as Good or Better Results to Multiple-Instance Learning*. *2015 IEEE International Conference on Computer Vision (ICCV)*, σελίδες 2605–2613, 2015.
- [34] M. Sun, T. X. Han, Ming-Chang Liu και A. Khodayari-Rostamabad. *Multiple Instance Learning Convolutional Neural Networks for object recognition*. *2016 23rd International Conference on Pattern Recognition (ICPR)*, σελίδες 3270–3275, 2016.
- [35] Abdullah Nahid, Ali Mehrabi και Yinan Kong. *Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering*. *BioMed Research International*, 2018:1–20, 2018.
- [36] Sawon Pratiher και Subhankar Chattoraj. *Manifold Learning & Stacked Sparse Autoencoder for Robust Breast Cancer Classification from Histopathological Images*. *ArXiv*, abs/1806.06876, 2018.
- [37] Joshua Tenenbaum, Vin Silva και John Langford. *A Global Geometric Framework for Nonlinear Dimensionality Reduction*. *Science (New York, N.Y.)*, 290:2319–23, 2001.
- [38] N. Bayramoglu, J. Kannala και J. Heikkilä. *Deep learning for magnification independent breast cancer histopathology image classification*. *2016 23rd International Conference on Pattern Recognition (ICPR)*, σελίδες 2440–2445, 2016.
- [39] V. Gupta και A. Bhavsar. *Breast Cancer Histopathological Image Classification: Is Magnification Important?* *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, σελίδες 769–776, 2017.

- [40] Lior Rokach. *Ensemble-based classifiers*. *Artif. Intell. Rev.*, 33:1–39, 2010.
- [41] Shallu και Rajesh Mehra. *Breast cancer histology images classification: Training from scratch or transfer learning?* *ICT Express*, 4(4):247 – 254, 2018.
- [42] V. Gupta και A. Bhavsar. *Partially-Independent Framework for Breast Cancer Histopathological Image Classification*. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, σελίδες 1123–1130, 2019.
- [43] Tianqi Chen και Carlos Guestrin. *XGBoost*. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [44] Julien Meynet και Jean Philippe Thiran. *Information theoretic combination of pattern classifiers*. *Pattern Recognition*, 43(10):3412 – 3421, 2010.
- [45] Xia Li, Xi Shen, Yongxia Zhou, Xiuhui Wang και Tie Qiang Li. *Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)*. *PLOS ONE*, 15(5):1–13, 2020.
- [46] Qicheng Lao και Thomas Fevens. *Case-Based Histopathological Malignancy Diagnosis using Convolutional Neural Networks*, 2019.
- [47] P. Jonnalagedda, D. Schmolze και B. Bhanu. *[Regular Paper] MVPNets: Multi-viewing Path Deep Learning Neural Networks for Magnification Invariant Diagnosis in Breast Cancer*. *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, σελίδες 189–194, 2018.
- [48] Zhongyi Han, Benzhen Wei, Yuanjie Zheng, Yilong Yin, Kejian Li και Shuo Li. *Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model*. *Scientific Reports*, 7(1):4172, 2017.
- [49] Xipeng Pan, Lingqiao Li, Huihua Yang, Zhenbing Liu, Yubei He, Zhongming Li, Yongxian Fan, Zhiwei Cao και Longhao Zhang. *Multi-task Deep Learning for Fine-Grained Classification/Grading in Breast Cancer Histopathological Images*, σελίδες 85–95. 2020.
- [50] François Chollet. *Xception: Deep Learning with Depthwise Separable Convolutions*, 2017.
- [51] Shallu Sharma και Dr. Rajesh Mehra. *Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—a Comparative Insight*. *Journal of Digital Imaging*, 33, 2020.
- [52] Majid Nawaz, Adel A. Sewissy και Taysir Hassan A. Soliman. *Multi-Class Breast Cancer Classification using Deep Learning Convolutional Neural Network*. *International Journal of Advanced Computer Science and Applications*, 9(6), 2018.
- [53] B. Xu, J. Liu, X. Hou, B. Liu, J. Garibaldi, I. O. Ellis, A. Green, L. Shen και G. Qiu. *Attention by Selection: A Deep Selective Attention Approach to Breast Cancer Classification*. *IEEE Transactions on Medical Imaging*, σελίδες 1–1, 2019.

- [54] Ziba Gandomkar, Patrick C. Brennan και Claudia Mello-Thoms. *MuDeRN: Multi-category classification of breast histopathological image using deep residual networks*. *Artificial Intelligence in Medicine*, 88:14 – 24, 2018.
- [55] Juanying Xie, Ran Liu, Joseph Luttrell και Chaoyang Zhang. *Deep Learning Based Analysis of Histopathological Images of Breast Cancer*. *Frontiers in Genetics*, 10:80, 2019.
- [56] Md Zahangir Alom, Chris Yakopcic, Tarek M. Taha και Vijayan K. Asari. *Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network*, 2018.
- [57] Yun Jiang, Li Chen, Hai Zhang και Xiao Xiao. *Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module*. *PLOS ONE*, 14(3):1–21, 2019.
- [58] S. Pratiher και S. Chattoraj. *Diving Deep onto Discriminative Ensemble of Histological Hashing Class-Specific Manifold Learning for Multi-class Breast Carcinoma Taxonomy*. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, σελίδες 1025–1029, 2019.
- [59] Ljupco Todorovski και Sašo Džeroski. *Combining Classifiers with Meta Decision Trees*. *Machine Learning*, 50:223–249, 2003.
- [60] Jie Hu, Li Shen και Gang Sun. *Squeeze-and-Excitation Networks*. 2018.
- [61] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic και Tarek M. Taha. *Inception Recurrent Convolutional Neural Network for Object Recognition*, 2017.
- [62] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha και Vijayan K. Asari. *Improved Inception-Residual Convolutional Neural Network for Object Recognition*, 2017.
- [63] Ghulam Murtaza, Liyana Shuib, Ghulam Mujtaba και Ghulam Raza. *Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach*. *Multimedia Tools and Applications*, 79, 2020.
- [64] Sidharth Mishra, Uttam Sarkar, Subhash Taraphder, Sanjoy Datta, Devi Swain, Reshma Saikhom, Sasmita Panda και Menalsh Laishram. *Principal Component Analysis*. *International Journal of Livestock Research*, σελίδα 1, 2017.
- [65] Werner Dubitzky, Olaf Wolkenhauer, Kwang Hyun Cho και Hiroki Yokota, επιμελητές. *Kullback–Leibler Divergence*, σελίδες 1088–1088. Springer New York, New York, NY, 2013.
- [66] M. Jannesari, M. Habibzadeh, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi και I. Hajirasouliha. *Breast Cancer Histopathological Image Classification: A Deep Learning Approach*. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, σελίδες 2405–2412, 2018.
- [67] Majid Ali Nawaz, Adel A. Sewissy και Taysir Hassan A. Soliman. *Automated Classification of Breast Cancer Histology Images Using Deep Learning Based Convolutional Neural Networks*. 2018.

[68] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen και Max Welling. *Rotation Equivariant CNNs for Digital Pathology*. 2018.