



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Ανάπτυξη και αξιολόγηση συνελικτικών νευρωνικών
δικτύων για την κατηγοριοποίηση ιατρικών εικόνων**

Γεώργιος Ε. Γιαννούλης

Επιβλέπων

Ιωάννης Καλατζής Αναπληρωτής Καθηγητής

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2020



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE DEPARTMENT OF INFORMATICS &
TELECOMMUNICATIONS**

**POSTGRADUATE PROGRAM "INFORMATION TECHNOLOGIES IN MEDICINE AND
BIOLOGY"**

MASTER THESIS

**Development and evaluation of Convolutional Neural
Networks for classification of medical images**

Georgios E. Giannoulis

Supervisor Ioannis Kalatzis, Associate Professor

ATHENS

FEBRUARY 2020

MASTER THESIS

Development and evaluation of Convolutional Neural Networks for classification of medical images

Georgios E. Giannoulis

S.R.N.: ΠΙΒ101

SUPERVISOR: Ioannis Kalatzis, Associate Professor

**EXAMINATION
COMMITTEE:**

Ioannis Kalatzis,	Associate Professor
Spiros Kostopoulos,	Assistant Professor
Panteleimon Asvestas,	Associate Professor

February 2020

ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία εκθέτουμε τα βασικά στοιχεία της θεωρίας των νευρωνικών δικτύων με έμφαση στα συνελκτικικά νευρωνικά δίκτυα και με κύριο σκοπό μας τη μελέτη της κατηγοριοποίησης ιατρικών εικόνων. Πιο συγκεκριμένα χρησιμοποιήθηκαν εικόνες καρκίνου του μαστού που αφορούν σε πορογενές διηθητικό καρκίνωμα, οι οποίες επιθυμούμε να διαχωριστούν σε θετικές και αρνητικές στην ύπαρξη όγκου.

Η υλοποίηση του αλγόριθμου βασίζεται στην τεχνολογία των συνελκτικικών νευρωνικών δικτύων (ΣΝΔ) με μεγάλο βάθος. Τα ΣΝΔ παίρνουν ως είσοδο εικόνες και ανακαλύπτουν συγκεκριμένες πληροφορίες ή χαρακτηριστικά από τα δεδομένα, χωρίς ανθρώπινη παρέμβαση για το τι θεωρούμε σημαντικό για να διαχωριστούν οι εικόνες στις κατηγορίες που ανήκουν.

Τα ΣΝΔ μπορούν να έχουν διάφορες αρχιτεκτονικές με διαφορετικό βάθος, διαφορετικό πλήθος συνελίξεων, διαφορετικές συναρτήσεις ενεργοποίησης και πολλές άλλες παραμέτρους και τεχνολογίες. Συνεπώς τα τελευταία χρόνια μεγάλος όγκος έρευνας γίνεται για την ανακάλυψη αρχιτεκτονικών που είναι οι πλέον κατάλληλες για την κατηγοριοποίηση ιατρικών εικόνων. Στην παρούσα εργασία διερευνούμε ποιοτικά εννέα τροποποιήσεις σε μια βραβευμένη αρχιτεκτονική συνελκτικικών δικτύων, την VGG16.

Σκοπός μας είναι να βελτιώσουμε την ακρίβεια στην κατηγοριοποίηση της συγκεκριμένης κατηγορίας εικόνων, άλλα παράλληλα να κατανοήσουμε βαθύτερα τις δομικές παραμέτρους της αρχιτεκτονικής των ΣΝΔ, για ταχύτερη και πλέον αξιόπιστη εκπαίδευση αντίστοιχων δικτύων για παρόμοια προβλήματα. Επίσης εξετάζουμε τη μεταφοράς γνώσης από ήδη εκπαιδευμένα δίκτυα σε διαφορετικού είδους προβλήματα ώστε να βελτιώσουμε την ακρίβεια στο προαναφερθέν πρόβλημα που πραγματεύεται η παρούσα εργασία.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Συνελκτικικά Νευρωνικά Δίκτυα, Μηχανική Μάθηση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: ιατρική εικόνα, συνελκτικικά δίκτυα, νευρωνικά δίκτυα, διαγνωστικές μέθοδοι, καρκίνος στήθους

ABSTRACT

In this diploma thesis we present basic elements of neural network theory with emphasis on convolutional neural networks (CNN) with the main aim of studying the categorization of medical images. In particular, breast cancer images of the Invasive Ductal Carcinoma (IDC). This form of breast cancer consists of 80% of the cases of breast cancer. We aim to separate given images of tissue regarding to being positive or negative on IDC.

The implementation of the algorithms used, is based on the technology of CNN that exhibit a greater depth than used to be applied. CNN technology makes it possible to take a picture as input and discover specific information or features from the data the network deems important in order to separate the images into pre-specified categories that the images belong without any human input as to what differentiates the categories.

CNNs come in very different architectures, with different depth, different plurality of convolutions, different activation functions as well as many other technologies and techniques. In recent years, therefore, a great deal of research has been made to discover the most suitable architectures. In the present study we qualitatively and quantitative explore 9 modifications to an award-winning neural network architecture, the VGG16. Our aim is to evaluate various structural and quality parameters of a CNN in relation with how should we train such networks for better and faster results. Also, we focus on the transfer of knowledge from already trained networks which solved a different set of problems.

SUBJECT AREA: Neural Networks, Machine Learning

KEYWORDS: medical imaging, convolutional networks, neural networks, diagnostic methods, breast cancer

Θέλω να ευχαριστήσω τον επιβλέποντα Καθηγητή μου κ. Ιωάννη Καλατζή, που μου εμπιστεύτηκε το θέμα και με καθοδήγησε καθ' όλη τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας.

Επίσης θέλω να ευχαριστήσω όλη την οικογένεια μου για την στήριξη που μου παρείχε όλα αυτά τα χρόνια, τη γυναίκα μου που με παρότρυνε και με υπέμεινε καθ' όλη τη διάρκεια του μεταπτυχιακού προγράμματος, την πολυαγαπημένη μου αδερφή για τις διαρκείς συμβουλές τις και τέλος τους φίλους μου Σωτήρη και Hugo που ήταν και παραμένουν συνέχεια πλάι μου.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ.....	17
1.1 Τεχνητή Νοημοσύνη.....	17
1.2 Μηχανική Μάθηση – Γενικά στοιχεία.....	18
1.3 Νευρωνικά Δίκτυα – Γενικά στοιχεία.....	19
1.3.1 Βιολογικοί Νευρώνες.....	19
1.3.2 Τεχνητά Νευρωνικά Δίκτυα.....	21
2. Τεχνικές εκμάθησης στη μηχανική μαθηση.....	23
2.1 Γραμμική παλινδρόμηση.....	23
2.2 Ελαχιστοποίηση συνάρτησης.....	24
2.2.1 Έλεγχος κλίσης.....	25
2.3 Επιβλεπόμενη μάθηση.....	25
2.3.1 Βασικές δομές νευρωνικών δικτύων.....	26
2.4 Νευρωνικά Δίκτυα πολλαπλών επιπέδων.....	27
2.5 Αλγόριθμος Οπισθοδιάδοσης Σφάλματος (Backpropagation Algorithm).....	29
2.6 Stochastic Gradient Descent.....	33
2.7 Ορμή Σύγκλισης.....	35
3. Συνελικτικά νευρωνικά δίκτυα.....	37
3.1 Εισαγωγή.....	37
3.2 Δομή Συνελικτικών Νευρωνικών Δικτύων.....	39
3.2.1 Αρχιτεκτονική Συνελικτικών δικτύων.....	39
3.2.2 Είδη επιπέδων συνελικτικών δικτύων.....	42
3.3 Αξιολόγηση αρχιτεκτονικής.....	50
4. Το σύνολο δεδομένων.....	52
4.1 Βασικές Ιδιότητες Ιατρικής Εικόνας.....	52
4.2 Στοιχεία του Συνόλου Δεδομένων.....	52

5. Μεθοδολογία	56
5.1 Απλό Συνελικτικό Δίκτυο	56
5.2 Η αρχιτεκτονική του δικτύου VGG16	57
5.2.1 Η δομή του VGG16	58
5.2.2 Υλικό και λογισμικό εκτέλεσης πειραμάτων	61
5.2.3 Δομή φακέλων.....	61
5.3 Τροποποίηση Πρώτη, μεταβολή μόνο στο πλήρως συνδεδεμένο επίπεδο (VGG16#1).....	62
5.4 Τροποποίηση Δεύτερη, εκπαίδευση μόνο στο συνελικτικό επίπεδο, αρχικοποίηση βαρών από VGG16#1 (VGG16#2)	63
5.5 Τροποποίηση Τρίτη, διαγραφή τελευταίου συνελικτικού επιπέδου (VGG16#3).....	65
5.6 Τροποποίηση Τέταρτη, διαγραφή τελευταίου συνελικτικού επιπέδου, εκπαίδευση συνελικτικών επιπέδων (VGG16#4)	67
5.7 Τροποποίηση Πέμπτη, διαγραφή δύο τελευταίων συνελικτικών επιπέδων, εκπαίδευση πλήρως συνδεδεμένου δικτύου (VGG16#5)	68
5.8 Τροποποίηση Έκτη, διαγραφή δύο τελευταίων συνελικτικών επιπέδων, εκπαίδευση συνελικτικών επιπέδων δικτύου (VGG16#6).....	69
5.9 Τροποποίηση Έβδομη, εκπαίδευση μόνο 2 τελευταίων συνελικτικών επιπέδων (VGG16#7)	70
5.10 Τροποποίηση Όγδοη, εκπαίδευση μόνο τελευταίου συνελικτικού επιπέδου (VGG16#8)..	71
5.11 Τροποποίηση Ένατη, αρχικοποίηση με VGG16 και εκπαίδευση σε όλο το δίκτυο (VGG16#9)	72
6. Αποτελέσματα	74
6.1 Επισκόπηση υπαρχόντων αποτελεσμάτων	74
6.2 Αποτελέσματα για VGG16#1	75
6.3 Αποτελέσματα για VGG16#2	81
6.4 Αποτελέσματα για VGG16#3	84
6.5 Αποτελέσματα για VGG16#4	88
6.6 Αποτελέσματα για VGG16#5	91

6.7	Αποτελέσματα για VGG16#6	94
6.8	Αποτελέσματα για VGG16#7	97
6.9	Αποτελέσματα για VGG16#8	100
6.10	Αποτελέσματα για VGG16#9	104
7.	Συμπεράσματα	109
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	112
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	115
	Παράρτημα Ι.....	116
	ΑΝΑΦΟΡΕΣ	117

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1-1. Βιολογικός Νευρώνας και τα κύρια στοιχεία του	20
Εικόνα 1-2. Αντιδιαστολή βιολογικού με τεχνητό νευρώνα	22
Εικόνα 2-1. Δομικά στοιχεία υπολογιστικού νευρώνα	26
Εικόνα 2-2. Εικόνα δικτύου perceptron, με ανάδραση σφάλματος	30
Εικόνα 2-3. Gradient Descent στην επιφάνεια σφάλματος	31
Εικόνα 2-4. Πορεία σύγκλισης αλγορίθμου στις περιπτώσεις, (μπλε) batch gradient, (πράσινο) mini-batch gradient, (μωβ) stochastic gradient	35
Εικόνα 3-1. Σχηματική απεικόνιση της μεταβολής των δεδομένων εισόδου	38
Εικόνα 3-2. Βασική δομή πλήρως συνδεδεμένου νευρωνικού δικτύου	41
Εικόνα 3-3. Μετασχηματισμός του αρχικού "όγκου" χαρακτηριστικών, με βαθμιαία αύξηση του βάθους	41
Εικόνα 3-4. Παράδειγμα συνέλιξης και το αποτέλεσμα που έχει στην έξοδο. Είσοδος 7x7, πυρήνας συνέλιξης 3x3, για stride 2 και 0 padding.....	43
Εικόνα 3-5. Εφαρμογή διαφόρων μεγεθών πυρήνα στην είσοδο, με και χωρίς εφαρμογή padding (1 ^η περ. padding=0, stride=1, 2 ^η περ. padding=1, stride=2, 3 ^η περ. padding=1, stride=1)	44
Εικόνα 3-6. Βασική δομή inception, συμπίεση γνώσης από διάφορα μεγέθη και συνελίξεων.....	46
Εικόνα 3-7. Ένα block 5 επιπέδων, με ρυθμό αύξησης k=4. Κάθε επίπεδο παίρνει όλα τα προηγούμενα επίπεδα ως εισόδους.	47
Εικόνα 4-1. (αριστερά) Χαρακτηρισμός περιοχής εικόνας από ειδικό, (δεξιά) κομμάτια της εικόνας που έχουν επιλεγεί ως αρνητικά (κόκκινα) και θετικά (πράσινα).....	54
Εικόνα 4-2. Χαρακτηριστικές εικόνες, (πάνω) μέρος από εικόνες που έχουν αναγνωριστεί ως θετικά σε IDC, (κάτω) μέρος από εικόνες που έχουν χαρακτηριστεί ως αρνητικές στην ύπαρξη IDC	55
Εικόνα 5-1. Δομή Συνελικτικού Δικτύου	56
Εικόνα 5-2. Δομή της αρχιτεκτονικής του συνελικτικού δικτύου AlexNet, προκάτοχο του VGG16	57

Εικόνα 5-3. Αναπαράσταση δικτύου VGG16, με χρήση "όγκων" χαρακτηριστικών, και η πορεία των χαρακτηριστικών μέχρι την κατηγοριοποίηση	60
Εικόνα 5-4. Αναπαράσταση όλων των πράξεων με παρουσίαση και των λεπτομερειών του κάθε επιπέδου για το δίκτυο VGG16, για τα συνελκτικά επίπεδα.	60
Εικόνα 5-5. Απεικόνιση δομής VGG16#1	62
Εικόνα 5-6. Απεικόνιση δομής VGG#2	64
Εικόνα 5-7. Τα σχήματα που γίνονται αντιληπτά από τους νευρώνες, όσο αυξάνουμε το βάθος του δικτύου γίνονται πιο πολύπλοκα	65
Εικόνα 5-8. Απεικόνιση δομής VGG#3	66
Εικόνα 5-9. Απεικόνιση δομής VGG#4	67
Εικόνα 5-10. Απεικόνιση δομής VGG#5	68
Εικόνα 5-11. Απεικόνιση δομής VGG#6	70
Εικόνα 5-12. Απεικόνιση δομής VGG#7	71
Εικόνα 5-13. Απεικόνιση δομής VGG#8	72
Εικόνα 5-14 Απεικόνιση δομής VGG#9	73
Εικόνα 6-1. Accuracy και άλλες μετρικές για το δίκτυο VGG16#1	76
Εικόνα 6-2. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#1	78
Εικόνα 6-3. Accuracy και άλλες μετρικές για το δίκτυο VGG16#1	82
Εικόνα 6-4. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#2	83
Εικόνα 6-5. Accuracy και άλλες μετρικές για το δίκτυο VGG16#3	85
Εικόνα 6-6. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#3	86
Εικόνα 6-7. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#4	89
Εικόνα 6-8. Accuracy και άλλες μετρικές για το δίκτυο VGG16#4	89
Εικόνα 6-9. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#5	92

Εικόνα 6-10. Accuracy και άλλες μετρικές για το δίκτυο VGG16#5	93
Εικόνα 6-11. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#6.....	95
Εικόνα 6-12. Accuracy και άλλες μετρικές για το δίκτυο VGG16#6	96
Εικόνα 6-13. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#7.....	98
Εικόνα 6-14. Accuracy και άλλες μετρικές για το δίκτυο VGG16#7	99
Εικόνα 6-15. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#8.....	102
Εικόνα 6-16. Accuracy και άλλες μετρικές για το δίκτυο VGG16#8	103
Εικόνα 6-17. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#9.....	106
Εικόνα 6-18. Accuracy και άλλες μετρικές για το δίκτυο VGG16#9	107

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Περιγραφή των χαρακτηριστικών που χρησιμοποιήθηκαν ώστε να μπορεί να συγκριθεί το συνελικτικό νευρωνικό δίκτυο αναφοράς.....	75
Πίνακας 2. Αποτελέσματα σύγκρισης CNN με άλλες τεχνικές μηχανικής μάθησης	75
Πίνακας 3. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#1	75
Πίνακας 4. Αποτελέσματα accuracy για 3 διαφορετικές αρχικοποιήσεις μετά από μόλις 1 εποχή εκπαίδευσης.	77
Πίνακας 5. Αποτελέσματα της τροποποίησης VGG16#1 σε μορφή πίνακα σύγκρισης, 6 ^η Εποχή Εκπαίδευσης.....	79
Πίνακας 6. Μετρικές αποτελεσμάτων για την περίπτωση VGG#1, 6 ^η Εποχή Εκπαίδευσης	79
Πίνακας 7. Αποτελέσματα της τροποποίησης VGG16#1 σε μορφή πίνακα σύγκρισης, μετά την 44 ^η εποχή εκπαίδευσης.....	79
Πίνακας 8. Μετρικές αποτελεσμάτων για την περίπτωση VGG#1 , μετά την 44 ^η εποχή εκπαίδευσης	80
Πίνακας 9. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#2	81
Πίνακας 10. Μετρικές αποτελεσμάτων για την περίπτωση VGG#2 , μετά την 61 ^η εποχή εκπαίδευσης	83
Πίνακας 11. Αποτελέσματα της τροποποίησης VGG16#2 σε μορφή πίνακα σύγκρισης, μετά την 61 ^η εποχή εκπαίδευσης	84
Πίνακας 12. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#3	84
Πίνακας 13. Μετρικές αποτελεσμάτων για την περίπτωση VGG#3 , μετά την 300 ^η εποχή εκπαίδευσης	87
Πίνακας 14. Αποτελέσματα της τροποποίησης VGG16#3 σε μορφή πίνακα σύγκρισης, μετά την 300 ^η εποχή εκπαίδευσης	87
Πίνακας 15. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#4	88
Πίνακας 16. Μετρικές αποτελεσμάτων για την περίπτωση VGG#4 , μετά την 43 ^η εποχή εκπαίδευσης	90

Πίνακας 17. Αποτελέσματα της τροποποίησης VGG16#4 σε μορφή πίνακα σύγκρισης, μετά την 43ή εποχή εκπαίδευσης	90
Πίνακας 18. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#5	91
Πίνακας 19. Μετρικές αποτελεσμάτων για την περίπτωση VGG#5 , μετά την 300ή εποχή εκπαίδευσης	93
Πίνακας 20. Αποτελέσματα της τροποποίησης VGG16#5 σε μορφή πίνακα σύγκρισης, μετά την 300ή εποχή εκπαίδευσης	93
Πίνακας 21. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#6	94
Πίνακας 22. Μετρικές αποτελεσμάτων για την περίπτωση VGG#6 , μετά την 416ή εποχή εκπαίδευσης	96
Πίνακας 23. Αποτελέσματα της τροποποίησης VGG16#6 σε μορφή πίνακα σύγκρισης, μετά την 200ή εποχή εκπαίδευσης	97
Πίνακας 24. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#7	97
Πίνακας 25. Μετρικές αποτελεσμάτων για την περίπτωση VGG#7 , μετά την 137ή εποχή εκπαίδευσης	99
Πίνακας 26. Αποτελέσματα της τροποποίησης VGG16#7 σε μορφή πίνακα σύγκρισης, μετά την 137ή εποχή εκπαίδευσης	100
Πίνακας 27. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#8	101
Πίνακας 28. Μετρικές αποτελεσμάτων για την περίπτωση VGG#8 , μετά την 379ή εποχή εκπαίδευσης	103
Πίνακας 29. Αποτελέσματα της τροποποίησης VGG16#8 σε μορφή πίνακα σύγκρισης, μετά την 379ή εποχή εκπαίδευσης	104
Πίνακας 30. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#9	104
Πίνακας 31. Μετρικές αποτελεσμάτων για την περίπτωση VGG#9 , μετά την 66ή εποχή εκπαίδευσης	107
Πίνακας 32. Αποτελέσματα της τροποποίησης VGG16#9 σε μορφή πίνακα σύγκρισης, μετά την 66ή εποχή εκπαίδευσης	108

1. ΕΙΣΑΓΩΓΗ

1.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη (TN) ή Artificial Intelligence (AI) στα Αγγλικά, είναι μία πολύ ευρεία έννοια που σχετίζεται με πολλές επιστήμες όπως για παράδειγμα η επιστήμη της πληροφορικής, της ψυχολογίας, της φιλοσοφίας, της νευρολογίας, την επιστήμη των μηχανών ακόμη και την επιστήμη της γλωσσολογίας προκειμένου να γίνει εφικτή η σύνθεση μιας τεχνητής συμπεριφοράς που αναδεικνύει σημάδια ευφυΐας . Στον τομέα της πληροφορικής προσπαθούμε να δημιουργήσουμε αλγόριθμους και συσκευές που είτε να μιμούνται την ανθρώπινη συμπεριφορά υποδηλώνοντας την παρουσία κάποιας στοιχειώδους ευφυΐας, είτε να αντλούν έμπνευση από άλλους οργανισμούς όπως οι γενετικοί αλγόριθμοι ή να προσπαθούμε να δημιουργήσουμε βραχίονες που να συμπεριφέρονται όπως το ανθρώπινο χέρι (haptics). Κάποια ακόμα παραδείγματα αποτελούν η εκμάθηση και προσαρμογή σε κάποιο νέο περιβάλλον, η εξαγωγή συμπερασμάτων, η προσέγγιση τιμών, η πρόβλεψη και γενικότερα η δυνατότητα επίλυσης πολύπλοκων προβλημάτων στα οποία πολλές φορές ούτε εμείς γνωρίζουμε τη βέλτιστη λύση. Ένας από τους επικρατούντες ορισμούς της τεχνητής νοημοσύνης οφείλεται στον Τζον Μακάρθι, μια εξέχουσα φυσιογνωμία στον τομέα της θεωρητικής πληροφορικής ο οποίος ορίζει τον συγκεκριμένο κλάδο ως «η επιστήμη και μεθοδολογία της δημιουργίας νοούντων μηχανών»[1].

Στην TN ο σκοπός μας μπορεί να συνοψιστεί ως η εκμάθηση και προσαρμογή στο εκάστοτε περιβάλλον, υπολογιστών και μηχανών συνήθως για να πετύχουν ένα συγκεκριμένο σκοπό. Ανάλογα με τον επιθυμητό στόχο η TN μπορεί να χωριστεί σε ένα ευρύτερο σύνολο τομέων, παραδείγματος χάριν στην μηχανική εκμάθηση, τα συστήματα γνώσης, την προσέγγιση συναρτήσεων, την προσομοίωση της ανθρώπινης συμπεριφοράς μεταξύ πολλών άλλων. Σε μεγάλο αριθμό περιπτώσεων υπάρχει επικάλυψη με συναφή επιστημονικά πεδία όπως π.χ. την υπολογιστική όραση, σύνθεση εικόνας, σύνθεση και αναγνώριση φυσικής γλώσσας και ρομποτική, τα οποία μπορούν να θεωρηθούν πλέον ως ανεξάρτητα πεδία της σύγχρονης TN .

Ερευνητές του τομέα της TN έχουν ως σκοπό την κατασκευή λογισμικού και μηχανών ικανών να επιλύουν πραγματικά προβλήματα διαφόρων τύπων («ασθενής TN» [2], παρούσα κατάσταση), ενώ αρκετοί αποβλέπουν στην προσομοίωση της γενικευμένης ευφυΐας, την αποκαλούμενη και «ισχυρή TN» [2].

1.2 Μηχανική Μάθηση – Γενικά στοιχεία

Η μηχανική μάθηση (machine learning) αποτελεί έναν από τους σημαντικότερους πυλώνες που στηρίζεται η Τεχνητή Νοημοσύνη (TN). Ως σκοπό έχει την ανάπτυξη αλγορίθμων κατάλληλων, που θα δώσουν τη δυνατότητα της “εκμάθησης” στους υπολογιστές. Το λογισμικό που χρησιμοποιείται από τους υπολογιστές γίνεται πλέον ευέλικτο και προσαρμόσιμο σε πολύ μεγάλο εύρος προβλημάτων, με βάση την ανάλυση των δεδομένων που λαμβάνουν, αντί της κλασικής πλέον προσαρμογής τους με βάση την διαίσθηση του ερευνητή ή του μηχανικού λογισμικού που προγραμματίζει κάποιο σύστημα και προσπαθεί να βρει ουσιώδη χαρακτηριστικά που διακρίνουν τις εκάστοτε περιπτώσεις μεταξύ τους με τον καλύτερο δυνατό τρόπο. Όλη η ουσία της μηχανικής μάθησης συνοψίζεται στη χρήση αλγορίθμων, ικανών να αναγνωρίζουν μοτίβα στα δεδομένα προκειμένου να λάβουν αποφάσεις, βασιζόμενες στην στατιστική, τη θεωρία των πιθανοτήτων και την βελτιστοποίηση. Αποτέλεσμα της εκμάθησης μηχανών μπορεί να είναι υπηρεσίες όπως φίλτρα ανεπιθύμητης αλληλογραφίας, αναγνώριση κειμένου και φωνής, αξιόπιστες μηχανές αναζήτησης στο διαδίκτυο, αυτό-οδηγούμενα μέσα μεταφοράς, αξιόπιστες διαγνώσεις όπως στην παρούσα εργασία και άλλα.

Οι αλγόριθμοι TN χωρίζονται στη βιβλιογραφία [3] με βάση το κατά πόσο γνωρίζουμε εκ των προτέρων το αποτέλεσμα που θέλουμε για τα δοθέντα παραδείγματα στις τρεις παρακάτω κατηγορίες:

1. Επιτηρούμενη μάθηση (supervised learning), όπου ο αλγόριθμος με βάση αντίστοιχων κριτηρίων δημιουργεί μια συνάρτηση που απεικονίζει δεδομένες εισόδους (labeled examples) σε γνωστές εξόδους σύνολο εκπαίδευσης (training set), κάνοντας προβλέψεις και διορθώνοντας τις προβλέψεις σε περίπτωση λάθους, με τελικό στόχο τη γενίκευση της συνάρτησης αυτής για άγνωστες εισόδους που έχουν αντίστοιχα άγνωστη έξοδο (σύνολο ελέγχου).
2. Ημι-επιτηρούμενη μάθηση (semi-supervised learning), στην οποία τα δεδομένα εκπαίδευσης είναι συνδυασμός γνωστών και αγνώστων δειγμάτων στα οποία θέλουμε να εκτιμήσουμε ένα επιθυμητό αποτέλεσμα. Σε αυτή την περίπτωση το μοντέλο πρέπει να μάθει δομές από τα ίδια τα δεδομένα για να τα χωρίσει σε κατηγορίες και να κάνει προβλέψεις για το σε ποια κατηγορία ανήκουν τα άγνωστα ακόμα στοιχεία εισόδου. Στην παρούσα εργασία δεν χρησιμοποιήθηκε αυτό το μοντέλο, αν και χρησιμοποιείται συχνά στην πράξη [4] σε ιατρικά σύνολα

δεδομένων μιας και δεν είναι πάντα εύκολο (η οικονομικά εφικτό) να υπάρχει η τελική διάγνωση από ειδικό για κάθε δεδομένο εισόδου.

3. Μη Επιτηρούμενη μάθηση (unsupervised learning), όπου ο αλγόριθμος δημιουργεί ένα μοντέλο για ένα σύνολο εισόδων δίχως να γνωρίζει a priori τις επιθυμητές εξόδους (unlabeled examples) για το σύνολο εκπαίδευσης, προσπαθώντας να εξερευνήσει την ίδια την δομή των δεδομένων.

Στην παρούσα εργασία ασχολούμαστε με τη χρήση επιτηρούμενης μάθησης μιας και γνωρίζουμε για κάθε υπό-εικόνα που μας δίνεται τη διάγνωση που έχει γίνει για το συγκεκριμένο κομμάτι εισόδου από εξειδικευμένο ιατρό.

1.3 Νευρωνικά Δίκτυα – Γενικά στοιχεία

Τα Νευρωνικά Δίκτυα (ΝΔ) (Neural Networks), μπορούν να χωριστούν σε βιολογικά και τεχνητά ΝΔ. Τα τεχνητά ΝΔ αρχικά άντλησαν έμπνευση για τον τρόπο λειτουργίας τους από τα αντίστοιχα βιολογικά. Ο πρωταρχικός λόγος που οι ερευνητές του τομέα των ΝΔ [5] κάνουν τον παραλληλισμό ήταν για να βρουν εξηγήσεις για το πως δουλεύει ο ανθρώπινος εγκέφαλος μιας και ήδη γνώριζαν ότι είναι ένα σύστημα στο οποίο δουλεύει η νοημοσύνη. Τα Βιολογικά Νευρωνικά Δίκτυα (ΒΝΔ) αποτελούν κομμάτι του νευρικού συστήματος όλων των θηλαστικών καθώς και πολλών άλλων μορφών ζωής. Σε αυτό το σημείο μπορούμε να διακρίνουμε διαφορετικά και αρκετά πολύπλοκα συστήματα όπως ηλεκτρικά συστήματα, χημικά συστήματα, διαφορετικά είδη αισθητήρων κ.ά. τα οποία ήταν τεράστια πρόκληση να γίνει εφικτό να μπορέσουν να μοντελοποιηθούν και να παράξουν αντίστοιχα αποτελέσματα απλά με τη χρήση απλών πράξεων.

1.3.1 Βιολογικοί Νευρώνες

Κάθε κύτταρο αποτελεί μια στοιχειώδη μονάδα μεταφοράς και επεξεργασίας ηλεκτρικών σημάτων και χημικών διαδικασιών που ονομάζεται “νευρώνας”. Ο ανθρώπινος εγκέφαλος σύμφωνα με εκτιμήσεις περιλαμβάνει περίπου $2 \cdot 10^{10}$ νευρώνες [6] με διάφορα σχήματα, δομή, λειτουργικότητα και μέγεθος το οποίο εκτιμάται σε περίπου 100μm.

Στις ομοιότητες των νευρικών κυττάρων με τα υπόλοιπα περιλαμβάνονται:

1. Περιβάλλονται από κυτταρική μεμβράνη
2. Έχουν πυρήνα που περιλαμβάνει γονίδια

3.Περιέχουν κυτταρόπλασμα, μιτοχόνδρια και άλλα αντίστοιχα οργάνδια

4.Πραγματοποιούν βασικές κυτταρικές διεργασίες όπως πρωτεϊνική σύνθεση και παραγωγή ενέργειας

Διαφέρουν όμως στα κάτωθι:

1. Οι νευρώνες επικοινωνούν μεταξύ τους μέσω ηλεκτρικών και χημικών σημάτων.
2. Έχουν ειδικευμένα μέρη που ονομάζονται «δενδρίτες» μεταφέρουν ηλεκτρικά σήματα στον νευρώνα και τους «άξονες» οι οποίοι μεταφέρουν σήματα από τον νευρώνα προς άλλα μέρη του εγκεφάλου.
3. Περιλαμβάνουν δομές που δεν εντοπίζονται σε άλλα είδη, όπως τις “συνάψεις” και εκκρίνουν χημικά όπως οι «νευροδιαβιβαστές»



Εικόνα 1-1. Βιολογικός Νευρώνας και τα κύρια στοιχεία του

Οι συνάψεις ενώνουν τους νευρώνες και δίνουν την δυνατότητα ανταλλαγής του ηλεκτρικού δυναμικού μεταξύ των δενδριτών. Οι ανταλλαγές αυτές συμβαίνουν με σχετικά αργό ρυθμό της τάξης του 10^2 Hz. Με τον μέσο ενήλικο να έχει της τάξης του 10^{13} συνάψεις θα μπορούσαμε να εκτιμήσουμε την ταχύτητα επεξεργασίας του ανθρώπινου εγκεφάλου στα 10^{15} Hz. Συγκρίνοντας τα παραπάνω νούμερα με έναν τυπικό επεξεργαστή του σήμερα που βρίσκεται στα $3 \cdot 10^9$ παρατηρούμε ότι ο άνθρωπος

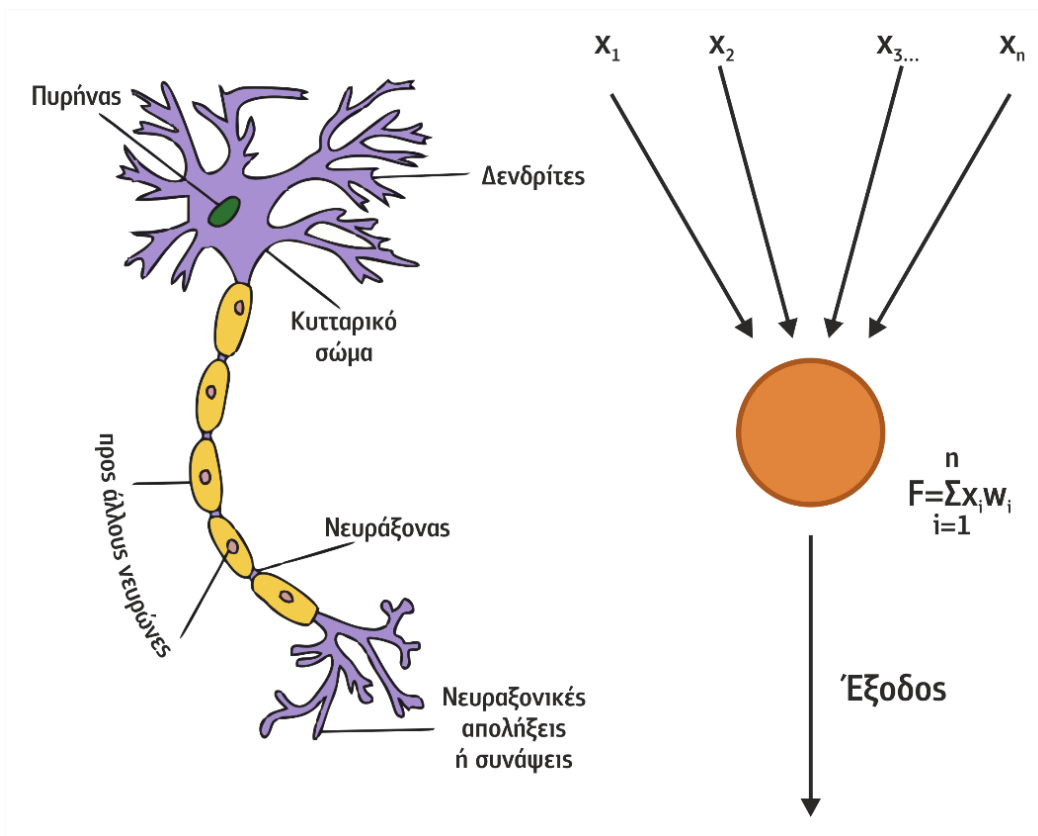
εγκέφαλος υπερτερεί σημαντικά, σε επεξεργαστική ισχύ, με τη διαφορά ότι δεν ενεργοποιείται ποτέ ολόκληρη η επεξεργαστική ισχύς του εγκεφάλου ταυτόχρονα.

1.3.2 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) μοντελοποιήθηκαν αρχικά ώστε να προσομοιάζουν τη λειτουργία του ανθρώπινου εγκεφάλου [7][8] υιοθετώντας την βασική ιδέα στην αρχιτεκτονική του, ενώ η εκπαίδευσή τους γίνεται με την ίδια λογική και στα ΒΝΔ, δηλαδή μέσα από ένα σύνολο παραδειγμάτων [9]. Όπως σε ένα παιδί με τα παραδείγματα από την πραγματική ζωή προσπαθεί να αντιληφθεί τον κόσμο το αντίστοιχο προσπαθούμε να κάνουμε πολύ πιο στοχευμένα στα ΤΝΔ Έτσι, τα κοινά χαρακτηριστικά των δυο κατηγοριών νοημοσύνης μπορούν να θεωρηθούν τα:

1. Τα σημεία συνάψεων των ΒΝΔ, ανάλογα με την ισχύ των δεσμών τους όπως και τα συναπτικά βάρη στα ΤΝΔ, χρησιμοποιούνται για την αποθήκευση της γνώσης που έχει αποκτηθεί.
2. Η εκπαίδευσή τους γίνεται μέσα από ένα σύνολο παραδειγμάτων, βοηθώντας τα να μάθουν το περιβάλλον τους.

Θα αναφερθούμε αναλυτικά και στο επόμενο κεφάλαιο όπου και θα εξηγηθούν και οι βασικοί αλγόριθμοι εκπαίδευσης των δικτύων που χρησιμοποιήθηκαν στην παρούσα εργασία, αλλά συνοπτικά, τα ΤΝΔ και τα ΒΝΔ αποτελούνται από ένα σύνολο νευρώνων, δηλαδή μιας μονάδας αποθήκευσης και επεξεργασίας πληροφορίας, τα τρία στοιχεία των οποίων είναι ένας αριθμός συνάψεων, ένας αθροιστής και μια συνάρτηση μεταφοράς (ενεργοποίησης) του νευρώνα. Ένας νευρώνας, μπορεί να έχει ένα μεγάλο αριθμό εισόδων αλλά η έξοδος του θα είναι πάντοτε μια, η οποία συνήθως αποτελεί είσοδο σε κάποιο άλλο νευρώνα, δημιουργώντας έτσι ένα δίκτυο από νευρώνες. Τα βάρη σε κάθε νευρώνα αντιπροσωπεύουν την χρήση που γίνεται σε αυτή τη σύναψη από τον νευρώνα, ενώ η συνάρτηση ενεργοποίησης καθορίζει την έξοδο σχετικά με τις σταθμισμένες εισόδους του νευρώνα. Στην παρακάτω εικόνα, παρατηρούμε σχηματικά τα ανωτέρω.



Εικόνα 1-2. Αντιδιαστολή βιολογικού με τεχνητό νευρώνα

2. ΤΕΧΝΙΚΕΣ ΕΚΜΑΘΗΣΗΣ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Σκοπός μας σε αυτό το κεφάλαιο είναι να παρουσιάσουμε τα απαραίτητα στοιχεία από τον κόσμο της μηχανικής μάθησης ώστε να γίνουν πλήρως κατανοητά τα κίνητρα, η θεωρία και η μεθοδολογία που χρησιμοποιήθηκε στην παρούσα εργασία. Να προσπαθήσουμε να κάνουμε κατανοητές τις τεχνικές με τις οποίους έχουμε προσπαθήσει να αντιμετωπίσουμε το πρόβλημα της διδασκαλίας του υπολογιστή από παραδείγματα. Θα αρχίσουμε με μερικούς βασικούς αλγόριθμους που έχουν χρησιμοποιηθεί αλλά που δίχως αυτή τη γνώση θα ήταν αδύνατο να περιγραφεί η λογική και η μορφή που παίρνουν οι συναρτήσεις που θέλουμε να εκτιμήσει το σύστημα μηχανικής μάθησης καθώς επίσης το κίνητρο για τις διάφορες συναρτήσεις και αρχιτεκτονικές που χρησιμοποιήθηκαν στα συνελκτικά δίκτυα.

2.1 Γραμμική παλινδρόμηση

Η γραμμική παλινδρόμηση αποτελεί μια απλή μορφή μηχανικής μάθησης με βάση την οποία σκοπό έχουμε να εισάγουμε την απαραίτητη ορολογία καθώς επίσης και τη δομή που έχουν αρκετά προβλήματα στον τομέα της επιτηρούμενης μάθησης. Ουσιαστικός σκοπός της τεχνικής της γραμμικής παλινδρόμησης (linear regression) [3] είναι η πρόβλεψη της τιμής του y δεδομένου κάποιου διανύσματος – εισόδου $x \in R^n$. Υποθέτουμε ότι μας δίνεται μεγάλος αριθμός από στιγμιότυπα του προβλήματος και μία τιμή που πρέπει να προβλέψουμε την οποία και γνωρίζουμε για κάθε στιγμιότυπο, όπου τα χαρακτηριστικά του i -οστού δείγματος συμβολίζονται με $x^{(i)}$ και τη γνωστή απάντηση ως $y^{(i)}$. Τελικός στόχος μας σε αυτό το σενάριο είναι να βρούμε μια συνάρτηση $h(x)$ τέτοια ώστε $y^{(i)} = h(x^{(i)})$ ή έστω να έχει μικρή απόκλιση. Επίσης θέλουμε αυτή η συνάρτηση για τιμές που δεν έχει δει να εκτιμά όσο το δυνατόν καλύτερο το $y^{(i)}$, δηλαδή να μπορεί να γενικεύει όπως συχνά αναφέρεται στην ορολογία της μηχανικής μάθησης.

Το να αποφασίσουμε τη μορφή της συνάρτησης που θα έχει η συνάρτηση $h(x)$ από μόνο του περιορίζει το πλήθος και τη μορφή των δυνατών λύσεων (που το ονομάζουμε “συναρτήσεις υπόθεσης”) που μπορεί να μας προκύψουν από το σύνολο όλων των συναρτήσεων. Αυτό μας οδηγεί στο πρόβλημα του διλήμματος προκατάληψης-διακύμανσης “bias-variance dilemma” [10] το οποίο πολύ απλουστευτικά μας λέει ότι όσο περιορίζουμε το πλήθος των δυνατών λύσεων και μικραίνουμε τη διακύμανση εισάγουμε προκατάληψη στη μορφή που θέλουμε να έχει η τελική μας λύση. Το παραπάνω είναι

ορθό να το κάνουμε όταν έχουμε την εμπειρία ή/και γνώση για την πραγματική μορφή που θέλουμε να έχει η συνάρτηση που προσπαθούμε να εκτιμήσουμε.

Επιστρέφοντας στο πρόβλημα μας, αρχικά στη γραμμική παλινδρόμηση υποθέτουμε (το πολύ περιοριστικό για τις δυνατές μορφές τις εξόδου μας) ότι η συνάρτηση εξόδου είναι μια γραμμική συνάρτηση των εισόδων μας. Σε αυτή τη συνάρτηση αυτό που ψάχνουμε είναι οι συντελεστές με τους οποίους θα πολλαπλασιάσουμε όλες τις εισόδους μας ώστε να υπολογίσουμε μια από αυτές τις συναρτήσεις $h(x, \theta) = \sum_i \theta_i x_i$. Σε αυτό το σημείο θα πρέπει να αποφασίσουμε με ποιο κριτήριο θα αποφασίσουμε ποιο από όλα τα θ μου δίνει μια $h(x, \theta)$ που να είναι η «καλύτερη δυνατή». Αυτό που ψάχνουμε είναι να ορίσουμε μια «συνάρτηση κόστους» (cost function) [11] που να υπολογίζει το σφάλμα που κάνει η συνάρτηση για κάθε είσοδο που γνωρίζουμε την έξοδο, δηλαδή για το σύνολο εκπαίδευσης της συνάρτησης μας (για το οποίο θα μιλήσουμε στη συνέχεια). Μια τυπική συνάρτηση κόστους είναι η ελαχιστοποίηση τετραγώνων σφάλματος (minimum square error) η οποία για το παραπάνω μοντέλο παίρνει την παρακάτω μορφή

$$J(\theta) = \frac{1}{2} \sum_l (h_{\theta} * x^{(l)} - y^{(l)})^2 = \frac{1}{2} \sum_l (\theta^{(l)} * x^{(l)} - y^{(l)})^2$$

2.2 Ελαχιστοποίηση συνάρτησης

Πλέον έχουμε ορίσει τη μορφή που έχουν οι συναρτήσεις που θέλουμε να προσεγγίσουν τα δεδομένα μας, και σκοπός μας είναι να επιλέξουμε ποια από αυτές «ταιριάζει» περισσότερο στα δεδομένα μας. Για να το κάνουμε αυτό θα επιλέξουμε εκείνη τη συνάρτηση που ελαχιστοποιεί τη συνάρτηση κόστους που έχουμε ορίσει για το πρόβλημα μας, δηλαδή να βρούμε κατάλληλο θ , το οποίο να ελαχιστοποιεί την $J(\theta)$. Συνήθως οι αλγόριθμοι ελαχιστοποίησης συναρτήσεων απαιτούν δυο πληροφορίες για την συνάρτηση $J(\theta)$: τον υπολογισμό της τιμής της σε οποιαδήποτε είσοδο θ και της κλίσης της $\nabla_{\theta} J(\theta)$ για όλες τις δυνατές τιμές του θ . Ανάλογα με τη φύση της ελαχιστοποίησης θα έχουμε ένα ολικό ή μερικό μέγιστο ή μια ικανοποιητική προσέγγιση. Αν εφαρμοστεί ο αλγόριθμος απληστίας (greedy algorithm) τότε μετακινούμαστε προς την κατεύθυνση που μας δείχνει η κλίση ότι η τιμή της συνάρτησης είναι μικρότερη, στο σημείο που βρισκόμαστε την κάθε στιγμή. Η κλίση μιας παραγωγίσιμης συνάρτησης, είναι ένα διάνυσμα που δείχνει στην κατεύθυνση της πιο απότομης μεταβολής ως προς το θ . Στο πρόβλημα της γραμμικής παλινδρόμησης για παράδειγμα με σχετικά απλούς αλγεβρικούς υπολογισμούς μπορούμε να υπολογίσουμε την παρακάτω έκφραση:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_l x_j^{(l)} (h_{\theta}(x^{(l)}) - y^{(l)})$$

για το διάνυσμα κόστους. Επέκταση της γραμμικής παλινδρόμησης είναι η λογιστική όπου πλέον παύουν οι εισοδοί να είναι γραμμικά εξαρτώμενες πράγμα που ξεφεύγει από τα όρια ανάλυσης της παρούσας εργασίας.

2.2.1 Έλεγχος κλίσης

Για τον υπολογισμό της κλίσης μίας συνάρτησης δεν είναι πάντα προφανές πως υπολογίζεται. Η πιθανότητα λογικών λαθών και σφαλμάτων υπολογισμού όπως υπερ(υπο)χείλισης κατά τη διάρκεια των υπολογισμών ενυπάρχει πάντοτε. Σε «εύκολες περιπτώσεις» όπως την παραπάνω περίπτωση που είναι τετραγωνική έχει γίνει μεγάλη ανάλυση στη βιβλιογραφία [12] και ο τύπος που προκύπτει είναι σχετικά απλός μιας και το πλήθος των παραμέτρων ταυτίζεται (είναι απλά μια παράμετρος παραπάνω για τον σταθερό όρο) με το πλήθος των αγνώστων. Το γεγονός ότι ο τύπος είναι απλός για την παράγωγο σε κάθε σημείο, δυστυχώς δεν μας εξασφαλίζει και ένα ολικό ελάχιστο για τη συνάρτηση ενδιαφέροντος. Γενικά σε οποιαδήποτε παραγωγίσιμη συνάρτηση μπορούμε να υπολογίσουμε την κλίση της (για μεγαλύτερη σταθερότητα) ως :

$$\frac{\partial J(\theta)}{\partial \theta} = \lim_{\varepsilon \rightarrow 0} \frac{J(\theta + \varepsilon) - J(\theta - \varepsilon)}{2\varepsilon}$$

Επιλέγουμε συνήθως τον παραπάνω τύπο για βελτίωση της ευστάθειας του υπολογισμού. Η επιλογή του παράγοντα ε για τον αριθμητικό υπολογισμό πρέπει να είναι τέτοια ώστε στη γλώσσα προγραμματισμού που χρησιμοποιούμε τα σφάλματα να μην υπερκεράσουν τη διαφορά της συνάρτησης για αυτό νούμερα της τάξης του 10^{-8} δουλεύουν καλά για πρακτικούς υπολογισμούς (ανάλογα βέβαια με την ακρίβεια των αριθμών που έχουν αποθηκευτεί στον υπολογιστή).

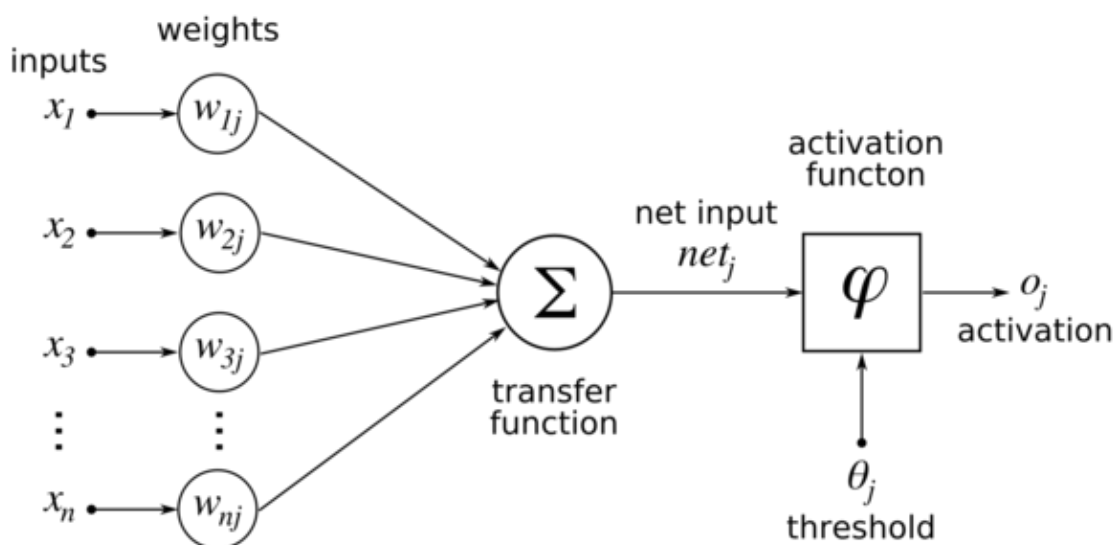
2.3 Επιβλεπόμενη μάθηση

Όπως συμβαίνει και στον άνθρωπο, προκειμένου να λειτουργήσει ορθά ένα ΝΔ, πρέπει να εκπαιδευτεί. Η εκπαίδευση συνίσταται στον προσδιορισμό των κατάλληλων συντελεστών βάρους μεταξύ των συνάψεων του ΝΔ και πραγματοποιείται με τη βοήθεια αλγορίθμων, κατάλληλων για την εκμάθηση του περιβάλλοντος για το εκάστοτε πρόβλημα και την βελτίωση της απόδοσής του. Η εκμάθηση ενός ΝΔ, γίνεται σταδιακά και σύμφωνα με καθορισμένους κανόνες. Μια επαναληπτική διαδικασία ρυθμίζει τα βάρη

στις συνάψεις μεταξύ των επιπέδων κι έτσι το ΝΔ αποκτά περισσότερη “γνώση”. Οι αλγόριθμοι που χρησιμοποιήθηκαν για την εκπαίδευση ενός ΝΔ αναλύονται στη συνέχεια.

2.3.1 Βασικές δομές νευρωνικών δικτύων

Ένα νευρωνικό δίκτυο αποτελείται από πολλούς νευρώνες συνδεδεμένους ώστε να επιτελούν πράξεις ώστε η έξοδος του ενός να αποτελεί είσοδο σε κάποιον άλλο (εκτός προφανώς από το τελευταίο επίπεδο) όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 2-1. Δομικά στοιχεία υπολογιστικού νευρώνα

Τα θ_j ονομάζονται κατώφλι (threshold) ή κατ’ άλλους αποκαλούνται και bias units και αντιστοιχούν σε βοηθητικούς όρους στη διαδικασία του να μην είναι γραμμική η συνάρτηση και να μπορεί να μετατοπιστεί από την αρχή των αξόνων. Σε επόμενα κεφάλαια θα επεκταθούμε στην εξήγηση τους. Το πρώτο επίπεδο (layer) (που βρίσκεται αριστερά) του δικτύου ονομάζεται επίπεδο εισόδου (input layer) και το τελευταίο επίπεδο (στα δεξιά) ονομάζεται επίπεδο εξόδου (output layer). Τα ενδιάμεσα επίπεδα ονομάζονται κρυφά επίπεδα (hidden layers) επειδή οι τιμές δεν εμφανίζονται στον παρατηρητή, απλά βοηθούν στον υπολογισμό της εξόδου. Επίσης να τονίσουμε ότι στο παράδειγμα που πραγματευόμαστε έχουμε n νευρώνες-εισόδου (δεν μετράμε το κατώφλι), ένα κρυφό νευρώνα και ένα νευρώνα εξόδου.

Κάθε νευρωνικό δίκτυο έχει ένα πλήθος από επίπεδα στα οποία η είσοδος τους μεταφέρεται σε κάθε επόμενο. Έστω ότι το πλήθος των επιπέδων του δικτύου μας είναι n_l . Αν συμβολίσουμε με L το κάθε επίπεδο τότε το L_1 είναι το πρώτο επίπεδο και L_{n_l} το τελευταίο επίπεδο ή επίπεδο εξόδου. Για κάθε επίπεδο εκτός του τελευταίου πρέπει να

έχουμε τις κάτωθι πληροφορίες: το βάρος κάθε νευρώνα κάθε επιπέδου $W^{(l)}$ καθώς και το κατώφλι κάθε νευρώνα επιπέδου θ_j . Το $W_{ij}^{(l)}$ χαρακτηρίζοντας την τιμή που συνδέει την έξοδο του νευρώνα i στον νευρώνα j στο επόμενο επίπεδο $l+1$. Επίσης με $\theta_i^{(l)}$ χαρακτηρίζουμε το αντίστοιχο κατώφλι για τον i νευρώνα στο επίπεδο $l+1$. Να διασαφηνισθεί εδώ ότι το κατώφλι του κάθε νευρώνα δεν έχει εισόδους σχετικά με τον νευρώνα και το επίπεδο. Στο παράδειγμά μας, έχουμε και τα κατώφλια (bias terms) που δεν έχουν εισόδους, καθώς πάντα θεωρούμε τη μονάδα ως είσοδο. Για αυτό το λόγο μπορούμε ακόμα και να τα ενσωματώσουμε με τα βάρη ως w_0 επεκτείνοντάς και όλες τις εισόδους μας με το 1 στην πρώτη θέση. Με n_i συμβολίζουμε τον αριθμό των νευρώνων στο επίπεδο (τα bias terms δεν μετρούνται). Γράφουμε $a_i^{(l)}$ για να χαρακτηρίσουμε τις ενεργοποιήσεις (activations) του νευρώνα στο επίπεδο. Δεδομένων των W και b παραμέτρων για κάθε επίπεδο, το ΝΔ προσδιορίζει την διαχωριστική συνάρτηση (hypothesis function) [13] η οποία εξάγει έναν πραγματικό αριθμό ή ένα διάνυσμα αριθμών που αντιστοιχούν στις κλάσεις εξόδου τις οποίες σκοπεύουμε να διαχωρίσουμε.

Αν συμβολίσουμε με $z_i^{(l)}$ το συνολικό αλλά και σταθμισμένο άθροισμα των εισόδων του i νευρώνα στο επίπεδο l , η εξίσωση που θα περιγράψει την έξοδο είναι η :

$$z_i^{(l)} = \sum_{j=1}^n W_{ij}^{(l-1)} x_j + b_i^{(l-1)}$$

Όταν το παραπάνω εφαρμοστεί από το αρχικό επίπεδο μέχρι και το τελευταίο τότε παίρνουμε την τελική έξοδο της συνάρτησης. Η παραπάνω εφαρμογή του τύπου για μια συγκεκριμένη είσοδο ονομάζεται *προς τα εμπρός διάδοση του σήματος* (forward propagation). Στο παραπάνω υποθέτουμε ότι σε κάθε επίπεδο συνδέονται όλοι οι κόμβοι του i επιπέδου με το $i+1$ επίπεδο. Το παραπάνω είναι ένα από τα βασικά στοιχεία που αλλάζει στα συνελκτικά δίκτυα.

2.4 Νευρωνικά Δίκτυα πολλαπλών επιπέδων

Έχει αποδειχτεί ότι όλες οι συνεχείς συναρτήσεις μπορούν να αναπαρασταθούν από ένα νευρωνικό δίκτυο με 2 κρυφά επίπεδα [14] με οσοδήποτε μικρό σφάλμα είναι επιθυμητό. Παρόλα αυτά όμως τα τελευταία χρόνια τα δίκτυα συνεχίζουν να έχουν όλο και περισσότερα επίπεδα [15]. Το οποίο δημιουργεί μία εύλογη απορία, για το λόγο για τον οποίο συμβαίνει. Μια αρχική διαισθητική απάντηση κατ' αναλογία είναι ότι για να υπολογίσουμε την δύναμη 2 αριθμών μπορούμε να το κάνουμε με διαδοχικούς πολ/μούς

τους οποίους μπορούμε να τους κάνουμε με διαδοχικές προσθέσεις. Παρόλα αυτά είναι αρκετά πιο δύσκολο και χρονοβόρο να υλοποιηθεί με αυτόν τον τρόπο. Επίσης δεν εκμεταλλευόμαστε την ίδια τη δομή του προβλήματος. Αντίστοιχα, μια ισοδύναμη συνάρτηση με 2 μόνο επίπεδα κρυφών νευρώνων θα απαιτούσε υπέρ-πολλαπλάσιο αριθμό νευρώνων και δεν θα μας επέτρεπε την «κριτική ματιά» για να καταλάβουμε για ποιο λόγο το κάθε βάρος έχει την τιμή που έχει. Στην περίπτωση των πολλαπλών νευρώνων μπορούμε να κατηγοριοποιήσουμε τα επίπεδα σε γενικές έννοιες (αφορά κυρίως τα συνελκτικά ΝΔ) ανάλογα με το τι πραγματοποιούν, όπως εύρεση ακμών ή εύρεση υφής (texture) [16] τουλάχιστον για τα πρώτα επίπεδα του δικτύου κάτι το οποίο μπορεί να μας οδηγήσει σε περαιτέρω βελτιώσεις. Σε αυτό το κεφάλαιο θα μιλήσουμε και για τα είδη των συναρτήσεων ενεργοποίησης που εφαρμόζονται συνήθως και τους λόγους που μπορούμε να επιλέξουμε την κάθε μία.

Για διευκόλυνση θα υποθέσουμε έναν νευρώνα με 3 μόνο εισόδους x_1, x_2, x_3 και αντίστοιχα το bias term. Σε αυτή τη περίπτωση η συνάρτηση εξόδου μπορεί να πάρει την παρακάτω μορφή:

$$h_{w,b}(x) = f(W^T x) = f\left(\sum_{i=1}^3 W_i x_i + b\right)$$

Η έξοδος του νευρώνα αυτού που συμβολίζεται με $h_{w,b}(x)$ χρησιμοποιεί την συνάρτηση f που ονομάζεται συνάρτηση ενεργοποίησης (activation function). Μια συχνή επιλογή για αυτή τη συνάρτηση είναι η σιγμοειδής συνάρτηση αν και υπάρχουν πλέον και άλλες δημοφιλείς επιλογές. Η σιγμοειδής συνάρτηση έχει τον παρακάτω τύπο :

$$f(z) = \frac{1}{1 + \exp(-z)}$$

Η παραπάνω συνάρτηση αντιστοιχεί το $\mathbb{R} \rightarrow \mathbb{R}$. Είναι μη γραμμική το οποίο είναι απαραίτητη προϋπόθεση για οποιαδήποτε συνάρτηση επιλεγεί, αλλιώς όλα τα «επίπεδα» που θα προσθέταμε θα μπορούσαν να αναχθούν σε ένα και μόνο επίπεδο.

Μια άλλη επιλογή είναι η υπερβολική εφαπτομένη η οποία έχει εφαρμοστεί σε μεγάλο μέρος της βιβλιογραφίας και έχει τον ακόλουθο τύπο:

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Τα τελευταία χρόνια έχει κερδίσει έδαφος σε σχέση με τις προηγούμενες δυο συναρτήσεις η «Διορθωμένη Γραμμική Συνάρτηση» (REctified Linear Function) και ολόκληρη η μονάδα που εφαρμόζει αυτή τη συνάρτηση ονομάζεται «Διορθωμένη Γραμμική Μονάδα» (REctified Linear Unit ή *RELU*) [16] της οποίας ο τύπος υπολογισμού φαίνεται παρακάτω.

$$f(z) = \max(0, z)$$

Η συγκεκριμένη συνάρτηση έχει το μειονέκτημα του να μην είναι παραγωγίσιμη σε όλο το πεδίο ορισμού της μιας και στο 0 δεν παραγωγίζεται. Αυτό όμως στην πραγματικότητα δεν είναι πρόβλημα μιας και μπορούμε να θεωρήσουμε ως παράγωγο ή το 0 ή το 1 χωρίς κάποιο ουσιαστικό πρόβλημα.

Ο βασικότερος λόγος που χρησιμοποιείται η παραπάνω συνάρτηση είναι ότι εξαλείφει το πρόβλημα της εξαφανιζόμενης παραγώγου [15]. Σε αυτό το πρόβλημα θα αναφερθούμε στην επόμενη ενότητα όταν θα έχει γίνει αναφορά στον αλγόριθμο οπισθοδιάδοσης σφάλματος (backpropagation algorithm).

2.5 Αλγόριθμος Οπισθοδιάδοσης Σφάλματος (Backpropagation Algorithm)

Ο αλγόριθμος backpropagation επέτρεψε να επιτευχθεί ένας συστηματικός τρόπος για την εκπαίδευση ενός νευρωνικού δικτύου το οποίο περιέχει περισσότερα από ένα κρυφά επίπεδα νευρώνων. Ουσιαστικά αυτό που επιτυγχάνουμε με την εφαρμογή του, είναι να καταμερίζουμε το σφάλμα που μπορούμε να μετρήσουμε στην έξοδο του δικτύου μας, σε κάθε ενδιάμεσο νευρώνα, και να μεταβάλουμε τα βάρη του δικτύου με σκοπό να ελαττωθεί αυτό το σφάλμα για το αντίστοιχο παράδειγμα εισόδου. Στη συνέχεια αναλύεται η λειτουργία του.

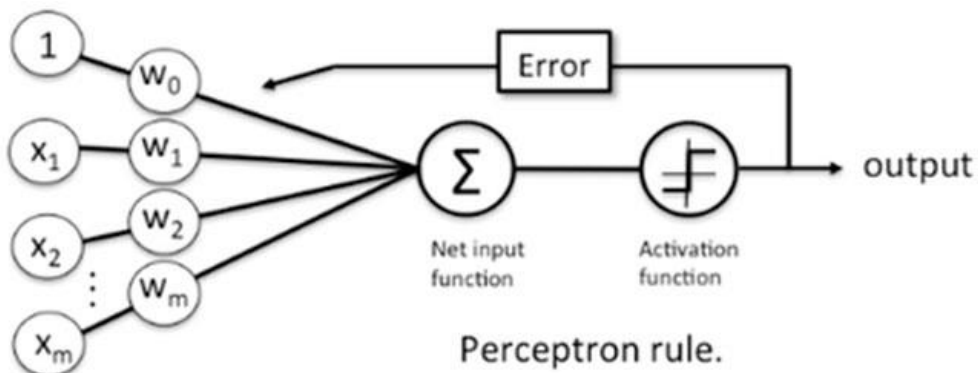
Έστω το σύνολο εκπαίδευσης που έχουμε αποτελείται από τα δείγματα $\{x^{(i)}, y^{(i)}\}$. Αρχικά θα υποθέσουμε ότι χρησιμοποιούμε την τεχνική της batch gradient descent (ελαχιστοποίηση με βάση την παράγωγο συνόλου), στην οποία παρατηρούμε όλα τα δείγματα και της αλλαγές που επιφέρουν στο δίκτυο και τις εφαρμόζουμε όλες μαζί αφού το δίκτυο έχει δει όλες τις εισόδους από το σύνολο εκπαίδευσης. Αυτή η τεχνική έχει το πλεονέκτημα ότι δεν επηρεάζεται το αποτέλεσμα με τη σειρά με την οποία δείχνουμε το κάθε δείγμα εισόδου στο δίκτυο. Αν αναλύθεί περισσότερο τη διαδικασία, έστω για ένα δείγμα εκπαίδευσης, προσδιορίζεται η συνάρτηση κόστους η οποία εξαρτάται από τα

βάρη και τα κατώφλια του δικτύου, την είσοδο και την έξοδο του δείγματος και παίρνει την μορφή:

$$J(W, b; x, y) = \frac{1}{2} \|h_{W,b}(x) - y\|^2$$

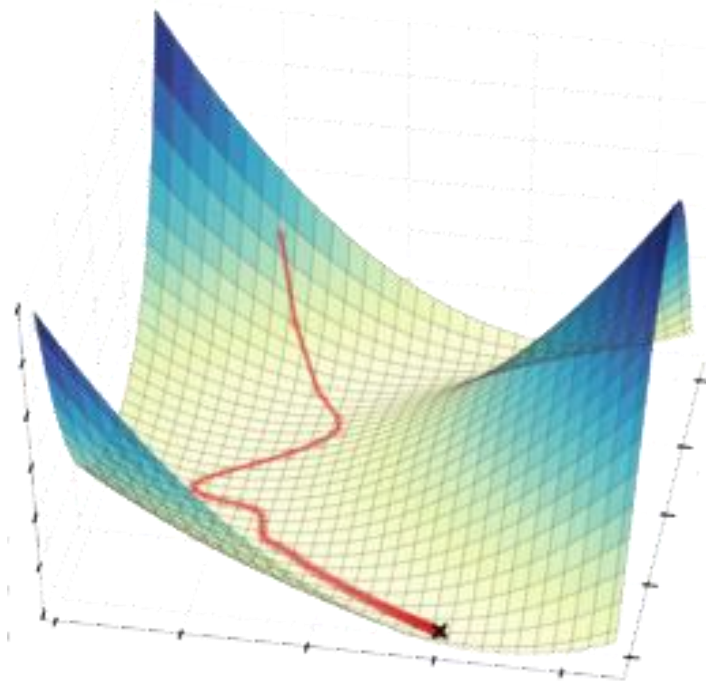
Στη συνάρτηση κόστους προστίθεται και ένας όρος ο οποίος «ποινικοποιεί» την αύξηση των βαρών ώστε το δίκτυο να μην «αποστηθίζει» το σύνολο εκπαίδευσης αλλά να έχει τη δυνατότητα να γενικεύει καλύτερα σε παραδείγματα με τα οποία δεν έχει εκπαιδευτεί. Αυτή η παράμετρος μπορεί να παραμένει αμετάβλητη με την πάροδο του χρόνου ή να προσαρμόζεται, κάτι που οδηγεί σε τελείως διαφορετική κλάση αλγορίθμων τους προσαρμοστικούς (adaptive) όπως ο Adagrad και ο Adam [17].

Στην παρακάτω εικόνα μπορούμε να δούμε το πώς θα συνέβαινε το παραπάνω για ένα μόνο νευρώνα, με τον κανόνα εκπαίδευσης του perceptron (είδος αισθητήρα TN).



Εικόνα 2-2. Εικόνα δικτύου perceptron, με ανάδραση σφάλματος

Για το πρόβλημα ταξινόμησης που θα αναλύσουμε θεωρούμε ότι έχουμε δύο κλάσεις για χάρην ευκολίας της ανάλυσης. Το δίκτυο μας στην αρχή βρίσκεται σε μια κατάσταση, με τα βάρη και τα κατώφλια να έχουν κάποιες τιμές. Στη βιβλιογραφία για να αποφύγουμε υπολογιστικά προβλήματα με τη συνάρτηση κόστους θέλουμε τα βάρη να έχουν μη μηδενικές τιμές αλλά κοντά στο 0. Για αυτό και σε αυτή την εργασία θα υποθέσουμε ότι τα αρχικά βάρη τα αρχικοποιούμε από τυχαίες τιμές κανονικής κατανομής με $\mu=0$ και $\sigma=0.01$ και στη συνέχεια εφαρμόζουμε τον αλγόριθμο βελτιστοποίησης όπως για παράδειγμα την batch gradient descent.



Εικόνα 2-3. Gradient Descent στην επιφάνεια σφάλματος

Επειδή η συνάρτηση που προσπαθούμε να ελαχιστοποιήσουμε είναι μη-κυρτή (non-convex) συνάρτηση, είναι ευαίσθητη σε τοπικά ακρότατα. Στην πράξη η gradient descent (ελαχιστοποίηση με βάση την παράγωγο) δουλεύει συνήθως πολύ καλά, και για να προσπαθήσουμε να αποφύγουμε το πρόβλημα μπορούμε να τρέξουμε το ίδιο πρόβλημα με διαφορετικές αρχικές συνθήκες και να κρατήσουμε το καλύτερο αποτέλεσμα. Τέλος, σημειώνουμε ότι είναι σημαντικό η αρχικοποίηση των παραμέτρων να γίνει τυχαία και να μη θέσουμε όλα τα στοιχεία ίσα με 0 μιας και τότε θα μάθουν το ίδιο ακριβώς φίλτρο σε όλα τα επίπεδα. Μετά από μια επανάληψη της gradient descent “ενημερώνονται” οι παράμετροι με τον τύπο που φαίνεται παρακάτω:

$$W_{ij}^{(l)} := W_{ij}^{(l)} - a \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$$

και τα αντίστοιχα βάρη με τον παρακάτω τύπο:

$$b_i^{(l)} := b_i^{(l)} - a \frac{\partial}{\partial b_i^{(l)}} J(W, b)$$

όπου έχουμε εισάγει την παράμετρο a που αντιστοιχεί στον ρυθμό εκμάθησης (learning rate). Η δυσκολία που κρύβει ο αλγόριθμος είναι στον υπολογισμό των μερικών παραγώγων των εσωτερικών επιπέδων του δικτύου. Για ολόκληρη την ανάλυση του αλγόριθμου και την απόδειξη των τύπων μπορείτε να ανατρέξετε εδώ [18]. Όταν

υπολογιστούν αυτές οι μερικές παράγωγοι, προκύπτει ότι για την κλίση της λαμβάνουμε τους παρακάτω τύπους για την ανανέωση των βαρών και των κατωφλιών αντίστοιχα:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)}$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)})$$

Παρατηρούμε ότι στην ουσία εφαρμόζεται ο ίδιος κανόνας για την ανανέωση των τιμών των παραμέτρων του δικτύου με τη μόνη διαφορά ότι στα βάρη θέλουμε να συμπεριλάβουμε και τον όρο για την κανονικοποίηση. Συνεπώς για να λάβουμε τις παραπάνω τιμές για μία είσοδο χρειάζεται ένα πέρασμα προς εμπρός (“forward pass”) έτσι ώστε να υπολογιστούν όλες οι ενεργοποιήσεις ενδιάμεσου δικτύου καθώς και του τελικού επιπέδου. Εν συνεχεία σε κάθε επίπεδο, υπολογίζεται ένας όρος σφάλματος ο οποίος ποσοτικοποιεί την ευθύνη του κάθε νευρώνα, χρησιμοποιώντας την πληροφορία του αμέσως επόμενου επιπέδου για την παραγωγή του σφάλματος σε αυτό το επίπεδο. Στα εσωτερικά επίπεδα υπολογίζουμε το σταθμικό άθροισμα σφαλμάτων των νευρώνων που προκάλεσαν αυτή την έξοδο.

Συνοπτικά μπορεί να συνοψιστεί ο αλγόριθμος με τα παρακάτω 4 βήματα:

1. Εκτελούμε ένα feed-forward πέρασμα, υπολογίζοντας τις ενεργοποιήσεις για όλα τα επίπεδα ως και L_{n_l} το επίπεδο εξόδου με χρήση των εξισώσεων για forward propagation με τις αντίστοιχες συναρτήσεις ενεργοποίησης.

2. Στο n_l επίπεδο εξόδου υπολογίζουμε το $\delta^{(n_l)} = -(y - a^{(n_l)}) \cdot f'(z^{(n_l)})$

3. Αρχίζοντας από το εγγύτερα στο τέλος $l = n_l - 1, \dots, 2$ θέτουμε: $\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)}) \cdot f'(z^{(l)})$

4. Τελικά υπολογίζουμε τις παραγώγους

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)}$$

Για τον υπολογισμό της παραγώγου εξαρτάται το ποια συνάρτηση ενεργοποίησης έχουμε χρησιμοποιήσει. Για τη RELU η συνάρτηση της παραγώγου είναι η βηματική συνάρτηση

(όπου στη τιμή 0 πρέπει να οριστεί μία εκ των τιμών 0 ή 1), κάτι που την καθιστά ιδανική για μείωση των απαιτούμενων υπολογισμών.

Στη συνέχεια αφού έχουμε τις παραπάνω τιμές για όλο το δίκτυο, κάνουμε την ανανέωση στα βάρη και τα κατώφλια ως εξής:

1. Αρχικοποιούμε τις $\Delta W^{(l)} = 0$, $\Delta b^{(l)} = 0$ σε όλα τα επίπεδα l .

Υπολογίζουμε με τη χρήση του back propagation αλγόριθμου τα παρακάτω

$$i = 1 \dots m,$$

$\nabla W^{(l)} J(W, b; x, y)$ και αντίστοιχα τα $\nabla b^{(l)} J(W, b; x, y)$ και τα προσθέτουμε στο ΔW το $\nabla W^{(l)}$ και στο Δb το $\nabla b^{(l)}$ πολλαπλασιασμένα με τον ρυθμό μάθησης.

2. Αφού έχουμε υπολογίσει τα παραπάνω ενημερώνουμε τις παραμέτρους

$$W^{(l)} = W^{(l)} - a \left[\left(\frac{1}{m} \Delta W^{(l)} \right) \lambda W^{(l)} \right]$$

$$b^{(l)} = b^{(l)} - a \left[\left(\frac{1}{m} \Delta b^{(l)} \right) \right]$$

Για την ολοκλήρωση της εκπαίδευσης πραγματοποιούμε επαναληπτικά τα παραπάνω βήματα για όσες εποχές είναι επιθυμητό. Για το πώς ελέγχεται το πότε διακόπτουμε την εκπαίδευση του δικτύου θα γίνει αναφορά σε επόμενο υπό-κεφάλαιο.

2.6 Stochastic Gradient Descent

Όπως περιγράφηκε παραπάνω ο αλγόριθμος backpropagation, θεωρήθηκε ότι στο δίκτυο περνάν όλα τα δείγματα του συνόλου εκπαίδευσης και στο τέλος υπολογίζουμε την ανανέωση των βαρών. Όταν συμβαίνει αυτό τότε ουσιαστικά εφαρμόζουμε μεθόδους υπολογισμού με πακέτα (batch methods) [19]. Στην πραγματικότητα τα δεδομένα για την εκπαίδευση ενός δικτύου με πολλές κλάσεις όσο έχουν αρχίσει να μεγαλώνουν τα δίκτυα και οι αντίστοιχες παράμετροι αυτού του είδους η τεχνική γίνεται όλο και περισσότερο μη πρακτική γιατί πρέπει να γίνουν πάρα πολλοί υπολογισμοί για ένα και μόνο βήμα στον υπολογισμό της νέας κατεύθυνσης των βαρών.

Ένα ακόμη πρόβλημα της βελτιστοποίησης κατά πακέτα (batch optimization) είναι ότι δεν μπορούμε να εκμεταλλευτούμε νέα παραδείγματα που έρχονται ενώ έχει ήδη αρχίσει η διαδικασία εκπαίδευσης το οποίο το κάνει μη πρακτικό για εφαρμογές που αποκτούμε συνέχεια νέα παραδείγματα εκπαίδευσης. Για τους παραπάνω λόγους πολύ συχνότερα καταφεύγουμε στη χρήση της τεχνικής Stochastic Gradient Descent (SGD) [20] που εξαλείφει τα παραπάνω προβλήματα θυσιάζοντας ένα «μικρό» μέρος θεωρητικής

αξιοπιστίας. Η χρήση της SGD από το ΝΔ δικαιολογείται από το υψηλό κόστος του backpropagation τρέχοντάς το σε όλο το σύνολο δεδομένων. Η SGD υπολογίζει τη νέα κλίση όχι υποβάλλοντας ολόκληρο το σετ αλλά ένα τυχαίο υποσύνολο του, το οποίο στην ακραία περίπτωση μπορεί να φτάσει και το ένα παράδειγμα του συνόλου εισόδου (αν και δεν προτείνεται από τη βιβλιογραφία [21] για λόγους σταθερότητας).

Με τη χρήση της ενημερώνονται οι παραμέτροι θ (βάρη και κατώφλια) της συνάρτησης κόστους με τη χρήση του παρακάτω τύπου για την standard gradient descent:

$$\theta = \theta - \alpha \nabla_{\theta} E[J(\theta)]$$

όπου το α είναι ο ρυθμός μάθησης που μπορεί να είναι σταθερά η παράμετρος που μεταβάλλεται με την πάροδο των εποχών εκπαίδευσης.

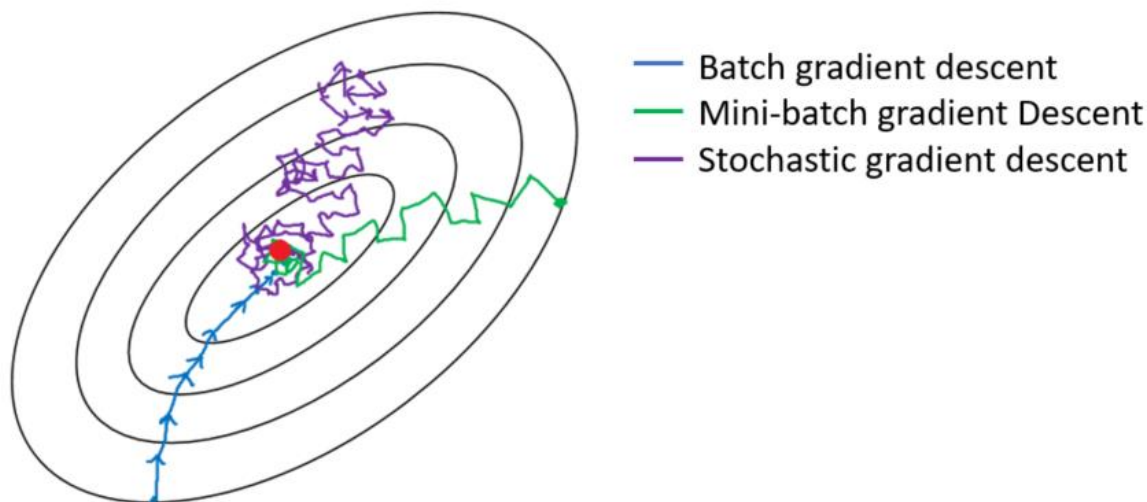
Για την SGD ο παραπάνω τύπος γίνεται για το υποσύνολο U του συνόλου εκπαίδευσης το οποίο επιλέγεται είτε τελείως τυχαία, είτε τυχαία από όλα τα δείγματα που δεν έχει δει σε αυτή την εποχή:

$$\theta = \theta - \alpha \nabla_{\theta} E_U[J(\theta)]$$

Επειδή θέλουμε να μικρύνουμε την διασπορά στην κατεύθυνση που θα κινηθούν οι παράμετροι προτιμούμε να έχουμε ένα σύνολο με παραδείγματα σε μέγεθος από 10 έως 1000 συνήθως, αλλά αυτό είναι μια παράμετρος που αποφασίζεται από τον ερευνητή που προσαρμόζει τον αλγόριθμο για το αντίστοιχο πρόβλημα.

Μια άλλη διαφορά μεταξύ των δύο παραπάνω μεθόδων είναι η παράμετρος α , ο ρυθμός μάθησης. Αυτό συμβαίνει γιατί επειδή είμαστε αρκετά πιο αβέβαιοι για το βήμα που θέλουμε να κάνουμε στην SGD σε σχέση με την standard εκδοχή, και επειδή το πλήθος των ανανεώσεων είναι πολλαπλάσιο μιας και θα κάνουμε πολλαπλές ανανεώσεις για την αντίστοιχη μία της standard SGD, το α αποκτά αντίστοιχα αρκετά μικρότερες τιμές ανάλογα με το μέγεθος του batch με το οποίο εκπαιδεύουμε. Μπορούμε να δούμε την σύγκριση στην παρακάτω απεικόνιση, όπου παρατηρούμε ότι η μπλε καμπύλη είναι ομαλή και πηγαίνει άμεσα αλλά αργά προς το (σε αυτή την περίπτωση μοναδικό) ελάχιστο, η mini-batch εκδοχή όπου παίρνουμε ένα υποσύνολο (πράσινη καμπύλη) παρατηρούμε ότι έχει μικρές αποκλίσεις αλλά αρκετά πιο γρήγορα καταλήγει προς το ελάχιστο και τέλος παρατηρούμε την stochastic gradient descent με ανανέωση ανά ένα παράδειγμα (μωβ καμπύλη) η οποία πηγαίνει προς το ελάχιστο αλλά μετά από πολύ περισσότερα βήματα και μετά από ακραίες μεταβολές στην κατεύθυνση προς την οποία

κινείται (το οποίο σε πραγματικά προβλήματα με πολλαπλά ελάχιστα σημαίνει ότι είναι αρκετά πιθανό να εγκλωβιστούμε σε κάποιο τοπικό ελάχιστο όχι και τόσο επιθυμητό) [17].



Εικόνα 2-4. Πορεία σύγκλισης αλγορίθμου στις περιπτώσεις, (μπλε) batch gradient, (πράσινο) mini-batch gradient, (μωβ) stochastic gradient

Όπως παρατηρούμε όσο πλησιάζουμε προς τη λύση τα βήματα πρέπει να είναι όλο και μικρότερα για να «κλειδώσει» ο αλγόριθμος κοντά στο σημείο τοπικού ελαχίστου που εκτιμούμε ότι βρίσκεται η λύση. Αυτή είναι η λογική που ο ρυθμός μάθησης α επιλέγεται ώστε μετά το πέρας κάποιων βημάτων (εποχών) να μειώνεται αντίστοιχα ή να μειώνεται σύμφωνα με μία συνάρτηση με βάση τον χρόνο (εποχή).

2.7 Ορμή Σύγκλισης

Μια επιπλέον παράμετρος που πρέπει να αποφασίσει ο αρχιτέκτονας που σχεδιάζει το νευρωνικό δίκτυο είναι η χρήση της ορμής. Η έμπνευση της παραπάνω μεθόδου οφείλετε στη φυσική και η βασική λογική της είναι η ακόλουθη. Όσο η μεταβολή της συνάρτησης μας είναι προς την κατεύθυνση που μεταβλήθηκε και προηγουμένως τότε πιθανότατα η τελική λύση θα είναι προς την ίδια κατεύθυνση συνεπώς σε κάθε επόμενο βήμα αυξάνεται η ταχύτητα με την οποία συγκλίνει η συνάρτηση. Αφού η έμπνευση της ήταν από την ταχύτητα συμβολίζεται και με u . Οι τύποι που χρησιμοποιούμε για να ανανεώσουμε το διάνυσμα των βαρών είναι οι εξής:

$$v = \gamma v + \alpha \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$$

$$\theta = \theta - v$$

Όπου με ν συμβολίζουμε την τρέχουσα ταχύτητα. Το ν είναι αντίστοιχο δάνυσμα σε μέγεθος με το θ το οποίο ανανεώνει. Το γ αποτελεί μια παράμετρο του πλήθους των επαναλήψεων που προηγούμενες ανανεώσεις της ταχύτητας θέλουμε να ληφθούν υπόψιν στην παρούσα ενημέρωση, με την παράμετρο αυτή να παίρνει τιμές το εύρος τιμών $[0, 1]$. Η επιλογή αρχικής τιμής είναι μέρος του σχεδιασμού του αλγόριθμου με μια συχνή τιμή να αποτελεί το 0.5. Το γ δεν είναι σταθερά αλλά το μεταβάλλουμε κατά τη διάρκεια των εποχών σε μεγαλύτερες τιμές όσο πλησιάζουμε περισσότερο προς τη λύση. Σε πιο προχωρημένες λύσεις μαθαίνουμε και τον παράγοντα γ κατά τη διάρκεια της εκπαίδευσης [22].

3. ΣΥΝΕΛΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

3.1 Εισαγωγή

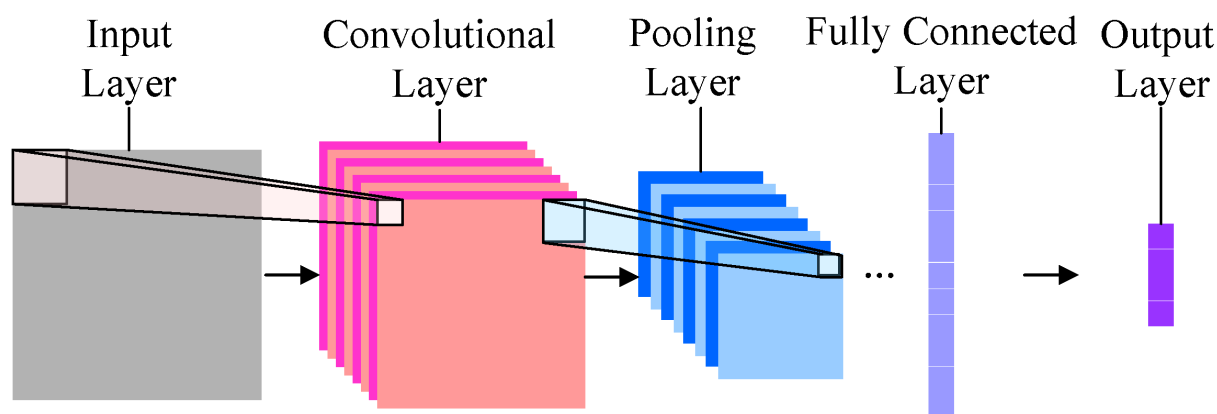
Ο βασικός λόγος που αφιερώσαμε το προηγούμενο κεφάλαιο στα νευρωνικά δίκτυα γενικότερα, είναι το γεγονός ότι πολλές από τις τεχνικές και αλγόριθμους που εφαρμόζονται και αναφέρθηκαν μπορούν αντίστοιχα να εφαρμοστούν στα συνελκτικά νευρωνικά δίκτυα (convolutional neural networks ή CNN). Τα CNN αποτελούνται από ένα ή συνήθως πολλά επίπεδα συνέλιξης (convolutional layers) τα οποία συχνά ακολουθούνται από αντίστοιχα επίπεδα υπέρ-δειγματοληψίας (pooling layers) και πιθανόν κοντά στην έξοδο ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα.

Η αρχιτεκτονική των CNN είναι τέτοια ώστε να εκμεταλλεύεται τη δομή των εικόνων στα οποία υπάρχει τοπικότητα στον χώρο, μιας και κοντινά pixel έχουν περισσότερη σχέση από αυτά που βρίσκονται μακρύτερα. Για να μπορέσουν να εκμεταλλευτούν αυτήν την τοπικότητα αντιστοιχίζονται κατάλληλα βάρη σε φίλτρα τα οποία «ασχολούνται» με ένα μικρό μέρος της εικόνας, το οποίο αναφέρεται στη βιβλιογραφία ως «πεδίο αντίληψης» (receptive field). Στη συνέχεια χρησιμοποιούμε μηχανισμούς δειγματοληψίας (pooling) ώστε να καταστήσουμε το δίκτυο ανθεκτικό σε διάφορους μετασχηματισμούς των εικόνων που ψάχνουμε όπως περιστροφή, αλλαγή της θέσης στην εικόνα, μεγέθυνση και άλλα. Επίσης για ένα δίκτυο CNN με ένα αντίστοιχο fully connected neural network (πλήρως συνδεδεμένο συνελκτικό δίκτυο) με τους ίδιους κόμβους σε κάθε επίπεδο έχει πολύ λιγότερες παραμέτρους, από ένα πλήρως συνδεδεμένο δίκτυο, κάτι που καθιστά την μάθηση πολύ πιο γρήγορη. Επίσης από τα φίλτρα που μαθαίνονται παίρνουμε πολύ χρήσιμες πληροφορίες για το ποια είναι τα χαρακτηριστικά που το ίδιο το δίκτυο εκτιμά ότι είναι χρήσιμα για τον διαχωρισμό των κλάσεων.

Σε ένα επίπεδο συνέλιξης η είσοδος (στο πρώτο επίπεδο) αποτελείται από εικόνα όπου έχει τρεις διαστάσεις, το ύψος, το πλάτος και ο αριθμός των καναλιών (για μια έγχρωμη εικόνα 3 ή 4 αν αποθηκεύουμε και διαφάνεια (transparency) ειδάλλως το βάθος αρχίζει με μέγεθος 1 για ασπρόμαυρη εικόνα). Το επίπεδο συνέλιξης αποτελείται από φίλτρα (kernels) μεγέθους μικρότερο από τη διάσταση της εικόνας στο ύψος και το πλάτος αλλά αφορούν όλο το βάθος στο οποίο εκτείνονται. Αυτό είναι σημαντικό να γίνει αντιληπτό, το γεγονός δηλαδή ότι εκμεταλλευόμαστε την τοπικότητα στο ύψος και στο πλάτος όμως για κάθε πυρήνα το βάθος του φίλτρου παραμένει το ίδιο με το βάθος εξόδου για το κάθε φίλτρο (προφανώς μιας και το πλήθος των φίλτρων παραμένει) και επηρεάζει ανεξάρτητα

το κάθε επίπεδο. Συνεπώς για να αυξηθεί το μέγεθος του βάθους προσθέτουμε αντίστοιχα το πλήθος των φίλτρων που θέλουμε. Αντίστοιχα η έξοδος του κάθε φίλτρου αποτελείται από βάθος 1 και πλάτος ανάλογα με το αν το φίλτρο μικραίνει, διατηρεί ή μεγαλώνει το μέγεθος της αρχικής εικόνας.

Ανάλογα με το μέγεθος των φίλτρων δημιουργείται μια τοπικά συνδεδεμένη δομή όπου το καθένα συνελίσσεται με κάθε εικόνα για να παράξουν χάρτες χαρακτηριστικών (feature maps). Κάθε χαρακτηριστικό υπό-δειγματοληπτείται τυπικά με mean ή max ή average pooling σε συνεχείς περιοχές όπου τυπικές τιμές που χρησιμοποιούνται είναι μεταξύ 2 και του 5 για μεγάλες εικόνες εισόδου. Πριν ή μετά το pooling layer (επίπεδο δειγματοληψίας) ακολουθεί μια προσθήκη bias και κάποιου είδους μη γραμμικότητα για κάθε έναν από τους χάρτες χαρακτηριστικών που παράγονται.



Εικόνα 3-1. Σχηματική απεικόνιση της μεταβολής των δεδομένων εισόδου

Στο τέλος μετά το πέρας αρκετών αντίστοιχων επιπέδων (με πιθανόν όχι απόλυτα γραμμική ακολουθία) έχουμε ένα σετ από χαρακτηριστικά για την εικόνα τα οποία συνήθως τα συνδέουμε σε ένα πλήρως συνδεδεμένο δίκτυο. Αυτό συμβαίνει γιατί θέλουμε στο τέλος για την αρχική μας εικόνα να λάβουμε υπόψιν όλα τα χαρακτηριστικά που έχουμε για ολόκληρη την εικόνα και όχι για κάθε ένα από τα μέρη της. Το πλήρως συνδεδεμένο δίκτυο έχει ως επίπεδο εξόδου (output layer) μέγεθος όσο και το πλήθος των κλάσεων που θέλουμε να διαχωρίζει. Αυτό δεν συμβαίνει πάντα μιας και μπορούμε να κάνουμε και άλλες κωδικοποιήσεις για την έξοδο κάτι που δε χρησιμοποιήθηκε στην παρούσα εργασία. Τα αποτελέσματα του τελευταίου επιπέδου μπορούν με τη χρήση της softmax συνάρτησης να «αντιστοιχιστούν» σε πιθανότητες που αντιστοιχούν στην εκτίμηση του αλγόριθμου για κάθε κλάση και συνεπώς, εκτός από την κλάση που έχει προβλέψει το CNN να έχουμε και την βεβαιότητα που έχει ο αλγόριθμος για αυτή την κλάση [23].

3.2 Δομή Συνελκτικών Νευρωνικών Δικτύων

Τα συνελκτικά νευρωνικά δίκτυα έχουν πολλές ομοιότητες με τα γενικά νευρωνικά δίκτυα που αναλύθηκαν στο προηγούμενο κεφάλαιο. Αποτελούνται από νευρώνες που έχουν εισόδους με αντίστοιχα βάρη και ένα κατώφλι. Κάθε νευρώνας λαμβάνει εισόδους, εκτελεί ένα εσωτερικό γινόμενο και το ακολουθεί με μια συνάρτηση μη γραμμική. Το συνολικό δίκτυο εξακολουθεί να εκφράζει μια ενιαία λειτουργία διαφοροποιήσιμης συνάρτησης που παράγει ένα αποτέλεσμα, στην περίπτωση της ταξινόμησης είτε την πιθανότητα ή ένα σκορ για το κατά πόσο ανήκει σε κάποια από τις κλάσεις. Από τα εικονοστοιχεία (pixels) της πρώτης εικόνας στο ένα άκρο έως τα αποτελέσματα της κατηγορίας στο άλλο άκρο του δικτύου. Εξακολουθούμε να έχουμε μια συνάρτηση απώλειας (συνήθως Softmax / cross-binary entropy / Support Vector Machines (SVM)) στο τελευταίο επίπεδο το οποίο είναι πλήρως συνδεδεμένο και εξακολουθούν να ισχύουν όλα όσα έχουμε αναφέρει σε ότι αφορά τις τεχνικές για την εκπαίδευση τους.

Ας εξετάσουμε τώρα τις διαφορές στις αρχιτεκτονικές συνελκτικών δικτύων. Τα CNN κάνουν τη ρητή παραδοχή ότι η είσοδος είναι εικόνες, που μας επιτρέπει να κωδικοποιούμε ορισμένες ιδιότητες στην αρχιτεκτονική τους. Μια πολύ σημαντική εξ αυτών είναι το γεγονός ότι υπάρχει μια εγγενής σχέση μεταξύ των χαρακτηριστικών αυτή της απόστασης μεταξύ τους στην εικόνα, την οποία και προσπαθούμε να εκμεταλλευτούμε με αντίστοιχα τοπικούς πυρήνες που αφορούν σε χαρακτηριστικά σε «κοντινά» pixel. Αυτά καθιστούν τότε πιο αποτελεσματική (λιγότερο κοστοβόρα σε πράξεις) τον εμπρόσθιο υπολογισμό του δικτύου και κυρίως μειώνουν σημαντικά την ποσότητα των παραμέτρων που χρειάζεται να εκπαιδευτούν στο δίκτυο.

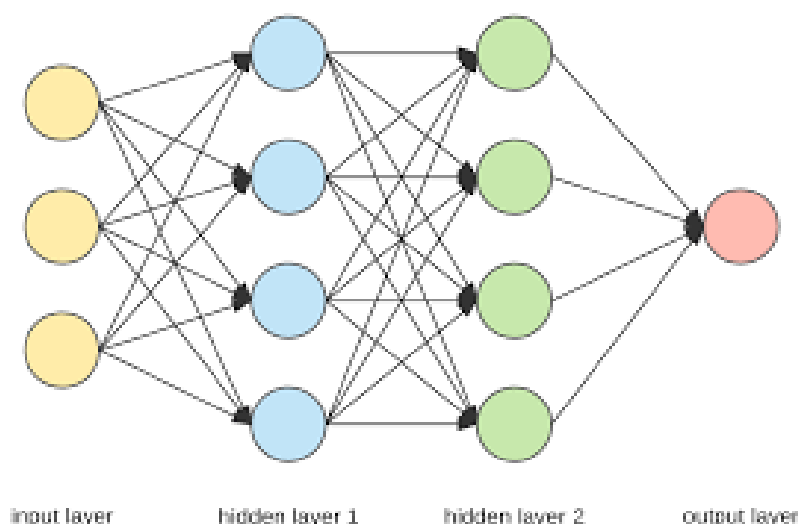
3.2.1 Αρχιτεκτονική Συνελκτικών δικτύων

Τα γενικά νευρωνικά δίκτυα λαμβάνουν μια είσοδο (ένα μόνο διάνυσμα) και τη μετατρέπουν μέσω μιας σειράς κρυφών συνελκτικών επιπέδων στην επιθυμητή έξοδο. Σε κάθε κρυμμένο στρώμα υπάρχει ένα σύνολο νευρώνων, όπου κάθε νευρώνας είναι πλήρως συνδεδεμένος με όλους τους νευρώνες στο προηγούμενο στρώμα και οι νευρώνες σε ένα μόνο στρώμα λειτουργούν εντελώς ανεξάρτητα και δεν μοιράζονται καμία σύνδεση. Συνεπώς για διαφορετικά κομμάτια τις εικόνας μαθαίνει διαφορετικά βάρη. Στο τελευταίο πλήρως συνδεδεμένο επίπεδο υπάρχει ένα "στρώμα εξόδου" και ουσιαστικά αντιπροσωπεύει τις βαθμολογίες της εκάστοτε κλάσης αν επιλεγεί η softmax συνάρτηση ή την επιλεχθείσα κλάση εξόδου ανάλογα με την υλοποίηση.

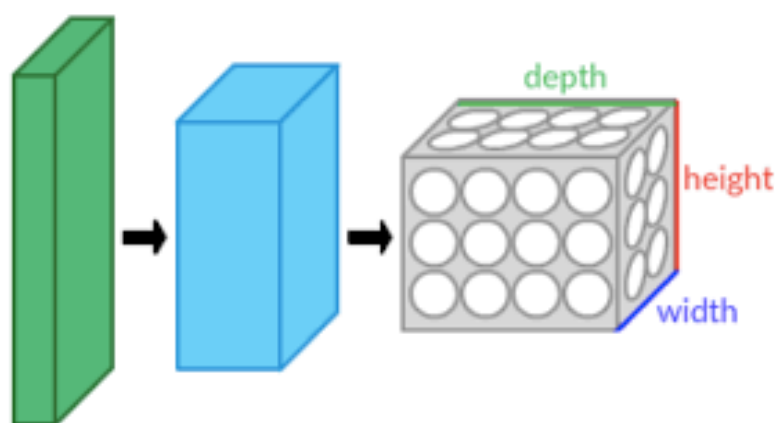
Τα κανονικά νευρωνικά δίκτυα επίσης δεν κλιμακώνονται καλά σε πλήρεις εικόνες με την ανάλυση που χρειαζόμαστε για ιατρικές εικόνες. Ας μιλήσουμε λίγο για παράδειγμα για το πολυχρησιμοποιημένο δίκτυο CIFAR-10 στο οποίο οι εικόνες έχουν μέγεθος μόνο $32 \times 32 \times 3$ (32 πλάτος, 32 ύψος, 3 κανάλια χρώματος), έτσι ώστε ένας μόνο πλήρως συνδεδεμένος νευρώνας σε ένα πρώτο κρυφό στρώμα ενός κανονικού νευρικού δικτύου θα είχε $32 * 32 * 3 = 3072$ βάρη. Το ποσό αυτό αν και φαίνεται διαχειρίσιμο, δεν μπορεί να κλιμακώνεται σε μεγαλύτερες εικόνες μιας και αυξάνεται το πλήθος των συνδέσεων (άρα και των βαρών) ραγδαία. Για παράδειγμα, μια εικόνα πάλι μικρή για πραγματικά δεδομένα αλλά με ελαφρώς πιο ρεαλιστικό μέγεθος έστω $300 \times 300 \times 3$, θα οδηγούσε σε νευρώνες που έχουν $300 * 300 * 3 = 270.000$ βάρη για έναν και μόνο νευρώνα. Είναι σαφές ότι αυτή η πλήρης συνδεσιμότητα είναι μη πρακτική και ο τεράστιος αριθμός παραμέτρων θα οδηγούσε σε μη ρεαλιστική δυνατότητα γενίκευσης και συνεπώς εκπαίδευσης.

Στα συνελκτικά νευρωνικά δίκτυα επωφελούμαστε από το γεγονός ότι η είσοδος αποτελείται από εικόνες και μας ενδιαφέρει να εκμεταλλευτούμε την τοπικότητα των χαρακτηριστικών των εικόνων, συνεπώς μπορεί να περιοριστεί η αρχιτεκτονική με έναν πιο λογικό τρόπο. Ειδικότερα, τα στρώματα ενός CNN έχουν νευρώνες διατεταγμένους σε 3 διαστάσεις: πλάτος, ύψος, βάθος. Ας σημειώσουμε ότι το βάθος εδώ αναφέρεται στην τρίτη διάσταση του τόμου ενεργοποίησης, όχι στο βάθος του πλήρους νευρωνικού δικτύου, το οποίο αναφέρεται στο συνολικό αριθμό των στρώσεων σε ένα δίκτυο.

Για παράδειγμα, οι εικόνες εισόδου στο CIFAR-10 είναι ένας «όγκος εισόδου ενεργοποιήσεων» και ο όγκος έχει διαστάσεις $32 \times 32 \times 3$ (πλάτος, ύψος, βάθος). Οι νευρώνες σε ένα στρώμα θα συνδεθούν μόνο σε μια μικρή περιοχή του στρώματος πριν από αυτό, αντί για όλους τους νευρώνες. Επιπλέον, το τελικό στρώμα εξόδου θα έχει διαστάσεις $1 \times 1 \times 10$, διότι μέχρι το τέλος της αρχιτεκτονικής του CNN θα μειώσουμε την πλήρη εικόνα σε ένα διάνυσμα βαθμολογίας κλάσης, διατεταγμένο κατά μήκος της διάστασης βάθους αντίστοιχου μεγέθους με το πλήθος των κλάσεων που έχουμε να προβλέψουμε (10 στην περίπτωση του CIFAR-10) που έχουμε 10 κλάσεις εξόδου. Μια απεικόνιση των παραπάνω σε ότι αφορά τη διαφορά με τα πλήρη συνδεδεμένα δίκτυα μπορεί να φανεί στις παρακάτω εικόνες:



Εικόνα 3-2. Βασική δομή πλήρως συνδεδεμένου νευρωνικού δικτύου



Εικόνα 3-3. Μετασχηματισμός του αρχικού "όγκου" χαρακτηριστικών, με βαθμιαία αύξηση του βάθους

Συνεπώς όπως βλέπουμε και στις παραπάνω εικόνες αν το εξετάζαμε προγραμματιστικά το παραπάνω παρατηρήσουμε ότι ένα CNN αποτελείται από επίπεδα με κάθε επίπεδο έχει μια απλή προγραμματιστική διεπαφή και σκοπό, να μετατρέψει έναν όγκο 3D εισόδου σε έναν αντίστοιχο 3D όγκο εξόδου με κάποια μη γραμμική διαφοροποιήσιμη συνάρτηση που μπορεί να έχει ή να μην έχει παραμέτρους. Σε αυτή τη μετατροπή το μέγεθος του όγκου καθώς διέρχεται το δίκτυο από επίπεδο σε επίπεδο διατηρεί ή μικραίνει το μέγεθος του ύψους και πλάτους, αλλά διατηρεί το ίδιο ή αυξάνει το βάθος του όγκου.

Χρησιμοποιούνται συνήθως τρεις βασικοί τύποι επιπέδων για την κατασκευή αρχιτεκτονικών CNN: Συνελκτικό Επίπεδο (Convolutional Layer), Επίπεδο δειγματοληψίας (Layer Pooling) και πλήρως συνδεδεμένο επίπεδο (Fully Layer Connected) (αντίστοιχο των γενικών νευρωνικών δικτύων). Στοιβάζονται αυτά τα επίπεδα

με αντίστοιχα μεγέθη ώστε να σχηματίσουν μια πλήρη αρχιτεκτονική του CNN . Η σειρά με την οποία τοποθετούνται το πλήθος τους και το μέγεθος καθώς και η συνάρτηση που χρησιμοποιείται σε κάθε επίπεδο είναι αυτό που διαφοροποιεί τις εκάστοτε αρχιτεκτονικές που υπάρχουν στη βιβλιογραφία [24][25] σε ότι αφορά τα συνελκτικά δίκτυα.

3.2.2 Είδη επιπέδων συνελκτικών δικτύων

Στην παρούσα ενότητα παρουσιάζουμε τα είδη των «μονάδων»(units) που εμφανίζονται συνήθως σε αρχιτεκτονικές που έχουν παρουσιαστεί για νευρωνικά δίκτυα.

3.2.2.1 Αρχικό επίπεδο

Στο αρχικό επίπεδο τοποθετείται η εικόνα η οποία έχει για κάθε εικόνα του συνόλου εκπαίδευσης συγκεκριμένο μέγεθος σε πλάτος, ύψος. Εδώ χρειάζεται προσοχή το πως είναι αποθηκευμένο το σύνολο των εικόνων. Σε πολλά σύνολα, η θέση που αποθηκεύεται το βάθος μπορεί να είναι είτε στην πρώτη ή στην τρίτη διάσταση. Επίσης παίζει ρόλο στην διάσταση του βάθους με ποια σειρά εκπαιδεύτηκε το δίκτυο, με RGB ή με BGR. Επίσης βασικό είναι όταν εκπαιδεύεται το δίκτυο συνήθως επιθυμούμε οι είσοδοι να είναι κεντραρισμένοι στο 0 με μικρή απόκλιση. Άρα πρέπει να αφαιρεθεί το μια σταθερή ποσότητα από κάθε χρώμα με αυτό που εκπαιδεύτηκε το αρχικό δίκτυο. Επίσης στη συνέχεια πρέπει να διαιρεθεί το κάθε χρώμα (συνήθως όπως και στην παρούσα εργασία με το 255) ώστε το αποτέλεσμα να είναι μέσα στο διάστημα $[-1,1]$.

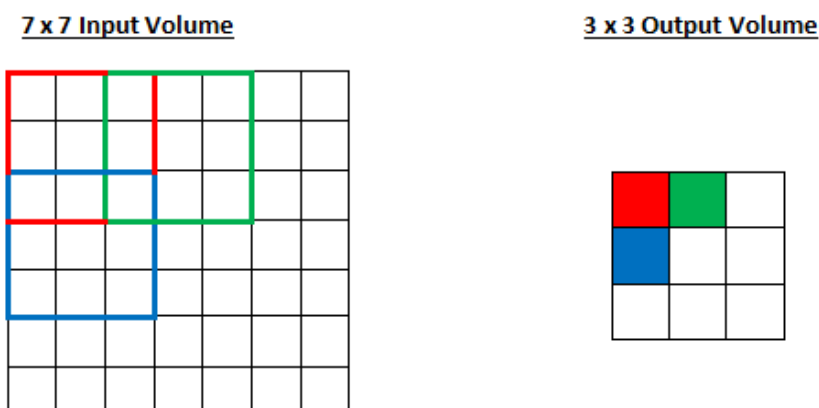
3.2.2.2 Συνελκτικό Επίπεδο (Convolutional Layer)

Σε ένα συνελκτικό επίπεδο οι παράμετροι του αντίστοιχου επιπέδου αποτελούνται από ένα σύνολο φίλτρων που μπορούν να μεταβάλουν τα βάρη τους ώστε να διδαχθούν. Κάθε φίλτρο είναι μικρό (κατά πλάτος και ύψος) σε αντιστοιχία με την είσοδο του, αλλά εκτείνεται μέσα από το πλήρες εύρος του επιπέδου εισόδου. Για παράδειγμα, ένα τυπικό φίλτρο σε ένα πρώτο στρώμα ενός CNN μπορεί να έχει μέγεθος $7 \times 7 \times 3$ (δηλαδή πλάτος και ύψος 7 pixel και 3 βάθος όσο και τα χρώματα της εικόνας ή 1 για ασπρόμαυρη).

Κατά τη διάρκεια της προς τα εμπρός διάδοσης, μετατοπίζουμε κάθε φίλτρο σε όλο το πλάτος και το ύψος του όγκου εισόδου και υπολογίζουμε το γινόμενο μεταξύ των αντίστοιχων θέσεων του φίλτρου και της εισόδου σε οποιαδήποτε θέση. Καθώς περνάμε το φίλτρο πάνω από το πλάτος και το ύψος του όγκου εισόδου, δημιουργούμε έναν 2-διαστάσεων χάρτη ενεργοποίησης που δίνει τις απαντήσεις του συγκεκριμένου φίλτρου σε κάθε θέση στο χώρο (ύψος, πλάτος) της αρχικής εικόνας. Το δίκτυο «βλέπει» σε αυτό

το πέρασμα σε ποια θέση της αρχικής εικόνας έχει ενεργοποιηθεί το κάθε φίλτρο και συνεπώς θα μάθει φίλτρα που ενεργοποιούνται όταν βλέπουν κάποιο τύπο οπτικού χαρακτηριστικού, όπως ακμές ή γωνίες ή σε υψηλότερα επίπεδα υφή (texture) ή ελλειπτικά μοτίβα [26]. Ας σημειωθεί ότι δεν έχουμε μόνο ένα φίλτρο άλλα πολύ περισσότερα σε κάθε επίπεδο, γιατί θέλουμε να εξετάσουμε τον «όγκο» εισόδου για διάφορα μοτίβα, πού θα είναι αντίστοιχα με το πλήθος των φίλτρων που θα προσθέσουμε [27][28]. Κάθε ένα από αυτά θα παράγει έναν ξεχωριστό χάρτη 2-διαστάσεων ενεργοποίησης (feature map) [29]. Όταν στοιβάζονται αυτοί οι χάρτες ενεργοποίησης (εννοείται τοποθετώντας τους πάντα στην ίδια σειρά) κατά μήκος της διάστασης βάθους παράγεται ο «όγκος» εξόδου.

Υπάρχουν κάποια σημαντικά χαρακτηριστικά που πρέπει να αναφερθούν σε αυτό το σημείο. Το πρώτο είναι το stride (διασκελισμός) το οποίο θα το αναφέρουμε από εδώ και στο εξής με την αγγλική ορολογία του. Το stride είναι μια από τις υπέρ-παραμέτρους που πρέπει να επιλέξει ο δημιουργός του νευρωνικού επιπέδου για κάθε ένα από τα συνελκτικά επίπεδα. Ουσιαστικά αφορά στο πόσο θα μετακινείται σε κάθε διάσταση ο πυρήνας της συνέλιξης. Αν απλά μετακινείται κατά ένα τότε ουσιαστικά η έξοδος θα προκύψει λίγο μικρότερη, όσο και το μέγεθος του φίλτρου συνέλιξης μείον ένα. Αν όμως επιλεγεί μεγαλύτερη τιμή (συνήθως για μεγαλύτερους σε μέγεθος πυρήνες) όπως π.χ. 3 τότε η έξοδος που προκύπτει είναι μικρότερη. Όποτε ένας βασικός λόγος ύπαρξης του stride είναι η μείωση του επιπέδου εξόδου.

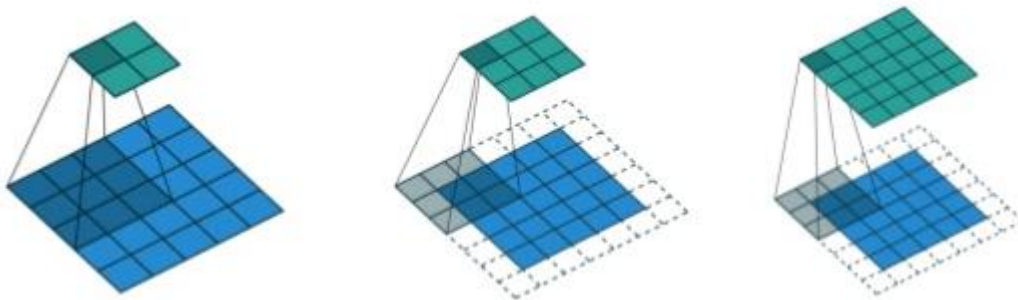


Εικόνα 3-4. Παράδειγμα συνέλιξης και το αποτέλεσμα που έχει στην έξοδο. Είσοδος 7x7, πυρήνας συνέλιξης 3x3, για stride 2 και 0 padding

Επίσης μπορούμε να παρατηρήσουμε ότι κομμάτια της εισόδου χρησιμοποιούνται για περισσότερες από μία εξόδους. Αυτό συμβαίνει γιατί μπορεί ένα χαρακτηριστικό να

βρίσκεται ανάμεσα σε 2 τετράγωνα. Σε αυτή την περίπτωση αν το stride ήταν όσο το μέγεθος του πυρήνα θα εντοπιζόταν. Οπότε είναι ευθύνη του σχεδιαστή του νευρωνικού δικτύου να εντοπίσει σε ποια επίπεδα θέλουμε μεγάλη ευαισθησία και σε ποια όχι για τον εντοπισμό ενός χαρακτηριστικού.

Μια άλλη υπερ-παράμετρος των νευρωνικών επιπέδων είναι το βοηθητικό συμπλήρωμα (padding) το οποίο θα το αναφέρουμε από εδώ και στο εξής με την αγγλική ορολογία του. Το padding το χρησιμοποιούμε για να ελέγξουμε το μέγεθος που θα έχει η έξοδος του εκάστοτε επιπέδου. Πιο συγκεκριμένα, είναι ένα πλήθος από στοιχεία στα οποία δίνεται μία αρχική αυθαίρετη τιμή τα οποία επικολλώνται στα άκρα της εικόνας, μόνο στις κατευθύνσεις του πλάτους και του ύψους. Η τιμή που δίνεται συνήθως είναι το 0 όμως ανάλογα με το πρόβλημα μπορεί είτε ενιαία να δοθεί μια άλλη τιμή ή απλά να αντιγράψουμε την τιμή του πιο κοντινού στοιχείου του για να μην αλλοιωθεί δραματικά η τιμή στην αντίστοιχη θέση σε σχέση με τις εσωτερικές. Στην πράξη ένας από τους πλέον συχνούς τρόπους που χρησιμοποιείται είναι εάν δεν είναι επιθυμητό να μειωθεί το μέγεθος της εικόνας. Σε αυτή την περίπτωση πρέπει να προσθέσουμε padding σε κάθε πλευρά $Padding = (Kernel Size - 1)/2$ στην περίπτωση που το stride είναι 1. Αυτός είναι και ο λόγος που οι πυρήνες που επιλέγονται έχουν περιττό πλήθος, ώστε να μπορεί να προστεθεί ως padding ακέραιο πλήθος στοιχείων.



Εικόνα 3-5. Εφαρμογή διαφόρων μεγεθών πυρήνα στην είσοδο, με και χωρίς εφαρμογή padding (1^η περ. padding=0, stride=1, 2^η περ. padding=1, stride=2, 3^η περ. padding=1, stride=1)

Ο γενικός τύπος για το μέγεθος της εξόδου που προκύπτει από την εφαρμογή ενός φίλτρου είναι ο ακόλουθος:

$$Διαστάση\ Εξόδου = \frac{Διάσταση\ Εισόδου - Διάσταση\ Πυρήνα + 2 * padding}{stride} + 1$$

Τέλος ένα ακόμα που πρέπει να προσεχθεί είναι η επιλογή του stride να είναι τέτοια ώστε να προκύπτει ακέραιος αριθμός από το παραπάνω κλάσμα αλλιώς θα είναι ακανόνιστη η έξοδος στο επόμενο επίπεδο.

Οι παραπάνω παράμετροι για κάθε φίλτρο του επιπέδου πρέπει να είναι οι ίδιες αλλιώς θα έχουμε διαφορετικό μέγεθος εξόδου για κάθε φίλτρο που το δυσκολεύει από άποψη κατανόησης, οργάνωσης και συμβολισμού χωρίς κάποιο πρακτικό όφελος. Είναι σαφές ότι εκμεταλλευόμαστε το γεγονός ότι αν ένα χαρακτηριστικό είναι σημαντικό σε ένα κομμάτι της εικόνας, μπορεί να είναι αντίστοιχα σημαντικό σε κάποιο άλλο μέρος της εικόνας.

Βασική επιτυχία των συνελκτικών δικτύων σε σχέση με τα πλήρως συνδεδεμένα επίπεδα είναι η οργανωμένη μείωση των συνδέσεων από επίπεδο σε επίπεδο, για λόγους τοπικότητας των χαρακτηριστικών. Αυτό προκαλεί μείωση των βαρών που πρέπει να υπολογιστούν και συνεπώς επίτευξη αντίστοιχου επιπέδου κατηγοριοποίησης με λιγότερα παραδείγματα.

Μια παρατήρηση για τα βάρη του πυρήνα συνέλιξης που προσπαθεί το δίκτυο να υπολογίσει, όταν στα σήματα ονομάζουμε συνέλιξη ο φορμαλισμός είναι ότι ο πυρήνας αντιστρέφεται, στη συνέχεια πολλαπλασιάζεται με το αρχικό σήμα και το τελικό ολοκλήρωμα (στην περίπτωση μας άθροισμα λόγω πεπερασμένου μεγέθους σήματος) είναι το αποτέλεσμα. Θα μπορούσαμε να υποθέσουμε ότι ο πίνακας βαρών του συνελκτικού δικτύου είναι τοποθετημένος ήδη αντίστροφα όποτε και θα έχουμε την ίδια ακριβώς πράξη. Επειδή όμως στη πράξη εφαρμόζεται πολύ συχνά και είναι «οπτικά όμορφο» όταν απεικονίζουμε το φίλτρο να βλέπουμε τι σχήμα έχει, στον ίδιο προσανατολισμό με τα δεδομένα, έχει επιλεγεί να αποθηκεύεται δίχως να αντιστρέφεται.

3.2.2.2.1 Σύγχρονες εξελίξεις στα συνελκτικά δίκτυα

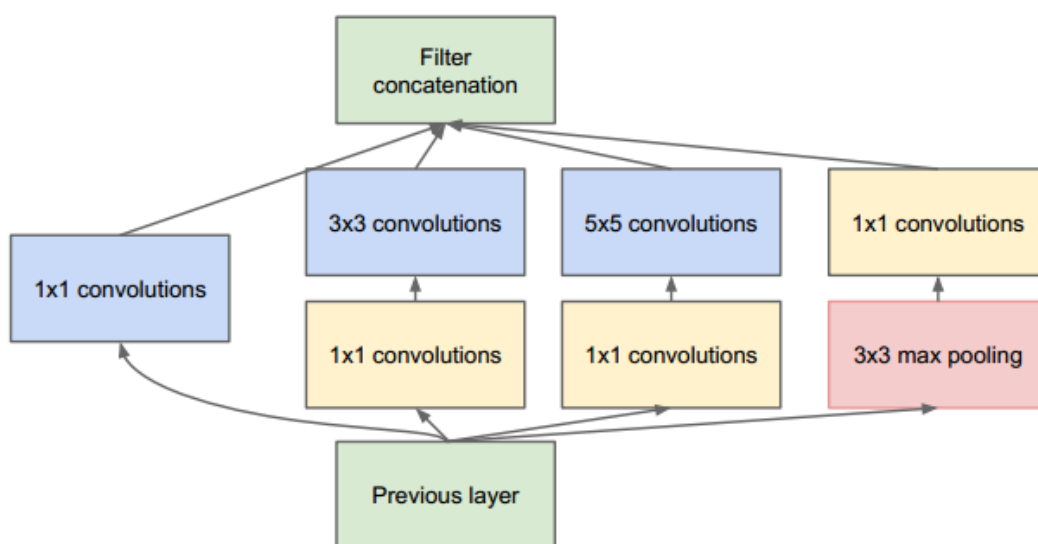
Παρότι μπορεί να φαίνεται στον αναγνώστη ότι η αρχιτεκτονική των CNN πλέον είναι δεδομένη και λίγα μπορούν να μεταβληθούν, υπάρχουν εκτός από αρχιτεκτονικές οι οποίες γίνονται βαθύτερες, και περισσότερη κατανόηση για τα διάφορα επίπεδα του CNN.

Ένα παράδειγμα, είναι οι διαστελλόμενες συνέλιξεις (dilation convolutions) [28]. Οι λογική πίσω από αυτές είναι ότι δεν είναι απαραίτητο ο πυρήνας της συνέλιξης να πολλαπλασιάζεται με μια συνεχή περιοχή της εικόνας, αλλά μπορεί να υπάρχουν κενά ανάμεσα στα pixels. Με αυτόν τον τρόπο ένα πυρήνας αφορά σε μεγαλύτερη περιοχή της εικόνας εισόδου και μπορεί να εντοπίζει χαρακτηριστικά που αφορούν μεγαλύτερες δομές

στην εικόνα, με πολύ λιγότερες παραμέτρους από τον συνδυασμό αρκετών επιπέδων συνεχόμενων πυρήνων.

Οι συνελίξεις 1×1 είναι ένα τέτοιο παράδειγμα [30]. Παρότι μπορεί να φαίνονται ανούσιες μιας και παίρνουν ένα «pixel» της εικόνας εισόδου, και το περνούν από μία συνάρτηση μη γραμμική. Δεν είναι προφανές το πλεονέκτημα, μέχρι να γίνει αντιληπτό ότι ο «όγκος» εισόδου είναι άλλη μια διάσταση, το βάθος. Συνεπώς ένα επίπεδο μη γραμμικό 1×1 αυτό που κάνει είναι, ότι στα ενδιάμεσα επίπεδα μπορεί να υπολογίσει μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών που έχουν ήδη υπολογιστεί για μία συγκεκριμένη περιοχή της εικόνας.

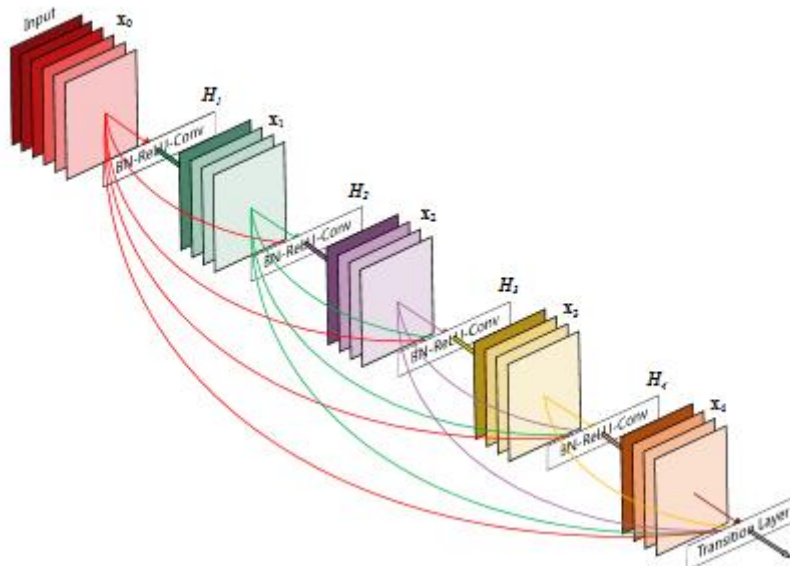
Μία λίγο πιο ανεπτυγμένη πρόσφατη εφαρμογή ήταν ένα επιπλέον επίπεδο “inception” που τοποθετήθηκε στο δίκτυο GoogLeNet [31]. Η βασική ιδέα είναι ότι αντί να επιλέγει σε κάθε θέση ποιο από «ζουμ» στην κάθε περιοχή της εικόνας χρειάζεται περισσότερο το 1×1 το 3×3 ή το 5×5 αποφάσισαν να τοποθετηθούν όλα τα παραπάνω. Αυτό όμως θα απαιτούσε τρομακτικά πολύ χώρο. Οπότε στο κάθε ένα από αυτούς τους πυρήνες δεν υπολόγιζαν τη συνέλιξη με ολόκληρο τον όγκο εισόδου, αλλά με ένα υποσύνολο το οποίο υπολογίζεται από μια μη γραμμικότητα 1×1 πολύ μικρού βάθους (στο παραπάνω δίκτυο επιλέχθηκαν 16 φίλτρα εξόδου). Με αυτό τον τρόπο «κατανοούν» με τη χρήση μικρού χώρου, πληροφορία από τα διάφορα ζουμ σε κάθε θέση εισόδου.



Εικόνα 3-6. Βασική δομή inception, συμπίεση γνώσης από διάφορα μεγέθη και συνελίξεων

Ένα άλλο ενδεχόμενο είναι η λήψη αποφάσεων σε συνδυασμό προηγούμενα με επόμενα επίπεδα όπως παρουσιάζεται στο δίκτυο DenseNet [32]. Σε αυτή την περίπτωση

παρατηρούμε ότι υπάρχουν συνδέσεις από προηγούμενα σε επόμενα επίπεδα μεγαλύτερου βάθους. Με αυτόν τον τρόπο μπορούμε να συνδέσουμε χαρακτηριστικά διαφόρων μορφών εκλέπτυνσης που όμως χρειάζεται ιδιαίτερη προσοχή στο να μην αυξηθούν εκθετικά οι συνδέσεις μεταξύ των επιπέδων. Η εικόνα παρακάτω δείχνει μία οπτική του δικτύου DenseNet.



Εικόνα 3-7. Ένα block 5 επιπέδων, με ρυθμό αύξησης $k=4$. Κάθε επίπεδο παίρνει όλα τα προηγούμενα επίπεδα ως εισόδους.

3.2.2.3 Επίπεδο Δειγματοληψίας (Pooling Layer)

Είναι συνηθισμένο πλέον στα περισσότερα νευρωνικά δίκτυα που είναι σε πρώτες θέσεις σε διαγωνισμούς αναγνώρισης εικόνας, το χαρακτηριστικό ότι υφίσταται ένα στρώμα δειγματοληψίας μεταξύ διαδοχικών επιπέδων συνέλιξης. Μία από τις λειτουργίες του είναι η σταδιακή μείωση του μεγέθους αναπαράστασης για τη μείωση του αριθμού των παραμέτρων και των υπολογισμών στο δίκτυο και, ως εκ τούτου, για τη βελτίωση της γενίκευσης με την αύξηση παραμέτρων [33].

Όμως διαισθητικά ο λόγος που τοποθετείτε ένα τέτοιο στρώμα είναι ώστε η εύρεση ενός χαρακτηριστικού να μην είναι συνυφασμένη με την ακριβή θέση. Αυτό επιτυγχάνεται παραδείγματος χάριν με το max pooling μιας και η ύπαρξη ενός χαρακτηριστικού στην περιοχή, ισοδυναμεί με αυξημένη απόκριση του φίλτρου, συνεπώς και με αυξημένη τιμή στο φίλτρο. Αυτό διατηρήσει μόνο τη μέγιστη τιμή σε εκείνη την περιοχή δίχως να μας ενδιαφέρουν οι μη αποκρίσεις.

Είναι σημαντικό να γίνει κατανοητό ότι το στρώμα δειγματοληψίας λειτουργεί ανεξάρτητα σε κάθε κομμάτι βάθους της εισόδου. Αυτό σημαίνει ότι το κάθε φίλτρο που εντοπίζει ένα χαρακτηριστικό μας επιστρέφει μια έξοδο. Δεν είναι επιθυμητό να συγκρίνουμε την μέγιστη τιμή από 2 «ανεξάρτητα» φίλτρα γιατί έτσι ουσιαστικά αντί να εντοπίζουμε χαρακτηριστικά τα μπερδεύουμε. Συνεπώς η δειγματοληψία σε αυτό το επίπεδο μειώνει το χωρικό επίπεδο του πλάτους και του ύψους αντίστοιχα με το μέγεθος του φίλτρου της δειγματοληψίας. Η πιο συνηθισμένη μορφή είναι ένα στρώμα δειγματοληψίας με φίλτρα μεγέθους 2x2 που εφαρμόζεται με ένα βήμα 2 υποδειγμάτων κάθε φέτα βάθους στην είσοδο κατά 2 κατά πλάτος και ύψος, μειώνοντας κατά 75% το πλήθος των ενεργοποιήσεων [34].

Οι διαφορετικές συναρτήσεις που έχουν προταθεί για τη χρησιμοποίηση σε αυτό το επίπεδο είναι οι ακόλουθες :

- **δειγματοληψία μεγίστου**

Στην περίπτωση αυτή από την περιοχή ενδιαφέροντος διατηρείται μόνο η μέγιστη τιμή. Είναι η πλέον χρησιμοποιημένη από όλες τις μεθόδους, γιατί διατηρεί μόνο την τιμή της απόκρισης του φίλτρου σε μία περιοχή που είναι η μέγιστη, άρα δεν επηρεάζεται από το κατά πόσο υπάρχουν και αδιάφορα χαρακτηριστικά στην περιοχή, αρκεί να υπάρχει αυτό που ενδιαφέρει το συγκεκριμένο φίλτρο. Βοηθάει και στην ταχύτερη εκμάθηση του δικτύου αφού όλη η εκμάθηση περνάει από τους κόμβους που έχουν εντοπίσει το συγκεκριμένο χαρακτηριστικό στην εικόνα.

- **δειγματοληψία μέσου όρου**

Στην περίπτωση αυτή επιλέγεται ο μέσος όρος από τα κελιά που βρίσκονται εντός του πυρήνα. Έτσι ομαλοποιείται το αποτέλεσμα που προκύπτει. Αυτό χρησιμοποιείται σε περιπτώσεις που τα χαρακτηριστικά που θέλουμε να υπολογίσουμε προκύπτουν από όλο το εύρος της εικόνας. Χρησιμοποιείται κυρίως σε περιπτώσεις παλινδρόμησης που το δίκτυο δεν επιλέγει κάποια κλάση ως έξοδο αλλά προσεγγίζει μια συνάρτηση και ενδιαφέρει η απόσταση από τη πραγματική συνάρτηση (κάτι που δεν συμβαίνει στην κατηγοριοποίηση).

- **δειγματοληψία μέσου (median)**

Στην περίπτωση αυτή επιλέγεται η μεσαία τιμή, και όχι ο μέσος όρος. Έχει το πλεονέκτημα (το οποίο για ορισμένες περιπτώσεις όπως η εικόνα αποτελεί μάλλον μειονέκτημα) ότι δεν επηρεάζεται από ακραίες αποκρίσεις του φίλτρου. Ένα

μειονέκτημα είναι ότι απαιτεί περισσότερο χρόνο να υπολογιστεί μιας και πρέπει να γίνει ταξινόμηση στις τιμές ώστε να υπολογιστεί η μεσαία και εισάγει πολυπλοκότητα στην υλοποίηση ειδικά για το βήμα του backpropagation αλγόριθμου.

3.2.2.4 Πλήρως συνδεδεμένο επίπεδο (Fully Connected Layer)

Ένα ακόμα συχνά χρησιμοποιούμενο επίπεδο που χρησιμοποιούνταν κατά κόρον σε παλαιότερες αρχιτεκτονικές αλλά παραμένει ακόμα σχετικό είναι το πλήρως συνδεδεμένο επίπεδο. Το επίπεδο αυτό επειδή διατηρεί όλες τις συνδέσεις μεταξύ των εισόδων και των εξόδων του, όταν χρησιμοποιείται είναι επιθυμητό να έχουν παραμείνει το δυνατόν τα πλέον καλά και λίγα χαρακτηριστικά λόγω ιδιαίτερως αργής εκμάθησης. Για αυτό τον λόγο χρησιμοποιείται πλέον μόνο στην έξοδο των συνελκτικών δικτύων όταν και έχουμε υπολογίσει τα χαρακτηριστικά μέσω συνελίξεων. Το επίπεδο αυτό πρέπει να έχει εισόδους όλα τα χαρακτηριστικά που έχουν υπολογιστεί από τα συνελκτικά επίπεδα του δικτύου. Τοποθετούνται τουλάχιστον 2 συνεχόμενα επίπεδα ώστε να μπορεί θεωρητικά να υπολογίσει οποιαδήποτε συνάρτηση.

Το μέγεθος της εξόδου του δικτύου πλήρους επιπέδου έχει όσες κλάσεις είναι επιθυμητό να υπολογιστούν (εκτός και εάν έχει επιλεγεί άλλου είδους κωδικοποίηση). Μια συχνή εξαίρεση είναι στην περίπτωση των 2 κλάσεων που είναι εφικτό να έχουμε μόνο μια έξοδο και ανάλογα με την τιμή εξόδου να αποφασίζουμε σε ποια έξοδο εκτιμά το δίκτυο ότι ανήκει. Στο τέλος αυτού του επιπέδου εφαρμόζεται η softmax function που έχει αναλυθεί στο προηγούμενο κεφάλαιο και μετατρέπει τα αποτελέσματα για όλες τις κλάσεις σε πιθανότητες, με το δίκτυο να αποφασίζει την κλάση με τη μεγαλύτερη πιθανότητα.

Τα τελευταία χρόνια το πλήρως συνδεδεμένο δίκτυο τείνει να αντικατασταθεί από πάλι ένα συνελκτικό δίκτυο αλλά με ελαφρώς διαφορετικά χαρακτηριστικά. Σε αυτήν την περίπτωση δεν επιλέγεται να έχουμε ένα μικρό συνελκτικό δίκτυο αλλά ένα δίκτυο με πυρήνα μεγέθους ίσο με το πλήθος των χαρακτηριστικών που είχαμε ως είσοδο στο πλήρως συνδεδεμένο δίκτυο. Για να μπορέσουμε να έχουμε αντίστοιχα 2 επίπεδα, αρχικά περνάμε τις εξόδους από ένα δίκτυο συνελίξεων 1×1 οπότε έχουμε στην είσοδο του συνελκτικού επιπέδου μια μη γραμμικότητα να εισάγεται. Έτσι σαν τελική έξοδο έχουμε ένα συνελκτικό επίπεδο του οποίου ο «όγκος» είναι 1×1 (πλήθος κλάσεων).

Το πλεονέκτημα που μας δίνει η παραπάνω μέθοδος σε σχέση με το πλήρως συνδεδεμένο επίπεδο είναι ότι αφού όλα τα επίπεδα του δικτύου αποτελούνται από

συνελκτικά και δειγματοληπτικά επίπεδα τα οποία είναι «ανθεκτικά» στην αλλαγή του μεγέθους εισόδου της εικόνας. Δηλαδή εάν έχουμε ένα εκπαιδευμένο δίκτυο το οποίο έχει εκπαιδευτεί για εικόνες 10x10x3 μπορούμε να του δώσουμε ως είσοδο εικόνες 20x20x3 το οποίο θα παράξει ως εξόδους μεγέθους 2x2x(πλήθος κλάσεων). Οπότε έχουμε 4 τιμές για κάθε έξοδο και μπορούμε να εκπαιδεύσουμε ένα ακόμα επίπεδο συνέλιξης 2x2x(πλήθος κλάσεων) ώστε να καταλήξουμε στο επιθυμητό 1x1x(πλήθος κλάσεων), έχοντας χρησιμοποιήσει και τα βάρη που έχει εκπαιδευτεί και για το τελευταίο επίπεδο. Εάν ήταν πλήρως συνδεδεμένο επίπεδο τότε τα βάρη του παύουν να έχουν νόημα και πρέπει να επανεκπαιδευτούν χαραμίζοντας τον χρόνο για την προηγούμενη εκπαίδευση του δικτύου. Στην περίπτωση μας δεν χρησιμοποιήθηκαν τα βάρη του πλήρως συνδεδεμένου επιπέδου, γιατί οι κλάσεις για τις οποίες έχει εκπαιδευτεί είναι φυσικές εικόνες και απέχουν πολύ από τις ιατρικές εικόνες που επιθυμούμε να ταξινομήσουμε στην παρούσα εργασία.

3.3 Αξιολόγηση αρχιτεκτονικής

Μετά την εκτέλεση των πειραμάτων χρειάζεται να αξιολογηθεί το ποια αρχιτεκτονική δίνει τα καλύτερα αποτελέσματα και να μπορούμε να συγκρίνουμε τις τεχνικές που δοκιμάστηκαν. Για αυτό το λόγο επιλέχθηκε η χρησιμοποίηση των μετρικών $F1$ [35][36] και BAC [37]. Για τον υπολογισμό των παραπάνω μετρικών χρειάζεται να έχουμε υπολογίσει 2 άλλες μετρικές την ακρίβεια (precision) και την ανάκληση (recall).

Η λογική πίσω από την precision είναι πόσα είναι όντως σωστά μεταξύ αυτών που ο αλγόριθμος μας χαρακτήρισε ως σωστά. Υπολογίζεται ως

$$precision = True\ Positive / (True\ Positive + False\ Positive)$$

Δηλαδή εάν ένας αλγόριθμος εμφάνιζε ότι στα σωστά είναι μόνο ένα στοιχείο αυτό που είχε μεγαλύτερο σκορ τότε πιθανότατα, θα είχε precision 1 αφού τα FP(False Positive) θα ήταν 0. Συνεπώς δεν είναι δυνατόν να χρησιμοποιηθεί από μόνη της αυτή η μετρική για την αξιολόγηση ενός αλγόριθμου μηχανικής μάθησης.

Αντίστοιχα η recall υπολογίζει το πόσα σωστά μας επέστρεψε ο αλγόριθμος σε σχέση με όλα τα σωστά στο αρχικό δείγμα. Ο τύπος που υπολογίζεται είναι ο ακόλουθος:

$$recall = True\ Positive / (True\ Positive + False\ Negative)$$

Αντίστοιχα σε αυτή την περίπτωση ο αλγόριθμος εάν υποθέσει ότι όλα τα δείγματα ανήκουν στην κατηγορία ενδιαφέροντος, τότε το recall είναι 1. Οπότε ούτε η

συγκεκριμένη μετρική μπορεί να χρησιμοποιηθεί από μόνη της αυτή η μετρική για την αξιολόγηση ενός αλγόριθμου μηχανικής μάθησης. Αν παρατηρήσουμε όμως οι 2 παραπάνω μετρικές μας βγάζουν τέλεια αποτελέσματα σε αντιδιαμετρικά ακραίες περιπτώσεις και συνεπώς ένας συνδυασμός τους μπορεί να χρησιμοποιηθεί για την αξιολόγηση τους.

Ένας τέτοιος συνδυασμός είναι και οι μετρική F1. Συγκεκριμένα η F1 (η αλλιώς F-measure) μετρική υπολογίζεται ως:

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

Η παραπάνω μετρική παίρνει τιμές από 0 έως 1, και λαμβάνει υπόψιν και τις 2 καταστάσεις που προαναφέρθηκαν. Χρησιμοποιείται ευρέως στη βιβλιογραφία για την αξιολόγηση μεθόδων μηχανικής μάθησης και λόγω της σχετικά απλής ερμηνείας της.

Η άλλη μετρική που χρησιμοποιήθηκε στην παρούσα εργασία είναι η εξισοροπημένη ακρίβεια (Balanced Accuracy ή BAC). Ο λόγος που χρησιμοποιήσαμε αυτή τη μετρική είναι επειδή οι δύο κλάσεις εισόδου δεν είναι ισάξια αναπαρεστημένες και κατά συνέπεια δεν θα ήταν ορθό να κρίνουμε έναν αλγόριθμο μόνο με την χρήση της «απλής» ακρίβειας (accuracy). Αυτό συμβαίνει γιατί εάν όπως στην περίπτωση μας έχουμε δύο κλάσεις που έχουμε τετραπλάσιες εικόνες για τη μια κλάση, θα μπορούσαμε με έναν απλό ταξινομητή που αποφάσιζε την πλέον πολυπληθή κλάση να έχουμε ακρίβεια της τάξης του 75% αντί για το 50% που διαισθητικά θα θέλαμε. Για αυτό το λόγο χρησιμοποιήθηκε η BAC, μιας και ποινικοποιεί αντιστρόφως ανάλογα με το μέγεθος της κλάσης τα σφάλματα κάθε εισόδου.

4. ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ

4.1 Βασικές Ιδιότητες Ιατρικής Εικόνας

Στην επιλογή ιατρικών εικόνων υπάρχει μεγάλη ποικιλομορφία στο είδος των εικόνων, το μέγεθος, την ανάλυση καθώς και στην ύπαρξη ή όχι χρώματος. Επίσης σε πολλές περιπτώσεις έχουμε υψηλής συσχέτισης εικόνες, π.χ. σε μαγνητικές τομογραφίες δεν μπορούμε να θεωρήσουμε την κάθε τομή ανεξάρτητη εικόνα (για τον ίδιο ασθενή). Στην παρούσα εργασία επιλέξαμε να ασχοληθούμε με εικόνες που αφορούν σε καρκίνο του μαστού. Η επιλογή έγινε έχοντας πολλαπλούς σκοπούς και αναλύεται στην παρακάτω παράγραφο.

Αρχικά θέλαμε το ιατρικό πρόβλημα με το οποίο θα ασχοληθούμε να έχει έντονο ιατρικό ενδιαφέρον, τόσο για την ιατρική κοινότητα αλλά και ει δυνατόν για σημαντικό πλήθος ασθενών. Επίσης συμπεριλαμβανόταν στους σκοπούς μας η ύπαρξη ερευνών ώστε να έχουμε ένα εξωτερικό σημείο αναφοράς για την ποιότητα των αποτελεσμάτων που θα λαμβάναμε. Επίσης λόγω της φύσης των νευρωνικών δικτύων απαιτούνταν ένα σύνολο δεδομένων το οποίο θα είχε σημαντικό πλήθος εικόνων ώστε να καταστεί δυνατή η εκπαίδευση του δικτύου, μιας και το πλήθος των εικόνων που απαιτούνται για αξιόλογα αποτελέσματα είναι αρκετά υψηλό.

4.2 Στοιχεία του Συνόλου Δεδομένων

Όλα τα παραπάνω συνηγόρησαν στην επιλογή για την παρούσα διπλωματική εργασία του συνόλου δεδομένων για τον καρκίνο του μαστού και συγκεκριμένα για το πορογενές διηθητικό καρκίνωμα (Invasive Ductal Carcinoma) ή για συντομία IDC . Το IDC αποτελεί τον πλέον συχνό φαινότυπο για τον καρκίνο του μαστού, ενώ αποτελεί περίπου το 80% των περιπτώσεων που εμφανίζεται [38][39]. Η συγκεκριμένη ασθένεια ταυτοποιείται από ειδικούς μέσω οπτικής εξέτασης διαφανειών οι οποίες έχουν χρωματιστεί με αιματοξυλίνη και ηωσίνη. Η αξιολόγηση έγινε από ειδικούς ώστε σε κάθε εικόνα να χαρακτηριστούν τα μέρη του ιστού στα οποία εμφανίζεται διηθητικός καρκίνος και σε ποια ο ιστός είναι ανεπηρέαστος.

Για τη συγκεκριμένη ασθένεια υπάρχει ένα αρκετά ικανοποιητικό από άποψη μεγέθους αλλά και ποιότητας εικόνων σύνολο δεδομένων. το οποίο έχει εικόνες από 162 ασθενείς. Πρόσβαση στη βάση δεδομένων και τα αρχεία μπορεί να πραγματοποιηθεί και από τον ακόλουθο σύνδεσμο: <https://www.kaggle.com/paultimothymooney/predict-idc-in-breast->

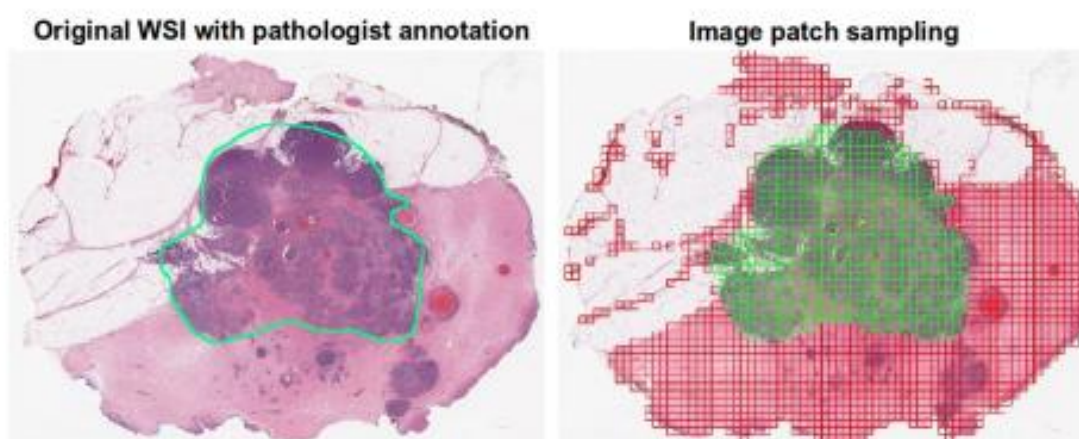
[cancer-histology-images/data](#) [38][40]. Το σύνολο των δεδομένων συλλέχθηκαν στο Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey [40]. Όλα τα slides ψηφιοποιήθηκαν σε 40x μεγέθυνση (0.25 μm/pixel resolution). Λόγω του γεγονότος ότι οι εικόνες είναι πολύ μεγάλου μεγέθους επειδή αρχικά χρησιμοποιήθηκαν για τον τοπικό εντοπισμό όπου εμφανίστηκε ο όγκος, στην εργασία συλλογής κάθε λήψης φωτογραφιών πλήρους μεγέθους (Whole Slide Imaging ή WSI) υποβλήθηκε σε υπό-δειγματοληψία με λόγο 16 προς 1 σε τελική ανάλυση 4 μm / pixel. Ο χαρακτηρισμός των θετικών περιοχών σε IDC έγινε από εξειδικευμένο ιατρό «χειρωνακτικά» όπως φαίνεται και στις εικόνες που παρατίθενται παρακάτω. Οι σημειώσεις των ειδικών έγιναν ελεύθερα προσβάσιμες μέσω του λογισμικού Imagescope της εταιρείας Aperio. Οι σημειώσεις του ειδικού ιατρού έγιναν σε ανάλυση εικόνας με έως 2x zoom, οπότε και δεν είναι απολύτως λεπτομερής, κάτι που δυσκολεύει τη εργασία, των αλγορίθμων κατηγοριοποίησης.

Το συνολικό πλήθος των εικόνων που υπάρχουν στο σύνολο δεδομένων είναι 277.524 (198.738 αρνητικές και 78.786 θετικές). Δυστυχώς όμως αναφέρεται ότι η επιλογή των ασθενών για τις εκάστοτε περιπτώσεις εκπαίδευσης έγινε τυχαία. Συνεπώς είναι αδύνατο να αναπαράγουμε ακριβώς την αναλογία των πειραμάτων όμως για τη καλύτερη δυνατή σύγκριση χρησιμοποιήθηκαν και στην παρούσα εργασία οι αντίστοιχη αναλογία για το σύνολο εκπαίδευσης, όπου επιλέξαμε το ικανό ποσοστό εικόνων για την εκπαίδευση και τον έλεγχο του μοντέλου. Στην περίπτωση μας όμως επιλέξαμε να δώσουμε μεγαλύτερη σημασία στο test set γιατί κυρίαρχος σκοπός μας ήταν να αντιλαμβανόμαστε κατά τη διάρκεια της εκπαίδευσης τη συμπεριφορά του ίδιου του αλγόριθμου. Τα τελικά νούμερα που χρησιμοποιήθηκαν στην παρούσα εργασία είναι τα ακόλουθα: 72.675 εικόνες (20.914 θετικές και 51.761 αρνητικές) για εκπαίδευση, 11.561 εικόνες (3.327 θετικές και 8.234 αρνητικές) για τελική αξιολόγηση και 193.291 (54.545 θετικές και 138.743 αρνητικές) για σετ δοκιμών, καθώς και για δοκιμές για έλεγχο παραμέτρων πριν την κάθε τροποποίηση. Ο εντοπισμός των περιοχών με ακρίβεια στα οποία εντοπίζεται προσβεβλημένος ιστός είναι μια διαδικασία απαιτητική σε χρόνο αλλά και σε πόρους μιας και χρειάζεται υψηλής εξειδίκευσης προσωπικό και με τις αρχικές φωτογραφίες να είναι σε υψηλή ανάλυση 3800x2950x3. Ο ακριβής διαχωρισμός των περιοχών αν και χρονοβόρος είναι κρίσιμος σε ότι αφορά το τελικό αποτέλεσμα που θέλουμε να πετύχουμε μιας και είναι πολύ σημαντικό να εκπαιδευτεί το δίκτυο με σωστά δεδομένα. Παρόλα αυτά η ακριβής οριοθέτηση των περιοχών σε επίπεδο pixel ακόμα και από ειδικούς αποτελεί πρόκληση και είναι σχεδόν αδύνατο να είναι απολύτως ακριβής.

Συνεπώς είναι βέβαιο ότι υπάρχουν και κάποιες εικόνες οι οποίες είναι εσφαλμένα ταξινομημένες στο σύνολο δεδομένων, όμως το ποσοστό αυτό είναι αρκετά μικρό. Επίσης στην εργασία αναφοράς [38] οι εικόνες που έχουν έστω και ένα μικρό ποσοστό καρκινικού ιστού τοποθετήθηκαν στην κατηγορία καρκίνου, το οποίο εν μέρει είναι λογικό αν και μάλλον δυσκολεύει την «κατανόηση» του δικτύου.

Επειδή οι εικόνες που έχουμε είναι όλες από ασθενείς, η κατηγορία της μη ύπαρξης καρκίνου αποτελείται από εικόνες οι οποίες βρισκόταν σε ιστό ασθενή αλλά που δεν είχαν προσβληθεί. Συνεπώς ο κάθε ασθενής συνεισφέρει και στις 2 κατηγορίες θετικές-αρνητικές εικόνες. Τα κομμάτια που δεν συνεισφέρουν στην αναγνώριση της ασθένειας δεν εισάγονται στο σύνολο εκπαίδευσης σε καμία από τις κατηγορίες. Το πλήθος των εικόνων που αντιστοιχούν σε προσβεβλημένο από την ασθένεια για κάθε ασθενή προφανώς διαφέρει.

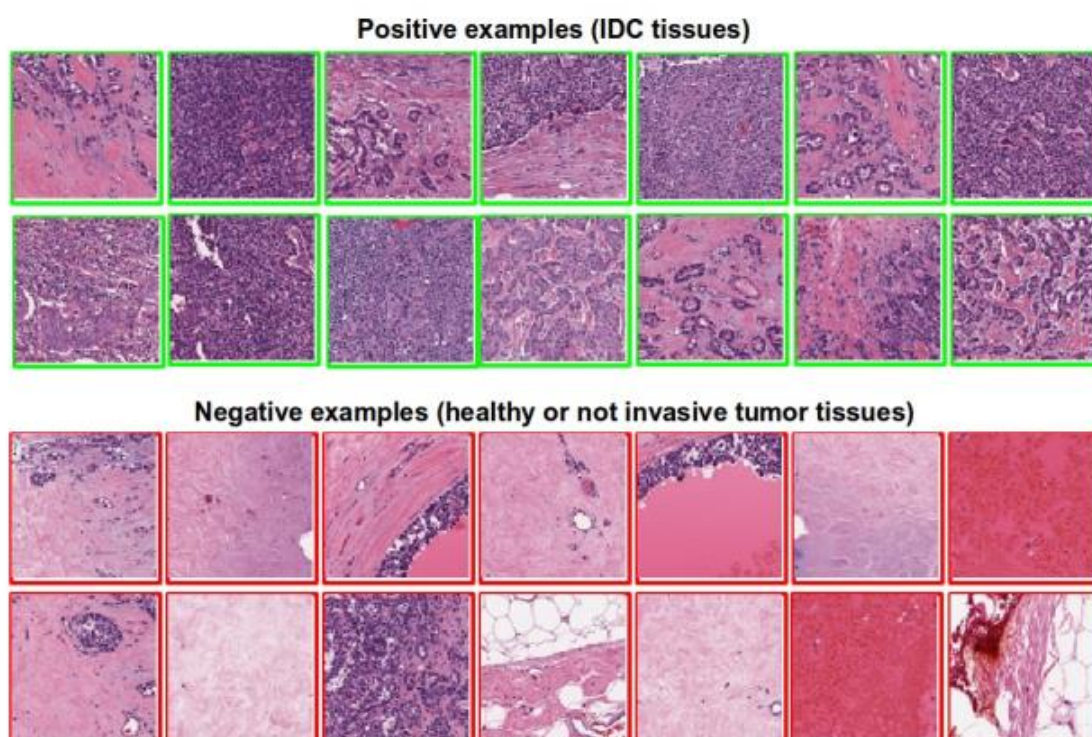
Επίσης στις περισσότερες εικόνες το μεγαλύτερο ποσοστό της εικόνας αποτελείται από υγιή ιστό, συνεπώς οι δύο κατηγορίες είναι ανισομερείς σε αναλογία της τάξης του 4 υγιής προς 1 με στοιχεία κακοήθους καρκίνου. Επίσης περιοχές οι οποίες είναι αδιάφορες ιστολογικά δεν κατατάχτηκαν σε καμία από τις 2 κατηγορίες, για τη μείωση του θορύβου στη διακριτική ικανότητα του αλγορίθμου. Αυτό γίνεται αντιληπτό από την παρακάτω εικόνα όπου οι κόκκινες περιοχές αποτελούν τμήματα μη ύπαρξης καρκινικού ιστού ή καλοήγη ιστό ενώ τα πράσινα αποτελούν παρουσία καρκινικού ιστού κακοήγη. Τα μη χρωματισμένα δεν συμπεριλαμβάνονται στην έρευνα μιας και δεν αποτελούν μέρος της έρευνας.



Εικόνα 4-1. (αριστερά) Χαρακτηρισμός περιοχής εικόνας από ειδικό, (δεξιά) κομμάτια της εικόνας που έχουν επιλεγεί ως αρνητικά (κόκκινα) και θετικά (πράσινα)

Το σύνολο δεδομένων που χρησιμοποιούμε στην παρούσα εργασία αρχικά χρησιμοποιήθηκε [38] για την κατηγοριοποίηση τμημάτων των αρχικών εικόνων 50x50x3 με τελικό σκοπό της εργασίας να είναι η οριοθέτηση των περιοχών που εμφανίζεται πορώδης διηθητικός καρκίνος, όπως φαίνεται και στην παραπάνω εικόνα.

Συνεπώς μια βελτίωση στο επίπεδο ταξινόμησης των εικόνων αυτομάτως θα βελτιώσει και την οριοθέτηση των περιοχών αν και αυτό δεν αποτελεί βασικό στόχο της παρούσης εργασίας. Στην παρακάτω εικόνα παρατηρούμε κάποια από τα πλέον χαρακτηριστικά παραδείγματα των δύο κατηγοριών.



Εικόνα 4-2. Χαρακτηριστικές εικόνες, (πάνω) μέρος από εικόνες που έχουν αναγνωρισθεί ως θετικά σε IDC, (κάτω) μέρος από εικόνες που έχουν χαρακτηριστεί ως αρνητικές στην ύπαρξη IDC

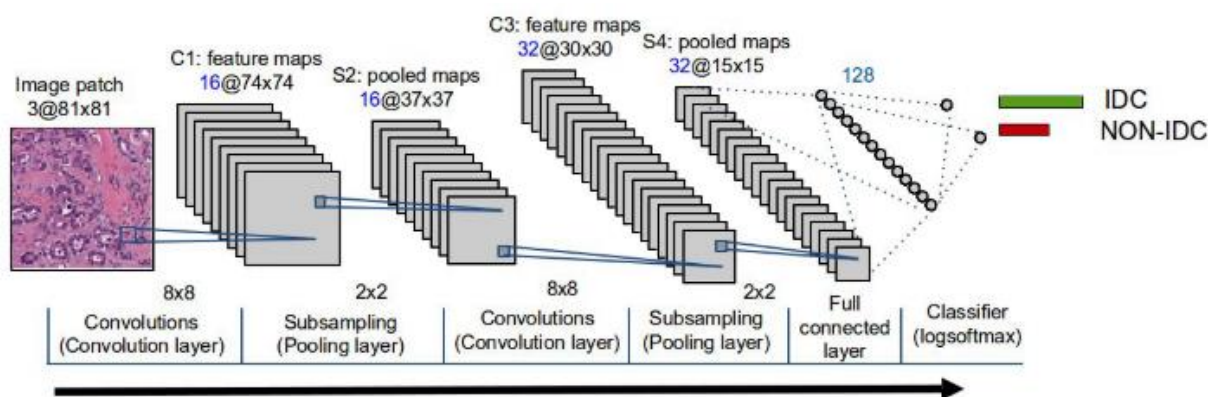
5. ΜΕΘΟΔΟΛΟΓΙΑ

Στην παρούσα εργασία σκοπός μας ήταν να συγκριθούν διάφορες μορφές συνελκτικών νευρωνικών δικτύων και να συγκριθούν με άλλες μεθοδολογίες που έχουν ήδη ελεγχθεί για το συγκεκριμένο σύνολο δεδομένων στη βιβλιογραφία. Επίσης, εξετάζουμε με κριτική ματιά τις μεθοδολογίες που είναι πιθανότερο στο μέλλον σε αντίστοιχα σύνολα δεδομένων να έχουν καλύτερα αποτελέσματα. Στο παρόν κεφάλαιο θα εξηγήσουμε τα πειράματα που έλαβαν χώρα, και κυρίως το λόγο για τον οποίο έγινε η επιλογή των συγκεκριμένων μεθοδολογιών. Τελικός σκοπός είναι να μπορούν να εξαχθούν συμπεράσματα για τις ενδεικνυόμενες μεθοδολογίες γενικότερα, σε ιατρικά σύνολα δεδομένων.

5.1 Απλό Συνελκτικό Δίκτυο

Αρχικά για να ορίσουμε ένα baseline για τα πειράματα θέλαμε να υπάρχει ένα απλό συνελκτικό δίκτυο ώστε να δούμε οι υπόλοιπες υλοποιήσεις κατά πόσο βελτιώνουν την κατηγοριοποίηση. Επειδή στην εργασία αναφοράς [38] της παρούσης εργασίας έχει ήδη γίνει υλοποίηση ενός CNN που το θεωρούμε και ως baseline για την παρούσα εργασία δεν κρίθηκε απαραίτητο η επανάληψη της υλοποίησης. Οπότε χρησιμοποιήθηκαν τα αποτελέσματα που παρουσιάζονται στην εργασία που αναφέρθηκε παραπάνω. Η παραπάνω ομάδα έχει ασχοληθεί επισταμένως με το πρόβλημα που πραγματευόμαστε όπως μπορεί να φανεί και από προηγούμενες δημοσιεύσεις [41]. Στη συγκεκριμένη εργασία εκτός του δικτύου CNN εξετάστηκαν και άλλοι αλγόριθμοι μηχανικής μάθησης όπως τυχαία δάση που θα εξετάσουμε σε επόμενο κεφάλαιο.

Για την υλοποίηση του απλού συνελκτικού δικτύου επιλέξαν ένα δίκτυο με 2 συνελκτικά και ένα πλήρες επίπεδο, τα χαρακτηριστικά του οποίου φαίνονται στην παρακάτω εικόνα.

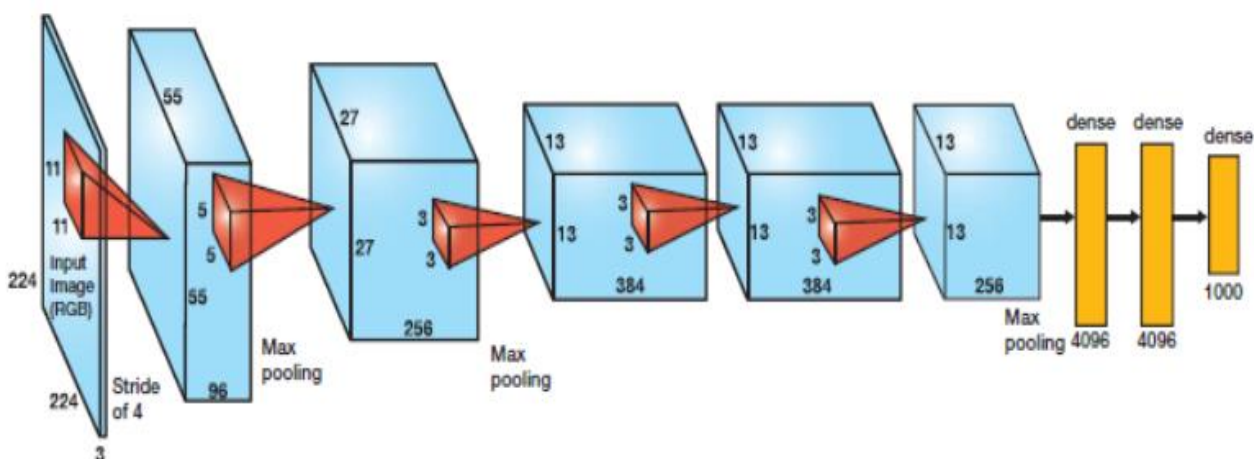


Εικόνα 5-1. Δομή Συνελκτικού Δικτύου

Στο 1^ο επίπεδο συνέλιξης με πυρήνα 8x8, στη συνέχεια max pooling με 2x2 πυρήνα, στη συνέχεια 2^ο επίπεδο συνέλιξεων με πυρήνα πάλι 8x8, στη συνέχεια πάλι επίπεδο max pooling με 2x2. Τελικά πλήρως συνδεδεμένο δίκτυο με 128 features, και στη συνέχεια log-softmax συνάρτηση για την επιλογή της τελικής απόφασης του δικτύου. Στο πρώτο επίπεδο, υπολογίζει 16, στο 2^ο 32 και ως είσοδο στο πλήρες δίκτυο 128 feature maps. Σε όλες τις περιπτώσεις εφαρμόστηκε stride 1 και μηδενικό padding με τυχαία αρχικοποίηση των βαρών από κανονική κατανομή.

5.2 Η αρχιτεκτονική του δικτύου VGG16

Αρχικά, θα διευκολύνει αρκετά την κατανόηση της δομής για το δίκτυο VGG16 (Visual Geometry Group 16) η εξήγηση της δομής του προκατόχου του και πηγή έμπνευσης της αρχιτεκτονικής του, το CNN AlexNet [29][39]. Η αρχιτεκτονική του προκατόχου του VGG16 [42] είναι η εξέλιξη του AlexNet, του δικτύου που κατ' ουσίαν έβαλε τα συνελκτικά νευρωνικά δίκτυα ξανά στο προσκήνιο της έρευνας στο επίπεδο της μηχανικής μάθησης. Η δομή του φαίνεται στο παρακάτω σχήμα, καθώς και το μέγεθος των πυρήνων του κάθε επιπέδου. Βασικό χαρακτηριστικό του δικτύου ήταν η ύπαρξη αρκετά μεγάλων πυρήνων 11x11 και stride 4 στο πρώτο επίπεδο. Επίσης στα τελευταία επίπεδα υπάρχουν πλήρως συνδεδεμένα επίπεδα και το επίπεδο εξόδου για 1000 κατηγορίες που χρησιμοποιούσε για την κατηγοριοποίηση εικόνων αντικειμένων καθημερινότητας όπως σκύλος, αεροπλάνο και άλλες αντίστοιχες.



Εικόνα 5-2. Δομή της αρχιτεκτονικής του συνελκτικού δικτύου AlexNet, προκατόχο του VGG16

Το παραπάνω δίκτυο έχει ως μέγεθος εικόνας εισόδου (224,224,3) το οποίο αποτελεί πολύ μεγαλύτερο από το σύνολο δεδομένων που χρησιμοποιήσαμε στην παρούσα εργασία. Στις προδιαγραφές του δικτύου αναφέρεται ότι οι εικόνες που μπορεί να δεχτεί

ως είσοδο είναι κατ' ελάχιστον (150x150x3) έτσι ώστε στο τελευταίο επίπεδο που αποτελεί είσοδο στα πλήρως συνδεδεμένα επίπεδα να μην εκμηδενίζεται μέσω των pooling layers. Επίσης σε αυτό το δίκτυο χρησιμοποιήθηκαν relu συναρτήσεις ενεργοποίησης αντί για τη σιγμοειδή ή την υπερβολική εφαπτομένη, πράγμα καινοτόμο και ριζοσπαστικό για την εποχή του. Επίσης εισήγαγε τη χρήση της τεχνικής του dropout που έχει ήδη εξηγηθεί, με $p=0.5$, μετά από κάθε πλήρως συνδεδεμένο επίπεδο στο επίπεδο εκπαίδευσης.

5.2.1 Η δομή του VGG16

Το VGG16 δίκτυο, που προέρχεται από την ομάδα VGG του πανεπιστημίου της Οξφόρδης [43] υπάρχει αντίστοιχη φιλοσοφία με το AlexNet με κάποιες βελτιώσεις όσον αφορά κυρίως τη δομή του δικτύου.

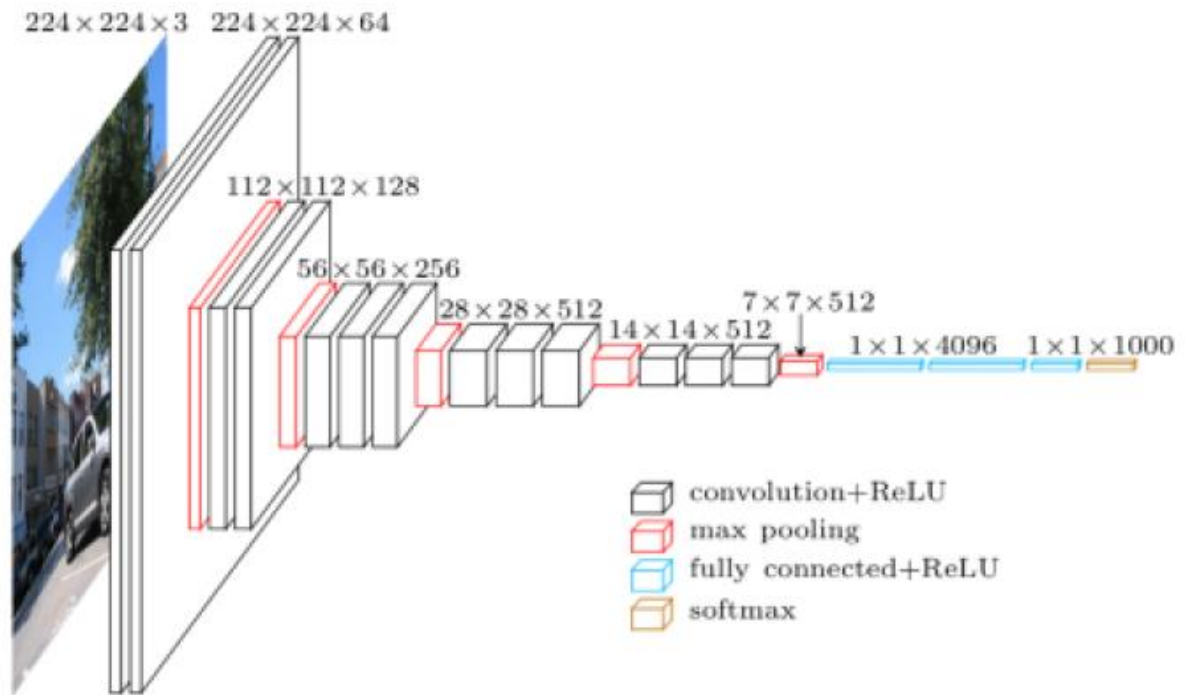
Αρχικά εγκαταλείφθηκε η χρήση μεγάλων πυρήνων συνέλιξης (11x11 και 5x5) και στη θέση τους εισήχθησαν μόνο 3x3 φίλτρα σε μεγαλύτερο βάθος δικτύου. Αυτό γίνεται κατανοητό εάν θεωρήσουμε 3 επίπεδα 3x3 με stride 1 τα οποία δημιουργούν ένα receptive field μεγέθους 7 και το πλήθος των παραμέτρων είναι $3 \cdot (3 \cdot 3 \cdot \text{βαθος}^2)$ ενώ με το αντίστοιχο 7x7 receptive field σε ένα επίπεδο έχουμε $49 \cdot \text{βάθος}^2$ που αποτελεί αρκετά μεγαλύτερο πλήθος παραμέτρων, το οποίο εκτινάσσεται όταν μιλάμε για receptive field μεγέθους 11x11 που υπάρχει στο AlexNet. Επίσης με τα 3x3 φίλτρα με stride 1 σε σχέση με τα 11x11 φίλτρα με stride 4 μπορούν να διατηρηθούν πολύ πιο λεπτομερής πληροφορίες για την αρχική εικόνα σε επόμενα επίπεδα, πράγμα πολύ επιθυμητό, ειδικά για ιατρικές εικόνες όπου η υφή και η λεπτομέρεια είναι αυτή που διαφοροποιεί τις κατηγορίες, σε αντίθεση με την αναζήτηση μεγάλων δομών όπως αντικείμενα ή ζώα.

Για να είμαστε απόλυτα σαφείς το δίκτυο VGG16 εκπαιδεύτηκε πάνω στο σύνολο δεδομένων ImageNet, το οποίο έχει 14.187.122 εικόνες από 1000 κατηγορίες από αντικείμενα του φυσικού αλλά και του τεχνητού μας περιβάλλοντος. Έχει χρησιμοποιηθεί από πάρα πολλές εργασίες γενικότερα, αλλά και για ιατρικά δεδομένα, και αποτελεί σημείο αναφοράς για την εκπαίδευση και σύγκριση αρχιτεκτονικών συνελκτικών νευρωνικών δικτύων, αλλά και άλλων τεχνικών μηχανικής μάθησης [44].

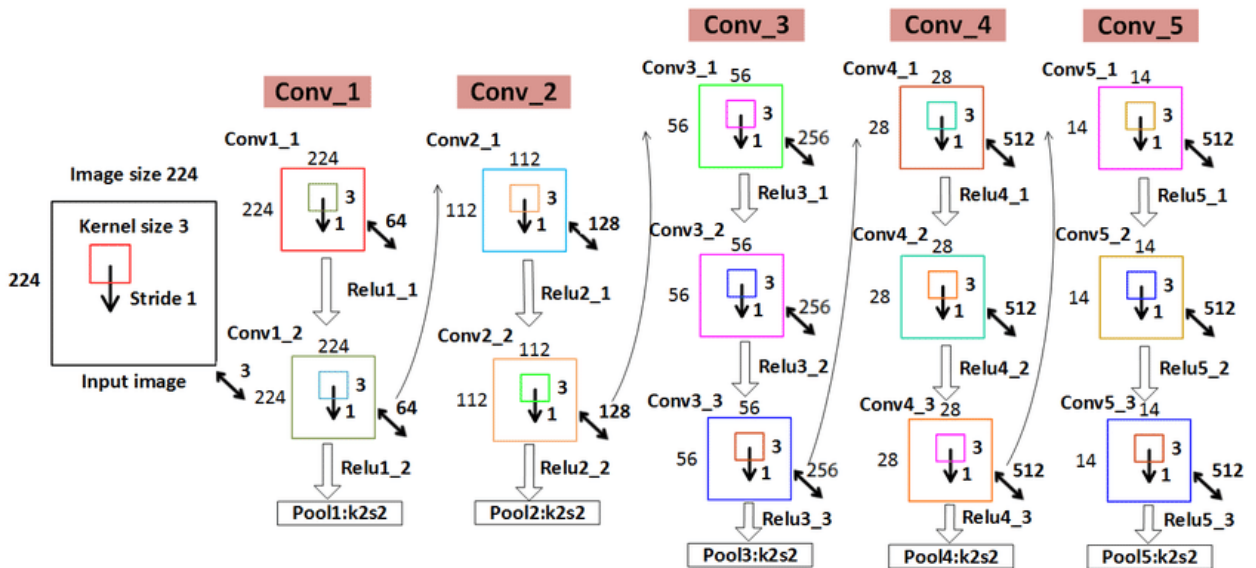
Στη δομή του VGG16 υπήρχε το ριζοσπαστικό το οποίο πλέον έχει διαδοθεί σε άλλα δίκτυα (αναφορά), ότι κάνει πολλαπλές συνέλιξεις στο ίδιο επίπεδο με διαφορετικούς πυρήνες, πριν να εισάγει το pooling layer 2x2 (το οποίο μειώνει το μέγεθος για το επόμενο επίπεδο). Αυτό έχει τα παρακάτω αποτελέσματα για το δίκτυο το οποίο προκύπτει.

Αρχικά αυξάνει τη μη γραμμικότητα αφού σε ένα επίπεδο υπάρχουν 2 ή και 3 μη γραμμικές συναρτήσεις που εφαρμόζονται. Το οποίο αρχικά μπορεί να μην φαίνεται ιδιαίτερα σημαντικό αλλά επειδή πλέον η χρήση της relu είναι καθολική στη δομή του δικτύου, σημαίνει ότι η μη γραμμικότητα είναι αρκετά μικρότερη από άλλες συναρτήσεις όπως η σιγμοειδής ή η υπερβολική εφαπτομένη ως συναρτήσεις ενεργοποίησης κάτι που μετριάξει πλέον αυτό το πρόβλημα. Μια επιπλέον προβληματική κατάσταση είναι ότι ένας νευρώνας που αρχικά έχει πολύ μικρή τιμή μπορεί να παραμείνει ανενεργός για πάντα [45]. Συνεπώς για να μπορέσει να αναπαραστήσει τις μη γραμμικότητες είναι πολύ χρήσιμη η πολλαπλή χρήση μη γραμμικότητας. Προφανώς η εκπαίδευση όταν υπάρχουν 3 επίπεδα συνελίξεων αντί για ένα καθυστερεί αρκετά περισσότερο και ο αριθμός των παραμέτρων του συστήματος αυξάνεται. Αυτός είναι και ο βασικός λόγος που η συγκεκριμένη αρχιτεκτονική δικτύου μειονεκτεί σε εφαρμογές που χρειάζονται είτε online ενημέρωση των βαρών είτε έχουν μικρή χωρητικότητα μιας και μόνο οι παράμετροι απαιτούν περισσότερο του 0.5 GB. Το συνολικό πλήθος των παραμέτρων για το VGG16 είναι λίγο παραπάνω από 138 εκατομμύρια. Συνεπώς καθίσταται η εκπαίδευση του δικτύου να γίνει με GPU με αυξημένη μνήμη, διότι χρειάζεται και οι παράμετροι αλλά και οι μεταβολές των παραμέτρων καθώς επίσης και το σύνολο των δεδομένων μας να χωράει στη μνήμη του υπολογιστή ή της κάρτας γραφικών.

Όλα τα παραπάνω όμως με τη σημερινή τεχνολογία είναι διαχειρίσιμα και πλέον μπορούμε να εκπαιδεύσουμε σε «αποδεκτό» χρόνο ένα μοντέλο που να εξάγει ικανοποιητικά αποτελέσματα. Το πλεονέκτημα που έχουμε σε αυτό το δίκτυο είναι η ομοιομορφία στις πράξεις συνέλιξης που συμβαίνουν. Όπως μπορούμε να δούμε και στην παρακάτω εικόνα της δομής του δικτύου η μόνη πράξη συνέλιξης που λαμβάνει χώρα είναι συνέλιξη με πυρήνες 3x3. Επίσης στην παρακάτω εικόνα γίνεται εμφανές και η μεταβολή του «όγκου» των χαρακτηριστικών που μεταφέρονται από επίπεδο και η «μετατροπή» του ύψους και πλάτους σε βάθος.



Εικόνα 5-3. Αναπαράσταση δικτύου VGG16, με χρήση "όγκων" χαρακτηριστικών, και η πορεία των χαρακτηριστικών μέχρι την κατηγοριοποίηση



Εικόνα 5-4. Αναπαράσταση όλων των πράξεων με παρουσίαση και των λεπτομερειών του κάθε επιπέδου για το δίκτυο VGG16, για τα συνελκτικά επίπεδα.

https://www.researchgate.net/figure/Architecture-of-truncated-VGG-16_fig3_321065008

Παρατηρούμε στη δομή του δικτύου ότι στα δύο πρώτα επίπεδα (όπου ως επίπεδο ορίζουμε οποιονδήποτε μετασχηματισμό υπάρχει ανάμεσα σε pooling layers) βρίσκονται

2 συνελίξεις, ενώ στα 3 τελευταία επίπεδα υπάρχουν 3 συνελίξεις. Αυτό γιατί τα αρχικά επίπεδα, έχοντας σχετικά μικρό βάθος και πολύ βασικά χαρακτηριστικά δεν απαιτούν μεγάλη συσχέτιση στα χαρακτηριστικά τους. Σε αντίθεση με τα τελευταία επίπεδα που πλέον έχουν χαρακτηριστικά που περιέχουν αυξημένη πληροφορία από μεγαλύτερες περιοχές της αρχικής εικόνας.

Παρατηρώντας τη δομή του δικτύου και επειδή έχει το πλεονέκτημα ότι είναι αμιγώς γραμμική σε αντίθεση με άλλες αρχιτεκτονικές (όπως η Inception) βασικός σκοπός της διπλωματικής είναι η διερεύνηση του δικτύου καθώς και μεταβολών της δομής του δικτύου.

5.2.2 Υλικό και λογισμικό εκτέλεσης πειραμάτων

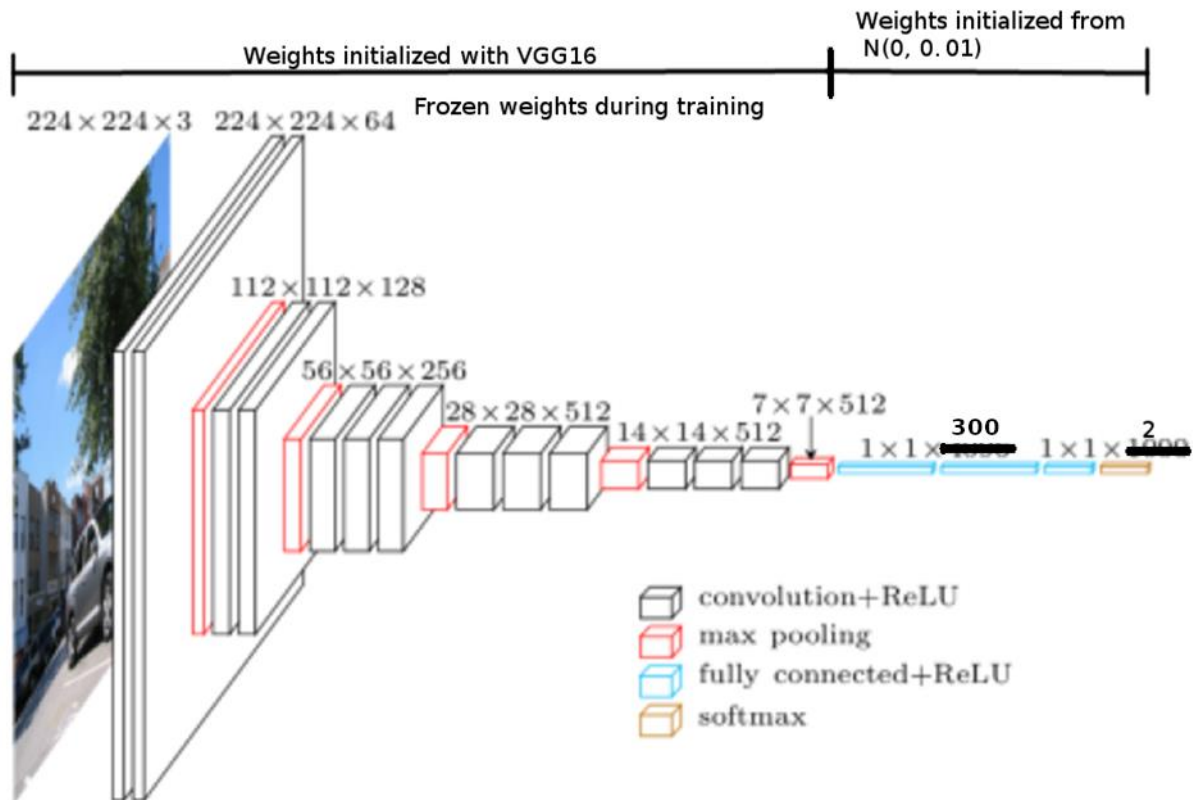
Τα πειράματα έτρεξαν σε λειτουργικό σύστημα Windows 10 v.1742. Ο επεξεργαστής που χρησιμοποιήθηκε είναι ο Intel Core i7-8700K @3.70 GHz, overclocked στα 4.6 GHz. Η μνήμη του συστήματος ήταν 2x 8GB @ 3.2 GHz, και η κάρτα γραφικών 1070 Ti Gigabyte NVIDIA με 8 GB DDR5. Το λογισμικό το οποίο χρειαστήκαμε ήταν η σουίτα anaconda με τη χρήση της python 3.7, τη βιβλιοθήκη keras στην έκδοση 2.15 και τη βιβλιοθήκη tensorflow στην έκδοση 1.7.1 . Για τα διαγράμματα χρησιμοποιήθηκε η βιβλιοθήκη matplotlib στην έκδοση 2.2.2 . Όλες οι παραπάνω βιβλιοθήκες που αναφέρονται έχουν και πολλές άλλες εξαρτήσεις από άλλες βιβλιοθήκες. Δεν κρίνεται σκόπιμο να αναφερθούν όλες αναλυτικά, επειδή με την εγκατάσταση είτε με τη χρήση του conda είτε με τη χρήση του pip, εγκαθίστανται αυτόματα.

5.2.3 Δομή φακέλων

Στον κώδικα που ακολουθεί την εργασία υπάρχει μια μεταβλητή src, στην οποία ορίζουμε το που θα βρίσκονται οι εικόνες με τις οποίες ο αλγόριθμος εκπαιδεύεται και ελέγχει την ακρίβεια στο test set. Στον φάκελο src θα πρέπει να βρίσκονται 3 φάκελοι με ονόματα training, validation, test. Μέσα σε κάθε φάκελο πρέπει να υπάρχουν 2 φάκελοι με ονόματα 0 και 1. Στον φάκελο με όνομα 0 πρέπει να υπάρχουν όλες οι φωτογραφίες για την περίπτωση μη ύπαρξης όγκου, ενώ στον φάκελο 1 πρέπει να υπάρχουν οι εικόνες για την περίπτωση ύπαρξης όγκου. Επίσης για τα μοντέλα που χρησιμοποιήσαν ως αρχικά βάρη, τα βάρη που βρίσκονται σε κάποιο άλλο μοντέλο, θα πρέπει να έχει τοποθετηθεί το αντίστοιχο αρχείο στον ίδιο φάκελο.

5.3 Τροποποίηση Πρώτη, μεταβολή μόνο στο πλήρως συνδεδεμένο επίπεδο (VGG16#1)

Η πρώτη εκδοχή που δοκιμάστηκε είχε τα ίδια ακριβώς βάρη με το αρχικό δίκτυο που έχει εκπαιδευτεί στο σύνολο δεδομένων ImageNet. Στη συγκεκριμένη περίπτωση δεν μεταβάλαμε καθόλου ούτε τη δομή του δικτύου αλλά ούτε και τα βάρη του, για όλα τα επίπεδα του συνελκτικού μέρους του δικτύου. Το μόνο που εκπαιδεύσαμε είναι τα τελευταία επίπεδα που αποτελούν το πλήρως συνδεδεμένο κομμάτι του δικτύου το οποίο και διαχωρίζει τις κλάσεις μιας και πλέον δεν έχουμε 1000 αλλά 2 κλάσεις, και τα βάρη για να τις ξεχωρίσουμε είναι οπωσδήποτε διαφορετικά. Συνεπώς αρχικοποιήθηκαν τα 2 πλήρως συνδεδεμένα επίπεδα με τυχαίες τιμές από κανονική κατανομή με $\mu=0$ και $\sigma=0.01$.



Εικόνα 5-5. Απεικόνιση δομής VGG16#1

Επίσης επειδή το αρχικό δίκτυο είχε ως σκοπό τον διαχωρισμό 1000 κλάσεων χρειαζόταν μεγαλύτερο πλήθος νευρώνων και συνεπώς χαρακτηριστικών, για να διακρίνει τις κατηγορίες. Στην περίπτωση που μελετήθηκε στην παρούσα εργασία είχαμε μόνο 2 κλάσεις και μετά από δοκιμές στο test set, αποφασίσαμε τα 2 πλήρως συνδεδεμένα επίπεδα να αποτελούνται από 300 νευρώνες αντί για 4096 που είχε το αρχικό δίκτυο, το

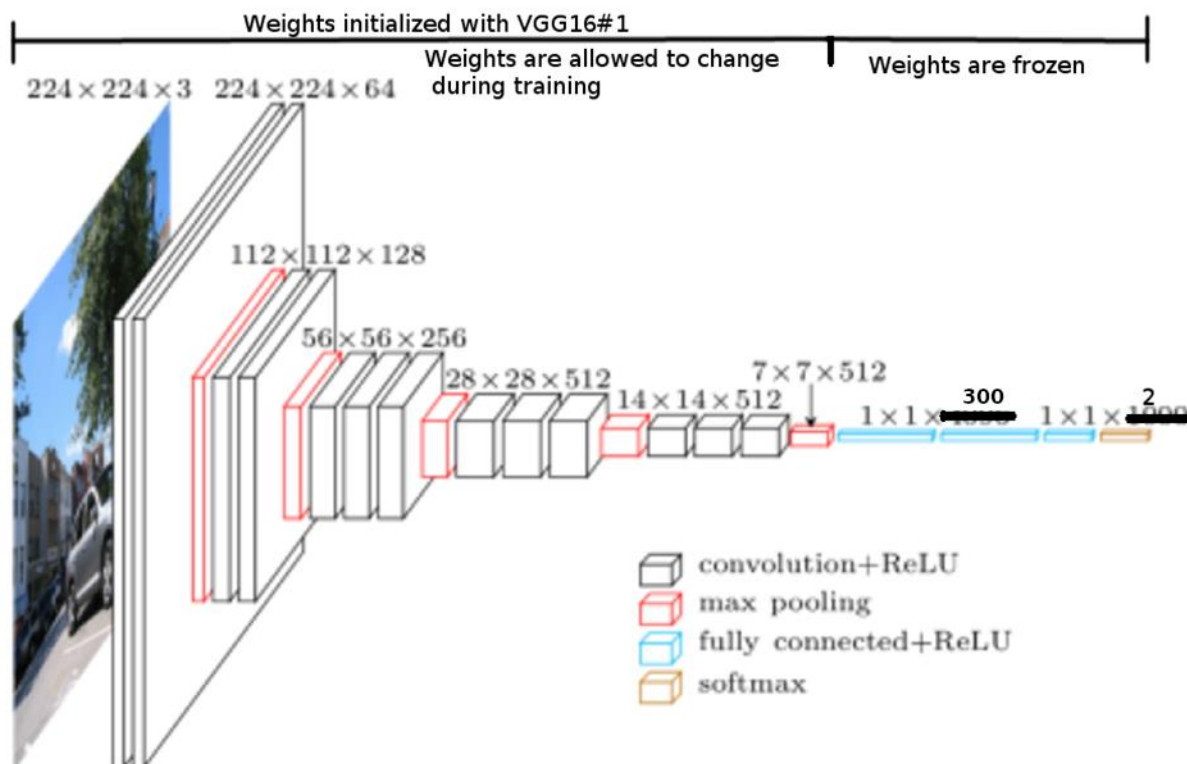
οποίο μείωσε αρκετά τον χρόνο εκπαίδευσης και κατά 96% το πλήθος των συνάψεων μεταξύ των 2 επιπέδων και το πλήθος των παραμέτρων αυτού του τμήματος του δικτύου. Στη συνέχεια ένα πλήρες επίπεδο, το οποίο συνδέεται με μια softmax συνάρτηση για να αντιστοιχίζουμε τα αποτελέσματα σε πιθανότητες. Τέλος επειδή δεν έχουμε καμία κατηγορία ίδια από τις υπάρχουσες του δικτύου είναι υποχρεωτικό να εκπαιδεύσουμε τα πλήρως συνδεδεμένα επίπεδα του δικτύου.

Ο λόγος που επιλέξαμε να ελέγξουμε αυτή την υλοποίηση, ήταν ότι ουσιαστικά αποτελεί ένα baseline για τους αλγόριθμους που δοκιμάσαμε και χρησιμοποιούν μεταφορά γνώσης (transfer learning) [46]. Αυτό γιατί στα νευρωνικά δίκτυα είναι αρκετά απλή η εφαρμογή του transfer learning στην συγκεκριμένη περίπτωση, μιας και ουσιαστικά συνίσταται στη μεταφορά των βαρών ενός δικτύου ως αρχικών βαρών στο δίκτυο που χρησιμοποιείται, με τον περιορισμό βέβαια ότι τα 2 δίκτυα έχουν ίδια αρχιτεκτονική, όπως συμβαίνει στην δική μας περίπτωση.

5.4 Τροποποίηση Δεύτερη, εκπαίδευση μόνο στο συνελκτικό επίπεδο, αρχικοποίηση βαρών από VGG16#1 (VGG16#2)

Η δεύτερη εκδοχή που δοκιμάσαμε είχε τα ίδια ακριβώς βάρη με το δίκτυο VGG16#1 σε όλη την έκταση του δικτύου. Στη συγκεκριμένη τροποποίηση δεν μεταβάλαμε τη δομή του δικτύου καθόλου, το μόνο που επιτρέψαμε να μεταβληθεί είναι τα βάρη του. Το μέρος του δικτύου που επιτρέψαμε να εκπαιδευτεί ήταν όλο το συνελκτικό μέρος του δικτύου, με τα

βάρη στα πλήρως συνδεδεμένα επίπεδα να παραμένουν αμετάβλητα από το VGG16#1.



Εικόνα 5-6. Απεικόνιση δομής VGG#2

Κατά κάποιον τρόπο μπορούμε να το δούμε σαν τη συνέχεια του VGG#1, με την εκπαίδευση του δικτύου να εστιάζεται πλέον στο συνελκτικό κομμάτι και μόνο. Το παρόν ήταν αδύνατο να εκτελεστεί αν προηγουμένως δεν είχε ολοκληρωθεί το VGG#1 μιας και δεν είχαμε στη διάθεση μας αξιόπιστα-εκπαιδευμένα βάρη για το πλήρως συνδεδεμένο μέρος του δικτύου. Συγκρινόμενο με το VGG#1 μπορούμε να δούμε κατά πόσο τα χαρακτηριστικά του VGG16 δικτύου στην αρχική τους μορφή, σε σχέση με αυτά που επιτρέπουμε να μεταβληθούν και να προσαρμοστούν για το συγκεκριμένο πρόβλημα μας βελτιώνουν τη διακριτική ικανότητα κατηγοριοποίησης του δικτύου. Να υπενθυμιστεί σε αυτό το σημείο ότι στην παρούσα τροποποίηση αρχικοποιούμε τις τιμές του δικτύου με τα «χαρακτηριστικά» του VGG#1 και συνεπώς μεταφέρουμε γνώση από το VGG#1 αλλά και από το αρχικό δίκτυο μιας και τα βάρη στο συνελκτικό επίπεδο του VGG#1 παρέμειναν αμετάβλητα. Ουσιαστικά πρόκειται για το πρώτο δίκτυο που σε όλο το βάθος του δικτύου έχει εκπαιδευτεί με τα παραδείγματα του training set.

5.5 Τροποποίηση Τρίτη, διαγραφή τελευταίου συνελκτικού επιπέδου (VGG16#3)

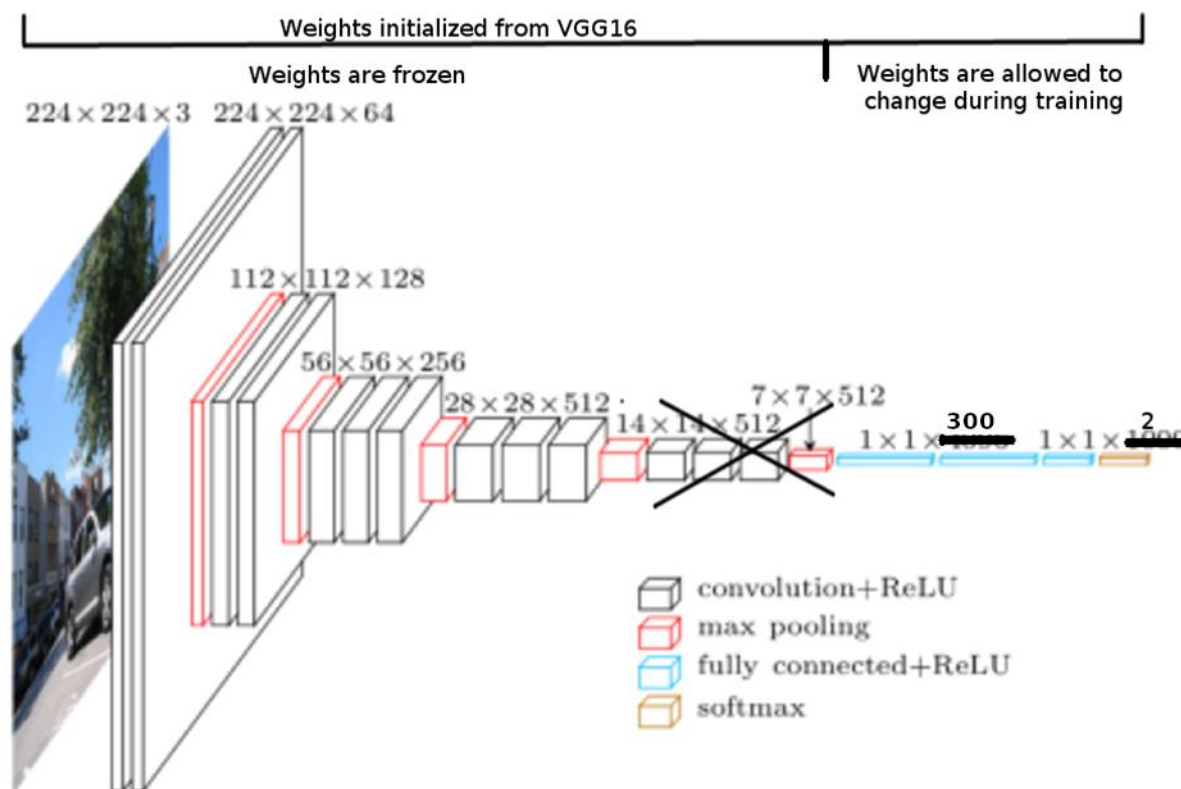
Σκοπός του αρχικού δικτύου ήταν η ταξινόμηση 1000 κατηγοριών. Συνεπώς το πλήθος και η διαφορετικότητα των χαρακτηριστικών ήταν πλέον κρίσιμης σημασίας. Γνωρίζουμε ότι στα πρώτα συνελκτικά επίπεδα τα χαρακτηριστικά τα οποία υπολογίζονται είναι πολύ πιο «βασικά» και απλά στα χαρακτηριστικά που υπολογίζονται σε σχέση με αυτά που υπολογίζονται σε μεταγενέστερα επίπεδα, όπως μπορούμε να δούμε και στην κλασική πλέον στη βιβλιογραφία εικόνα που ακολουθεί [29].



Εικόνα 5-7. Τα σχήματα που γίνονται αντιληπτά από τους νευρώνες, όσο αυξάνουμε το βάθος του δικτύου γίνονται πιο πολύπλοκα

Θέλαμε να δούμε εάν το 5^ο συνελκτικό επίπεδο του VGG16 δικτύου το οποίο αφήνει αμετάβλητο το βάθος του δικτύου είναι «απαραίτητο» για το διαχωρισμό 2 κλάσεων αντί 1000, και εάν είναι κατά πόσο. Δηλαδή θέλουμε να ελέγξουμε κατά πόσο μια μείωση του βάθους και ιδιαίτερα στα τελευταία επίπεδα έχει αρνητική επίδραση στην απόδοση του δικτύου, σε περίπτωση που η «χωρητικότητα» για το πλήθος των κλάσεων δεν είναι ο πρωτεύων παράγοντας.

Γνωρίζουμε ότι το δίκτυο VGG16 έχει εκπαιδευτεί σε άλλα δεδομένα, ενδεχομένως τα πρώτα επίπεδα του να μην χρειάζονται εκπαίδευση και ίσως η εκπαίδευση που θα γίνει να έχει αρνητικά αποτελέσματα στο επιθυμητό μας σκοπό (της αύξησης των μετρικών F1 και BAC). Οποτε στο δίκτυο που προτείνουμε σε αυτή την τροποποίηση «παγώσαμε» το δίκτυο για τα πρώτα επίπεδα, και το αφήσαμε να εκπαιδεύσει μόνο στα τελευταία επίπεδα που είναι πλήρως συνδεδεμένα.



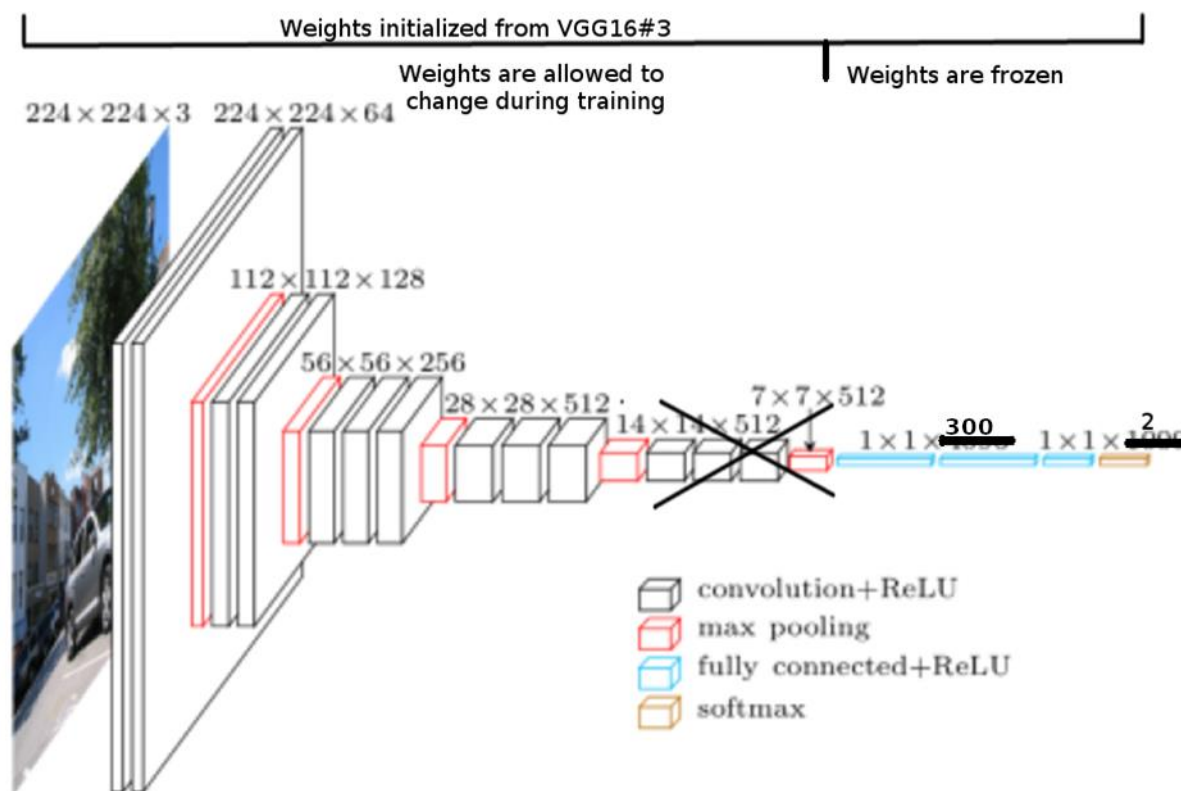
Εικόνα 5-8. Απεικόνιση δομής VGG#3

Σε αυτή την περίπτωση το δίκτυο αρχικοποιήθηκε με τα βάρη του αρχικού VGG16 δικτύου για τα πρώτα 4 επίπεδα, με το πλήρως συνδεδεμένο επίπεδο να αρχικοποιείται με τυχαίες τιμές από κανονική κατανομή με $\mu=0$ και $\sigma=0.01$. Στη συνέχεια επιτράπηκε η εκπαίδευση όλων των επιπέδων του δικτύου. Η μόνη αλλαγή ήταν ότι το τελευταίο επίπεδο (4^ο πλέον μιας και τα 3 συνελκτικά φίλτρα του 5^{ου} επιπέδου έχουν αφαιρεθεί), συνδέεται με το πλήρως συνδεδεμένο μέρος του δικτύου. Συνεπώς επειδή τα χαρακτηριστικά που έρχονται στο πλήρες συνδεδεμένο δίκτυο έχουν μεταβληθεί (με την αφαίρεση ενός επιπέδου και 3 συνελκτικών φίλτρων) δεν είναι λογικό να διακρατήσουμε τα βάρη. Για αυτό το λόγο αρχικοποιήθηκε με τα βάρη του ImageNet μόνο στα συνελκτικά επίπεδα.

Κύριος στόχος αυτής της τροποποίησης, ήταν να ελέγξουμε κατά πόσο το μεγάλο βάθος του δικτύου μπορεί να μειωθεί χωρίς αντίκτυπο στα αποτελέσματα, επειδή στο πρόβλημα που προσπαθούμε να λύσουμε στην παρούσα εργασία έχει 2 κατηγορίες αντί για 1000 που είχε το πρόβλημα στο οποίο εκπαιδεύτηκε το δίκτυο.

5.6 Τροποποίηση Τέταρτη, διαγραφή τελευταίου συνελκτικού επιπέδου, εκπαίδευση συνελκτικών επιπέδων (VGG16#4)

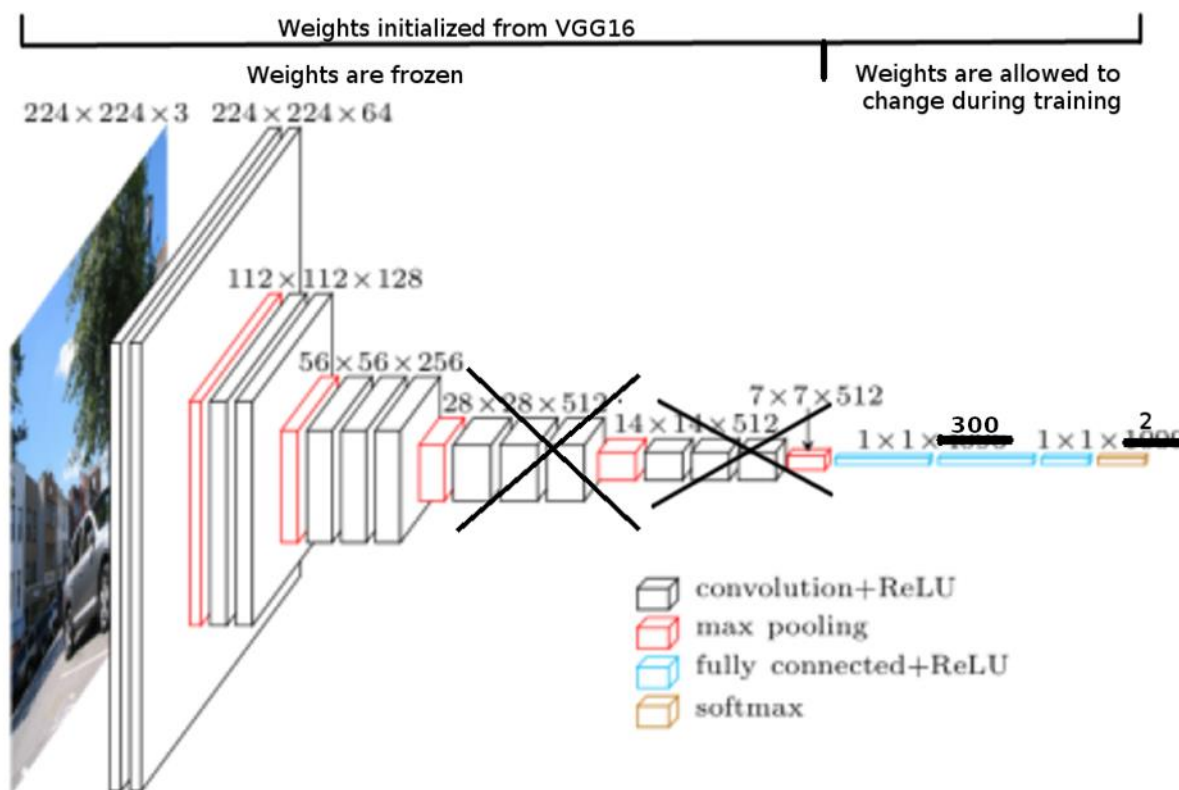
Στην παρούσα τροποποίηση θέλαμε να εξετάσουμε την επίδοση του δικτύου VGG#3, αφού όμως του επιτρέψουμε να εκπαιδευτεί και στα κομμάτι του δικτύου που παράγει τα «χαρακτηριστικά» του, δηλαδή και στα συνελκτικά επίπεδα. Να υπενθυμίσουμε ότι έχει 4 συνελκτικά επίπεδα αντί για 5 που είχε το αρχικό VGG16. Επίσης κατά την εκπαίδευση του VGG#3 εκπαιδεύτηκε μόνο το πλήρως συνδεδεμένο μέρος, με το συνελκτικό μέρος να μένει ανέπαφο από το αρχικό VGG16. Αυτό που εκτελέσαμε στην παρούσα τροποποίηση είναι η εκπαίδευση μόνο του συνελκτικού μέρους, με το πλήρως συνδεδεμένο που έχει εκπαιδευτεί ανέπαφο.



Εικόνα 5-9. Απεικόνιση δομής VGG#4

Συνεπώς περιμένουμε πλέον που έχουν προσαρμοστεί και τα χαρακτηριστικά για το συγκεκριμένο πρόβλημα να έχουμε καλύτερη επίδοση από την επίδοση του δικτύου με το οποίο αρχικοποιήθηκε (VGG16#3). Μία άλλη οπτική γωνία για το λόγο που επιλέξαμε αυτή την τροποποίηση είναι ότι μελετούμε ότι εξέταζε το VGG16#2, αλλά με αρχικό επίπεδο το VGG14#3.

5.7 Τροποποίηση Πέμπτη, διαγραφή δύο τελευταίων συνελκτικών επιπέδων, εκπαίδευση πλήρως συνδεδεμένου δικτύου (VGG16#5)



Εικόνα 5-10. Απεικόνιση δομής VGG#5

Σκοπός του αρχικού δικτύου ήταν η ταξινόμηση 1000 κατηγοριών. Συνεπώς το πλήθος και η διαφορετικότητα των χαρακτηριστικών ήταν πλέον κρίσιμης σημασίας. Γνωρίζουμε ότι στα πρώτα συνελκτικά επίπεδα τα χαρακτηριστικά τα οποία υπολογίζονται είναι πολύ πιο «βασικά» και απλά σε σχέση με αυτά που υπολογίζονται σε μεταγενέστερα επίπεδα [29]. Θέλαμε να εξετάσουμε εάν αφαιρούνταν τα 2 τελευταία συνελκτικά επίπεδα του VGG16 δικτύου το οποίο πλέον μεταβάλλει σημαντικά το βάθος του δικτύου, πόσο μεγάλη θα ήταν η επίδραση στο αποτέλεσμα που θα λάβουμε. Με αυτό τον τρόπο θα μπορούμε να κρίνουμε αν το βάθος είναι απαραίτητο για το διαχωρισμό 2 κλάσεων αντί 1000. Το παραπάνω γίνεται ιδιαίτερα σημαντικό αφού στα τελευταία επίπεδα «που δημιουργούν τα πλέον πολύπλοκα χαρακτηριστικά» θα αφαιρεθούν και θα πρέπει να συμπυκνωθεί η γνώση σε επίπεδα που βρίσκονται πιο κοντά στην είσοδο.

Γνωρίζουμε ότι το δίκτυο VGG16 έχει εκπαιδευτεί σε άλλα δεδομένα, ενδεχομένως τα πρώτα επίπεδα του να μην χρειάζονται εκπαίδευση και ίσως η εκπαίδευση που θα γίνει να έχει αρνητικά αποτελέσματα στο επιθυμητό μας σκοπό (της αύξησης των μετρικών F1

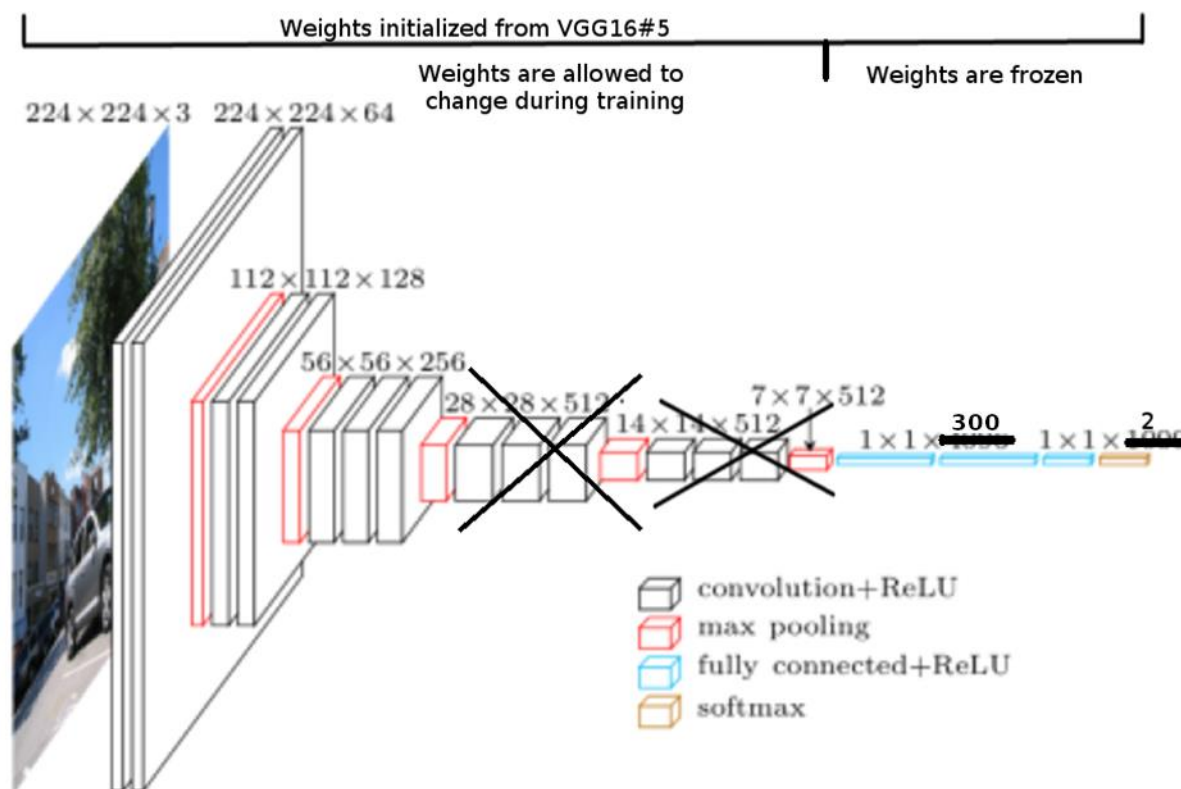
και BAC). Όποτε στο δίκτυο που προτείνουμε σε αυτή την περίπτωση «παγώσαμε» τα βάρη και αφαιρέσαμε τα συνελκτικά επίπεδα 4 και 5. Επίσης στο δίκτυο για τα πρώτα επίπεδα βάλουμε τα αρχικά βάρη και εκπαιδεύτηκε μόνο στα πλήρως συνδεδεμένα επίπεδα που έχουν αρχικοποιηθεί τυχαία.

5.8 Τροποποίηση Έκτη, διαγραφή δύο τελευταίων συνελκτικών επιπέδων, εκπαίδευση συνελκτικών επιπέδων δικτύου (VGG16#6)

Στην παρούσα τροποποίηση θέλαμε να εξετάσουμε την επίδοση του δικτύου VGG#5, αφού όμως του επιτρέψουμε να εκπαιδευτεί και στα «χαρακτηριστικά» του, δηλαδή και στα συνελκτικά επίπεδα. Να υπενθυμίσουμε ότι έχει πλέον μόλις 3 συνελκτικά επίπεδα αντί για 5 που είχε το αρχικό VGG16. Επίσης κατά την εκπαίδευση του VGG#5 εκπαιδεύτηκε μόνο το πλήρως συνδεδεμένο μέρος, με το συνελκτικό μέρος να μένει ανέπαφο, όπως ήταν από το VGG16.

Συνεπώς η τροποποίηση που εκτελέσαμε είναι η εκπαίδευση και του συνελκτικού μέρους, με το πλήρως συνδεδεμένο που έχει εκπαιδευτεί όπως εκπαιδεύτηκε από το VGG#5. Συνεπώς περιμένουμε πλέον που έχουν προσαρμοστεί και τα χαρακτηριστικά για το συγκεκριμένο πρόβλημα να έχουμε καλύτερη επίδοση από το πηγαίο δίκτυο VGG16#5. Κατ' ουσίαν εξετάζουμε ότι εξέταζε το VGG16#2 και το VGG16#4 αλλά με αρχικό επίπεδο το VGG14#5 με μικρότερο βάθος δικτύου.

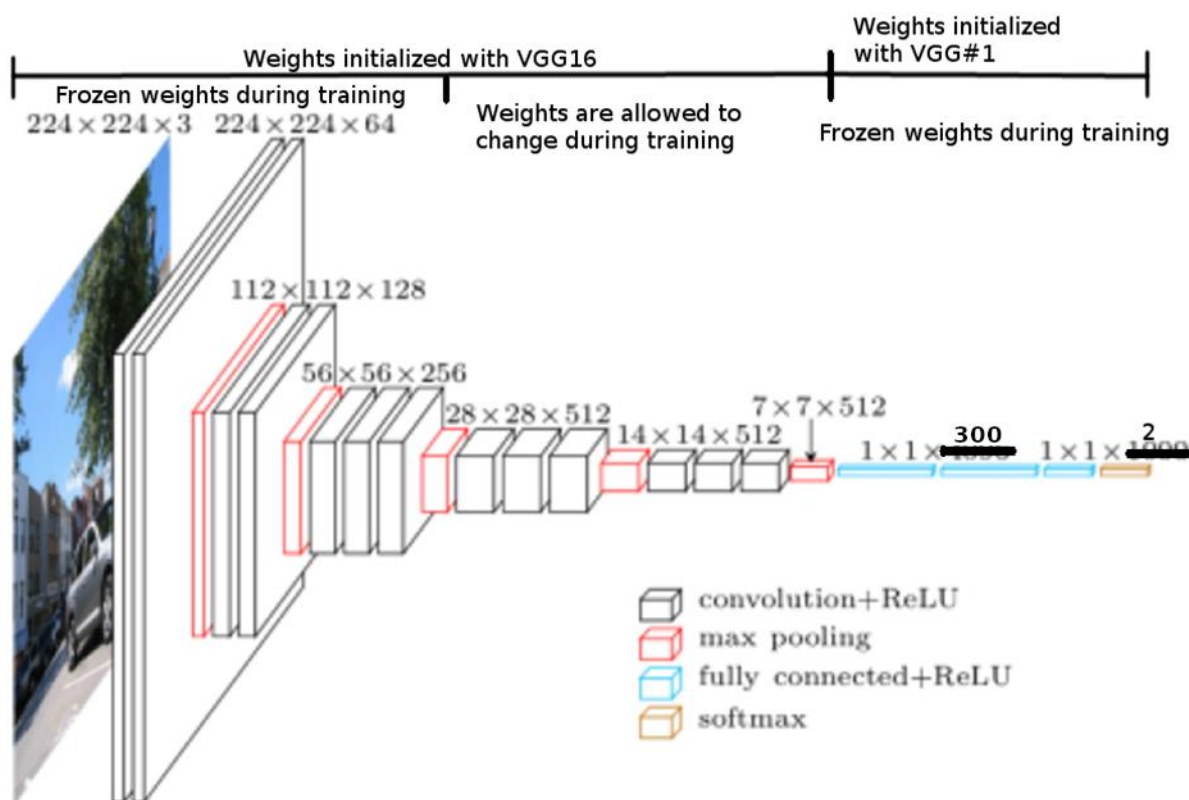
Ως συνέπεια, της μείωσης του πλήθους των παραμέτρων που πρέπει να εκπαιδευτούν, η ταχύτητα εκπαίδευσης είναι αρκετά αυξημένη σε σχέση με όλες τις προηγούμενες αντίστοιχες υλοποιήσεις.



Εικόνα 5-11. Απεικόνιση δομής VGG#6

5.9 Τροποποίηση Έβδομη, εκπαίδευση μόνο 2 τελευταίων συνελκτικών επιπέδων (VGG16#7)

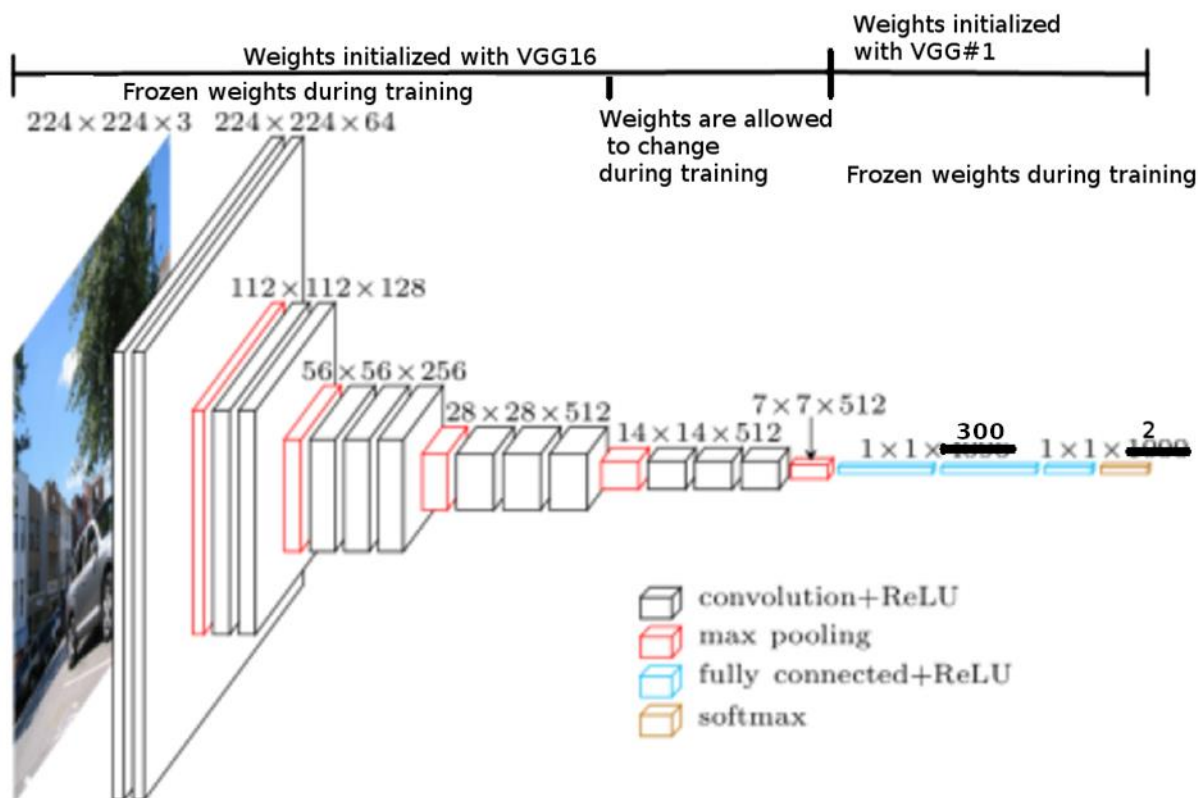
Στη συγκεκριμένη τροποποίηση εξετάζουμε το ρόλο που παίζει η εκπαίδευση συγκεκριμένων επιπέδων του δικτύου χωρίς να αφαιρέσουμε επίπεδα. Για να το πετύχουμε αυτό αρχικοποιήσαμε το δίκτυο με τα βάρη του VGG16#1, δηλαδή για τα επίπεδα που μας δημιουργούν τα «χαρακτηριστικά» κρατήσαμε τις τιμές ίδιες με το αρχικό δίκτυο VGG16 και για το πλήρως συνδεδεμένο επίπεδο πήραμε τα βάρη που αντιστοιχούν στην εκπαίδευση μόνο με το πλήρως συνδεδεμένο μέρος του δικτύου VGG16#1. Στη συνέχεια επιτρέψαμε την εκπαίδευση μόνο των 2 τελευταίων συνελκτικών επιπέδων του δικτύου με τη κυρίαρχη λογική να είναι ότι τα πρώτα επίπεδα εντοπίζουν στοιχειώδη μέρη της εικόνας συνεπώς ενδέχεται να μη χρειάζεται να εκπαιδευτούν. Οπότε για να κερδίσουμε σε χρόνο εκπαίδευσης αλλά και να επιβεβαιωθεί η υπόθεσή μας εκπαιδεύουμε μόνο τα 2 τελευταία συνελκτικά επίπεδα και θα συγκρίνουμε το κατά πόσο αυτού του είδους η εκπαίδευση προτείνεται για ιατρικά προβλήματα.



Εικόνα 5-12. Απεικόνιση δομής VGG#7

5.10 Τροποποίηση Όγδοη, εκπαίδευση μόνο τελευταίου συνελκτικού επιπέδου (VGG16#8)

Στη παρούσα τροποποίηση θέλαμε να εξετάσουμε το ρόλο που παίζει η εκπαίδευση πλέον μόνο ενός επιπέδου του δικτύου χωρίς την αφαίρεση επιπέδων. Για να το πετύχουμε αυτό πράξαμε ακριβώς όπως στο VGG16#7 με τη διαφορά ότι τώρα επιτρέπουμε μόνο σε ένα, στο τελευταίο συνελκτικό επίπεδο να εκπαιδευτεί. Η απόφαση αυτή είναι «ριζοσπαστική» και περιμένουμε ουσιαστικά να δούμε κατά πόσο θα υπάρξει καλύτερευση στα αποτελέσματα, και αν ναι κατά πόσο. Με αυτό τον τρόπο θέλουμε να δούμε τη βελτίωση αν εκπαιδύσουμε μόνο το συνελκτικό επίπεδο που παράγει τα πιο «αφηρημένα» χαρακτηριστικά κατά μόνας.



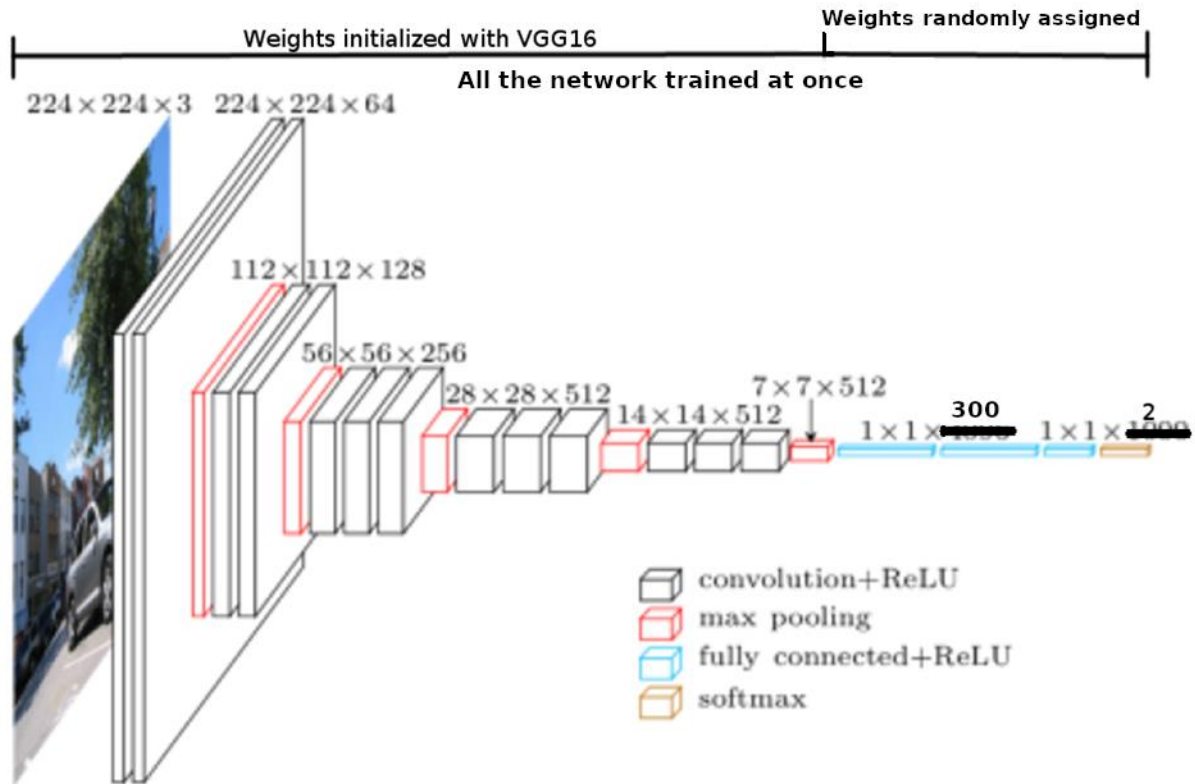
Εικόνα 5-13. Απεικόνιση δομής VGG#8

Σε αυτήν την περίπτωση μελετάμε την επίδραση ενός και μόνο επιπέδου στην αποτελεσματικότητα του δικτύου μας. Επιλέχθηκε το τελευταίο επίπεδο ώστε να μην υπάρχουν άλλα επίπεδα μετά από αυτό που να εξαρτώνται από τα βάρη του μιας και θα απαιτούσαν εκ νέου εκπαίδευση.

5.11 Τροποποίηση Ένατη, αρχικοποίηση με VGG16 και εκπαίδευση σε όλο το δίκτυο (VGG16#9)

Ο σκοπός αυτής της δοκιμής είναι να εξετάσουμε το «όφελος» της μεταφοράς γνώσης, έστω και μόνο στο συνελκτικό μέρος του δικτύου. Κατά κύριο λόγο στην παρούσα τροποποίηση ουσιαστικά εξετάζεται το κατά πόσο έχει νόημα να εκπαιδεύουμε τα δύο μέρη του δικτύου, συνελκτικό και πλήρως συνδεδεμένο, ξεχωριστά. Το γεγονός ότι στις μέχρι τώρα τροποποιήσεις εκπαιδεύτηκε πρώτα το πλήρως συνδεδεμένο μέρος του δικτύου και στη συνέχεια το συνελκτικό μέρος εισάγει μια χρονική πολυπλοκότητα στην εκπαίδευση. Επίσης δεν είναι βέβαιο ότι παράγει καλύτερα αποτελέσματα η μία τεχνική εκπαίδευσης από την άλλη. Συνεπώς, στην παρούσα τροποποίηση αφέθηκε το δίκτυο να

εξετάσει κατά πόσο τα βάρη του συνολικού δικτύου μπορούν να προλάβουν να προσαρμοστούν σε ένα ελάχιστο που να προσφέρει καλύτερα αποτελέσματα σε σχέση με το VGG16#1 και κυρίως το VGG16#2 όπου χρησιμοποιήσαμε την ίδια ακριβώς αρχιτεκτονική με διαφορά μόνο στο μέρος του δικτύου που επιτράπηκε να εκπαιδευτεί.



Εικόνα 5-14 Απεικόνιση δομής VGG#9

6. ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο παρόν κεφάλαιο αναλύουμε τα αποτελέσματα που λάβαμε για τις τροποποιήσεις των συνελκτικών δικτύων όπως αναλύθηκαν στο προηγούμενο κεφάλαιο. Αποτυπώνουμε όλες τις μετρικές και τα διαγράμματα που έχουμε λάβει ως αποτέλεσμα. Στα αποτελέσματα συμπεριλαμβάνονται οι στατιστικές που χρησιμοποιήσαμε για τον υπολογισμό των αποτελεσμάτων σε όλες τις περιπτώσεις που εξετάστηκαν, καθώς και οι γραφικές παραστάσεις που αφορούν σε κάθε μία από τις τροποποιήσεις του CNN που θέλαμε να εξετάσουμε την απόδοσή τους. Πιο συγκεκριμένα, έχουμε αποτυπώσει κατά τη διάρκεια εκπαίδευσης των CNN τη συνάρτηση απώλειας (loss function) που χρησιμοποιήθηκε καθώς και την ακρίβεια (accuracy) για κάθε μια από τις εποχές εκπαίδευσης.

Επίσης για κάθε μια από τις τροποποιήσεις του δικτύου VGG16 που δοκιμάστηκαν παρουσιάζεται ο πίνακας σύγχυσης (confusion matrix) για το τελικό σετ εξέτασης (validation set) στο οποίο και υπολογίζουμε τις μετρικές accuracy (για το validation set), BAC καθώς επίσης και την F1 για να μπορέσουμε να το συγκρίνουμε με μεθόδους που ήδη έχουν εφαρμοστεί στην βιβλιογραφία.

6.1 Επισκόπηση υπαρχόντων αποτελεσμάτων

Όπως αναφέρεται στο [38] χρησιμοποιήθηκαν εκτός από το CNN και άλλες μέθοδοι ταξινόμησης με χρήση κάποιων υπολογισμένων χαρακτηριστικών χωρίς τη χρήση συνελκτικών δικτύων. Μπορούμε να δούμε τα αποτελέσματα που παρήχθησαν από τη χρήση των αντίστοιχων χαρακτηριστικών στον αμέσως επόμενο πίνακα. Παρατηρούμε ότι το CNN που προτείνεται σε σχέση με τα υπόλοιπα χαρακτηριστικά που ελέγχονται στην εργασία παρέχει τα υψηλότερα ποσοστά και στην μετρική BAC αλλά και την μετρική F1. Αυτός είναι και ο λόγος που στα αποτελέσματα μας στην παρούσα εργασία δεν συγκρινόμαστε με τα υπόλοιπα χαρακτηριστικά παρά μόνο με τα αποτελέσματα που προέρχονται από το βασικό δίκτυο CNN όπως υλοποιήθηκε στην εργασία τους. Για λόγους πληρότητας της εργασίας, εξηγείται παρακάτω η αρχιτεκτονική της δομής CNN που χρησιμοποιήσαν οι συγγραφείς της εργασίας αναφοράς. Αρχικά στην εικόνα εφαρμόζουν 2 επίπεδα συνελίξεων με επίπεδο pooling και τελικά ένα πλήρως συνδεδεμένο επίπεδο στο οποίο εφαρμόζουν νευρώνες που χρησιμοποιούν τη λογιστική συνάρτηση όπως έχει ήδη εξηγηθεί στο κεφάλαιο 3.

Πίνακας 1. Περιγραφή των χαρακτηριστικών που χρησιμοποιήθηκαν ώστε να μπορεί να συγκριθεί το συνελκτικό νευρωνικό δίκτυο αναφοράς

Handcrafted feature	Visual property captured
Gray Histogram (GH) ³⁴	Luminance
Fuzzy Color Histogram (FCH) ³⁵	Color
HSV Color Histogram (HSVCH) ³⁶	Color
RGB Histogram (RGBH) ^{34, 37}	Color
JPEG Coefficient Histogram (JPEGCH) ³⁶	Color and Texture
Local Binary Partition Histogram (LBP) ³⁸	Texture
MPEG7 Edge Histogram (M7Edge) ³⁹	Texture
Haralick features ⁹	Nuclear textural (NT)
Graph-based features ⁹	Nuclear Architectural (NA)

Πίνακας 2. Αποτελέσματα σύγκρισης CNN με άλλες τεχνικές μηχανικής μάθησης

	Pr	Rc/Sen	Spc	F1	BAC
CNN	0.6540	0.7960	0.8886	0.7180	0.8423
FCH	0.7086	0.6450	0.9298	0.6753	0.7874
RGBH	0.7564	0.5956	0.9493	0.6664	0.7724
GH	0.7102	0.5240	0.9434	0.6031	0.7337
JPEGCH	0.7570	0.4646	0.9605	0.5758	0.7126
M7Edge	0.7360	0.4372	0.9585	0.5485	0.6979
NT	0.6246	0.2851	0.9547	0.3915	0.6199
LBP	0.7575	0.2291	0.9806	0.3518	0.6048
NA	0.6184	0.2413	0.9606	0.3472	0.6009
HSVCH	0.7662	0.2223	0.9821	0.3446	0.6022

6.2 Αποτελέσματα για VGG16#1

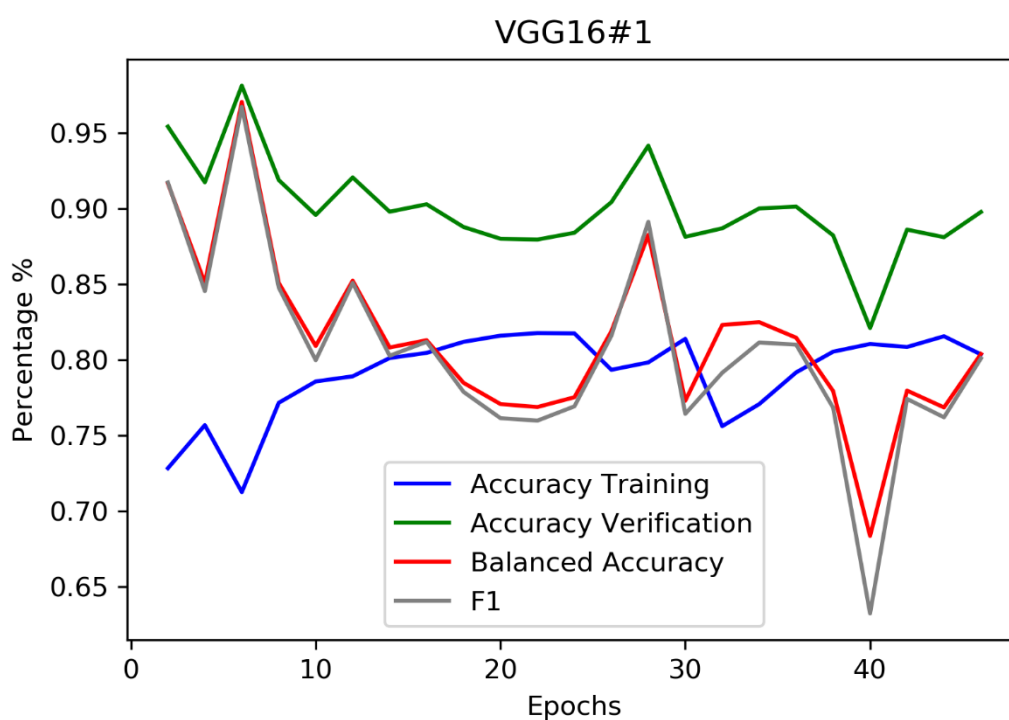
Στη συγκεκριμένη περίπτωση χρησιμοποιήσαμε τα βάρη του δικτύου VGG16 εκτός από το πλήρως συνδεδεμένο επίπεδο το οποίο αρχικοποιήθηκε με τυχαία βάρη και κατώφλια. Οι παράμετροι που χρησιμοποιήθηκαν για τον αλγόριθμο ελαχιστοποίησης είναι οι παρακάτω:

Πίνακας 3. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#1

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	48

Ταχύτητα εκμάθησης (learning rate)	0.00002
Χρήση ορμής εκμάθησης (momentum)	0.1
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Τα διαγράμματα που λάβαμε για την ακρίβεια κατά την εκπαίδευση του δικτύου, ελεγμένο και στο training set αλλά και στο testing set παρατίθεται παρακάτω.



Εικόνα 6-1. Accuracy και άλλες μετρικές για το δίκτυο VGG16#1

Παρατηρούμε ότι κατά τη διάρκεια της εκπαίδευσης αρχίζουμε από αρκετά χαμηλές τιμές training accuracy της τάξης του 75% που είναι αναμενόμενο μιας και η αναλογία των κλάσεων στο training set είναι σχεδόν 1 προς 4. Παρατηρούμε όμως προς μεγάλη μας έκπληξη ότι στα πρώτα δείγματα το accuracy validation είναι σε πολύ υψηλά επίπεδα και βαίνει μειούμενο μέχρι και την εποχή 40. Αυτή την εικόνα θα την περιμέναμε να τη δούμε μετά από overfitting το οποίο είναι αδύνατο στην περίπτωση μας μιας και συνέβη από

την αρχή της εκπαίδευσης μας. Για να διερευνήσουμε λίγο παραπάνω την «ιδιαιτερη» συμπεριφορά σε αυτή την περίπτωση δοκιμάσαμε άλλες 3 αρχικοποιήσεις για 3 εποχές και πήραμε τα ακόλουθα αποτελέσματα.

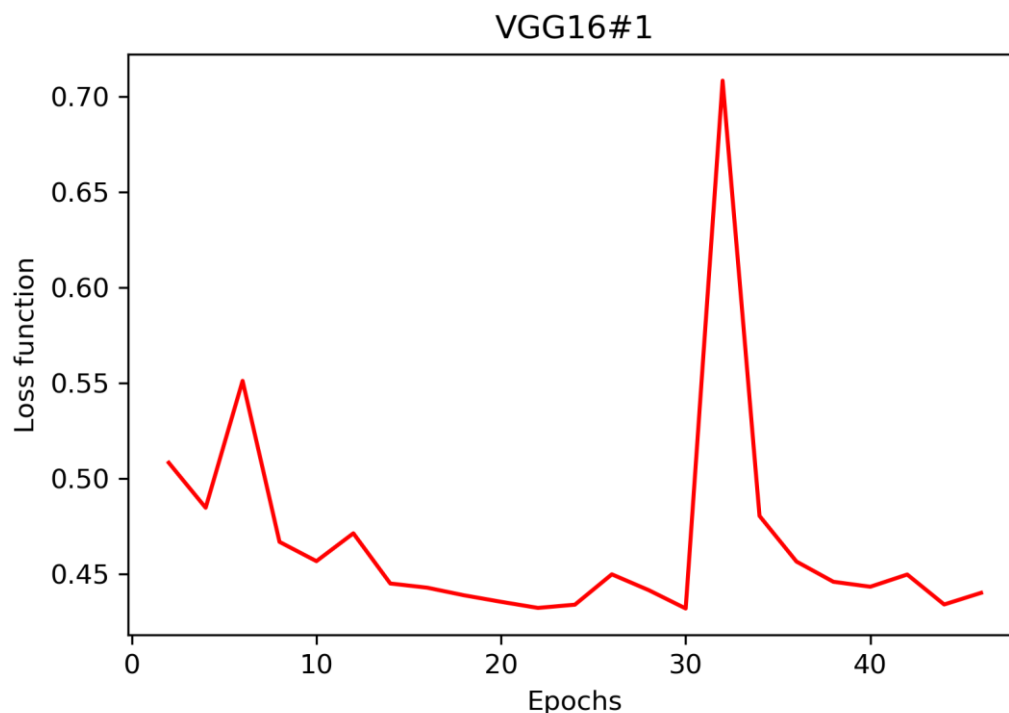
Πίνακας 4. Αποτελέσματα accuracy για 3 διαφορετικές αρχικοποιήσεις μετά από μόλις 1 εποχή εκπαίδευσης.

Training Accuracy	Validation Accuracy
0.7170	0.7364
0.7432	0.7580
0.7332	0.8291

Διακρίνεται αρκετά μεγάλη διαφορά στο validation set, ανάλογα με την αρχικοποίηση που γίνεται στο πλήρες συνδεδεμένο δίκτυο. Σε πλήρη αντίθεση το training accuracy που έχει πολύ μικρότερη μεταβλητότητα.

Επίσης διαπιστώνουμε ότι η αύξηση στο training set είναι σταθερή εκτός από το διάστημα εποχών από 31-33 όπου παρατηρούμε μια έντονη πτώση, γεγονός που αποδεικνύει το ασταθές και απότομο χαρακτήρα που επιδεικνύει το πεδίο εκπαίδευσης. Μία επιπλέον παρατήρηση είναι ότι οι μετρικές BAC και F1 έχουν πολύ μικρή διαφορά μεταξύ τους σε όλες τις εποχές. Αυτό είναι αναμενόμενο μιας και όσο περισσότερα δείγματα ταξινομούνται σωστά αυξάνουν και τις 2 στατιστικές. Επίσης το accuracy validation έχει ίδια μορφή αλλά παρουσιάζει μεγάλες διαφορές από τις καμπύλες των F1 και BAC και βρίσκεται καθ' όλη τη διάρκεια της εκπαίδευσης πάνω από και τις 2 μετρικές λόγω του ότι επηρεάζεται πολύ από τη κλάση που είναι περισσότερο αντιπροσωπευμένη στο training set.

Η εκπαίδευση μπορεί να φαίνεται ότι σταμάτησε σχετικά πρόωρα, όμως είχε τεθεί όριο 20 εποχών patience, όπου αν δεν παρατηρούνταν αύξηση στο παρατηρούμενο accuracy να σταματούσε η εκπαίδευση.



Εικόνα 6-2. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#1

Στη συνάρτηση απώλειας παρατηρούμε ότι στην αρχή η τιμή της μειώνεται με μικρή διακύμανση, όμως στην εποχή 30 παρατηρούμε ένα μια ραγδαία επιδείνωση η οποία όμως άμεσα επανέρχεται στα προηγούμενα επίπεδα εκπαίδευσης.

Σε αυτό το μοντέλο θα παρουσιάσουμε 2 από τα μοντέλα εκπαίδευσης καθώς και τις στατιστικές που λάβαμε λόγο του ιδιαίτερου σχήματος της συνάρτησης εκπαίδευσης. Το πρώτο μοντέλο που παρουσιάζουμε είναι αυτό που είχε την καλύτερη επίδοση και εντοπίζεται μόλις στην 6^η εποχή εκπαίδευσης, στην οποία και παρατηρούμε αύξηση της συνάρτησης απώλειας! Το γεγονός αυτό κατά κάποιον τρόπο αναδεικνύει το γεγονός ότι φτάσαμε σε αυτό το μοντέλο απρόσμενα και σίγουρα όχι με την εφαρμογή της λογικής μείωσης της συνάρτησης απώλειας.

Ακολουθεί το confusion matrix για το validation set με τη χρήση του τελικού μοντέλου. Να υπενθυμίσουμε σε αυτό το σημείο ότι το validation set περιέχει 11.561 εικόνες, εκ των οποίων 2944 θετικές και 8617 αρνητικές.

Πίνακας 5. Αποτελέσματα της τροποποίησης VGG16#1 σε μορφή πίνακα σύγκρισης, 6^η Εποχή Εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	3188	139
True Negative	75	8159

Οι στατιστικές που λάβαμε για το παραπάνω μοντέλο είναι οι ακόλουθες και είναι εντυπωσιακές, για ένα τόσο λίγο εκπαιδευμένο δίκτυο στο πλήρως συνδεδεμένο επίπεδο.

Πίνακας 6. Μετρικές αποτελεσμάτων για την περίπτωση VGG#1, 6^η Εποχή Εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0.9770	0.9582	0.9675	0.9707	0.9815

Το παραπάνω αποτέλεσμα είναι καλύτερο από όλα όσα έχουν παρουσιαστεί για το παραπάνω dataset στη βιβλιογραφία. Πάντως είναι τελείως απρόσμενο να επιτευχθεί ένα τέτοιο αποτέλεσμα με τόσο λίγη εκπαίδευση για αυτό το λόγο το αποδίδουμε κυρίως στην αρχικοποίηση του δικτύου. Είναι βέβαιο ότι το γεγονός πως το συνελκτικό μέρος του δικτύου είναι ήδη εκπαιδευμένο είναι προϋπόθεση για ένα τέτοιο αποτέλεσμα.

Πίνακας 7. Αποτελέσματα της τροποποίησης VGG16#1 σε μορφή πίνακα σύγκρισης, μετά την 44^η εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2076	758
True Negative	868	7859

Παρατηρούμε ότι το νευρωνικό δίκτυο έχει «αφιερώσει» πολύ από την διαχωριστική του ικανότητα στο να μην κάνει λάθη από την αρνητική κατηγορία στη θετική (type II error) λόγω της ανισότητας στο πλήθος των εικόνων των δύο κλάσεων. Αυτό φαίνεται από το γεγονός ότι έχουμε περίπου τετραπλάσιες εικόνες αρνητικές, όμως το πλήθος των εικόνων που έχουν κατηγοριοποιηθεί λάθος είναι αριθμητικά κοντά. Αυτή η «προτίμηση» του δικτύου έχει ιδιαίτερο αντίκτυπο στην μετρική BAC όπου οι κατηγορίες θεωρούνται

ισότιμες και τα λάθη από τη ολιγοπληθέστερη σε αντιπροσώπευση κατηγορία έχουν βαρύτερο αντίκτυπο.

Στη συνέχεια παρουσιάζουμε τα χαρακτηριστικά του μοντέλου που προέκυψε μετά από 44 εποχές, το τελευταίο μοντέλο πριν αρχίσει η συνάρτηση απώλειας να αυξάνεται. Το παραπάνω μοντέλο παρόλο που τα στατιστικά που επιδεικνύει είναι σαφώς υποδεέστερα του μοντέλου μετά από 6 εποχές εκπαίδευσης, είναι πιο αντιπροσωπευτικό αυτού που αναμενόταν με τη χρήση της παρούσας μεθόδου. Επίσης παρατηρούμε ότι η συνάρτηση απώλειας μέχρι και αυτό το μοντέλο έχει την τάση να μειώνεται που αποτελεί και το κριτήριο επιλογής αυτού του μοντέλου.

Το μοντέλο αυτό χρησιμοποιήθηκε για την αρχικοποίηση σε όσα από τα υπόλοιπα δίκτυα χρειαζόμασταν ένα πλήρες συνελικτικό δίκτυο εκπαιδευμένο αρχίζοντας από τα βάρη του VGG16. Ένας επιπλέον λόγος για το ότι επιλέξαμε αυτό το δίκτυο για να αρχικοποιήσουμε τα υπόλοιπα μοντέλα μας είναι ότι τα αποτελέσματα για το δίκτυο μετά την 6^η εποχή δεν επιδέχονται μεγάλης βελτίωσης, συνεπώς δεν θα μπορούσαμε ρεαλιστικά να κρίνουμε το εάν το μοντέλο βελτιώθηκε λόγω αρχικοποίησης ή λόγω του τρόπου εκπαίδευσης και της δομής του δικτύου. Επίσης ένας σημαντικότερος λόγος είναι ότι το μοντέλο στην εργασία καλό είναι να κρίνεται a priori των αποτελεσμάτων στο validation set μιας και εάν αποφασίζουμε με τα αποτελέσματα αυτά, κατ' ουσίαν το μετατρέπουμε σε test set.

Ακολουθούν τα αποτελέσματα για τις μετρικές που μας ενδιαφέρουν F1, BAC και accuracy καθώς επίσης και τα αποτελέσματα για τις μετρικές sensitivity, specificity και precision που απαιτούνται για τον υπολογισμό τους.

Πίνακας 8. Μετρικές αποτελεσμάτων για την περίπτωση VGG#1 , μετά την 44ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0,8998	0,6610	0,7620	0,7686	0,8812

Στα στατιστικά αυτού του μοντέλου παρατηρούμε παρατηρούμε δύο πολύ ενδιαφέροντα στοιχεία. Πρώτον, τη μεγάλη απόκλιση μεταξύ precision και recall εις βάρος του recall. Αυτό αναδεικνύει ότι, το γεγονός ότι αντιπροσωπεύεται η αρνητική κλάση με τετραπλάσια δείγματα, επηρεάζει το δίκτυο να μεροληπτεί υπέρ αυτής της κλάσης. Επίσης, παρατηρούμε είναι ότι η απόκλιση του BAC από το accuracy. Επειδή το BAC δίνει όμοια βαρύτητα και στις 2 κλάσεις και το δίκτυο «προσαρμόστηκε» λόγω loss function στην

κλάση με τα περισσότερα δείγματα, «τιμωρείται» από το γεγονός ότι οι κλάσεις θεωρούνται ισάξιες στο BAC.

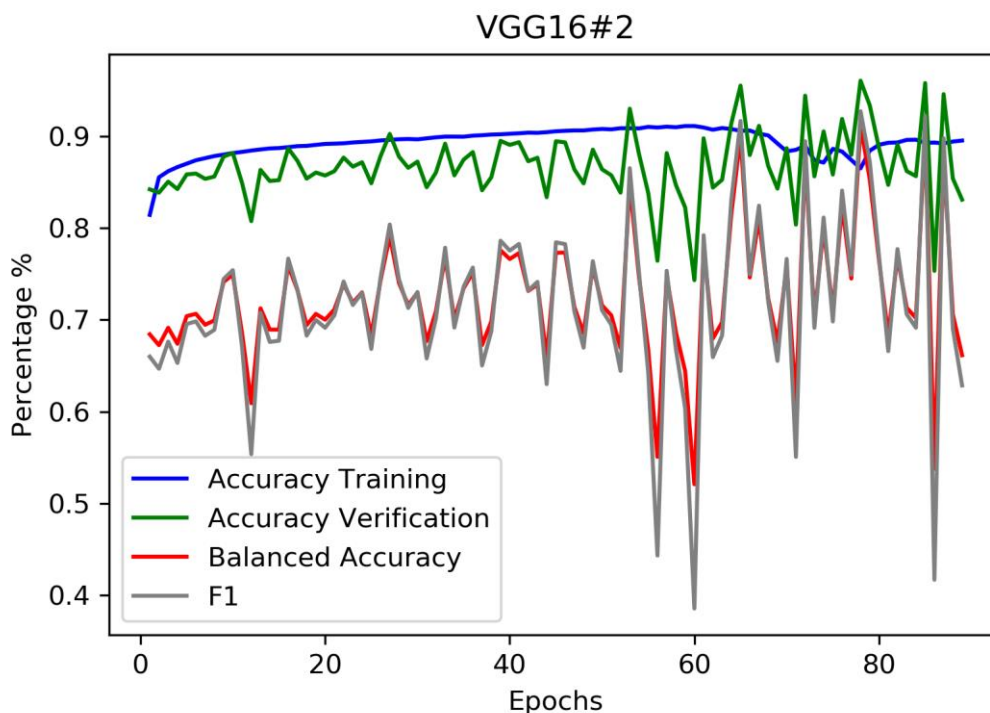
6.3 Αποτελέσματα για VGG16#2

Αρχικά να υπενθυμίσουμε τη λειτουργία που θέλουμε να εξετάσουμε με το δίκτυο VGG16#2. Ουσιαστικά κρατήθηκαν τα βάρη σε όλο το δίκτυο από το VGG16#1, όμως επειδή στο VGG16#1 τα βάρη του συνελκτικού μέρους δεν επηρεάστηκαν ουσιαστικά είναι σαν να μένουν αμετάβλητα από το αρχικό δίκτυο VGG16. Εν συνεχεία, αρχικοποιήθηκε το πλήρως συνδεδεμένο μέρος με τα βάρη του VGG16#1.

Πίνακας 9. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#2

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	89
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Τα διαγράμματα που λάβαμε για την ακρίβεια και για τις υπόλοιπες μετρικές που μας ενδιαφέρουν κατά την εκπαίδευση του δικτύου VGG16#2, ελεγμένο και στο training set αλλά και στο testing set παρατίθεται παρακάτω.

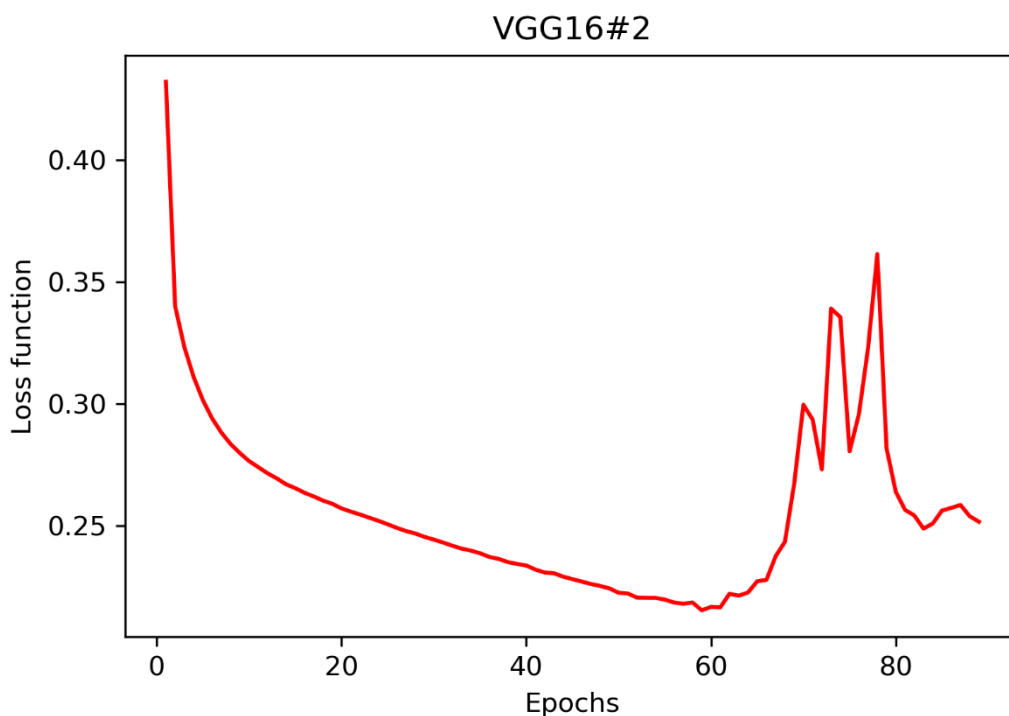


Εικόνα 6-3. Accuracy και άλλες μετρικές για το δίκτυο VGG16#1

Παρατηρούμε ότι κατά τη διάρκεια της εκπαίδευσης αρχίζουμε από υψηλότερες τιμές training accuracy σε σχέση με το VGG16#1, μιας και αρχίζουμε χρησιμοποιώντας την εκπαίδευση του και όχι με αρχικοποίηση τυχαίων τιμών.

Εξετάζοντας το διάγραμμα, είναι εμφανές ότι το διάγραμμα accuracy στο training set είναι εντυπωσιακά ομαλό και διαρκώς αυξανόμενο με μειούμενο ρυθμό, μέχρι και την εποχή 61. Αυτό καταδεικνύει ότι οι παράμετροι εκπαίδευσης που χρησιμοποιήθηκαν «ταιριάζουν» με το συγκεκριμένο πρόβλημα. Κατά το ίδιο διάστημα η accuracy στο validation set, φαίνεται να βελτιώνεται αλλά με κάποια διακύμανση. Μετά την εποχή 61 παρατηρούμε πτώση στο accuracy του training set, συνοδευόμενο με ραγδαίες μεταβολές στο validation accuracy και ακόμα πιο ραγδαίες αλλαγές στις μετρικές BAC και F1, το οποίο εξηγείτε με μεταβολή αρκετών στοιχείων της θετικής κλάσης στην αρνητική. Μας καταδεικνύει το παραπάνω ότι στο χώρο παραμέτρων έχει βρεθεί σε σημείο όπου συμβαίνουν μεγάλες αλλαγές στο αποτέλεσμα με μικρή αλλαγή στην είσοδο, οι οποίες όμως δεν είναι απαραίτητα θετικές. Αυτός ήταν και ο λόγος όπου θεωρήσαμε ως τέλος της εκπαίδευσης την εποχή 61 όπου και παρατηρούμε την υψηλότερη τιμή στην

εκπαίδευση του training set με ομαλό τρόπο όπως επίσης και τη μέγιστη μείωση στη loss function που μπορούμε να δούμε στην παρακάτω εικόνα.



Εικόνα 6-4. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#2

Όπως βλέπουμε και στη Loss function μετά την εποχή 69 έχουμε μεγάλη αύξηση με μη ομαλό τρόπο, που μας καταδεικνύει ότι η εκπαίδευση έχει παρεκτραπεί. Ένα επιπλέον στοιχείο είναι ότι παρόλο στο accuracy φαίνεται σαν να έχουμε κάποια θετικές και αρνητικές μεταβολές, το διάγραμμα της loss function είναι αμείλικτο και μας δείχνει ότι η loss function μόνο χειροτέρεψε μετά την εποχή 61.

Πίνακας 10. Μετρικές αποτελεσμάτων για την περίπτωση VGG#2 , μετά την 61ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0,9554	0,6766	0,7922	0,7798	0,8978

Παρατηρούμε ότι σε σχέση με το VGG16#1 έχουμε μια βελτίωση ειδικά στο precision της τάξης του 5% και μια μικρότερη αύξηση στο recall της τάξης του 1.5%. Τα παραπάνω οδηγούν το accuracy του δικτύου αυτού να αυξηθεί κατά 1.5%. Συνεπώς μπορούμε να συμπεράνουμε ότι το να εκπαιδεύσουμε τα χαρακτηριστικά του δικτύου ώστε να προσαρμοστούν στα χαρακτηριστικά του δικού μας προβλήματος οδηγεί σε βελτίωση του τελικού δικτύου. Η βελτίωση αυτή όμως παρατηρούμε ότι αν και είναι σημαντική, δεν είναι

όσο σημαντική θα περιμέναμε a priori. Αυτό καταδεικνύει τη δύναμη της μεταφοράς γνώσης καθώς επίσης και τη «στιβαρότητα» των χαρακτηριστικών που επιδεικνύεται από μια εκπαίδευση δικτύου σε δεδομένα γενικού τύπου.

Στη συνέχεια παραθέτουμε τον πίνακα σύγκρισης για το δίκτυο μετά την 61^η εποχή εκπαίδευσης. Παρατηρούμε χαρακτηριστικά, τη μεγάλη μείωση των false positive, με την παράλληλα μικρή αύξηση των false negative. Το παραπάνω είναι κάπως ανησυχητικό μιας και τα false negative σε ιατρικές εφαρμογές, όπως αυτή που ασχολούμαστε στο παρών πρόβλημα, είναι σημαντικότερα από τα false positive.

Πίνακας 11. Αποτελέσματα της τροποποίησης VGG16#2 σε μορφή πίνακα σύγκρισης, μετά την 61η εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2251	1076
True Negative	105	8129

Ως συμπέρασμα μπορούμε να συνάγουμε ότι η εκπαίδευση των συνελκτικών μερών ενός δικτύου μπορεί να πάρει προαιρετικό χαρακτήρα, σε περιπτώσεις που ο χρόνος δεν αποτελεί σημαντικό παράγοντα. Ειδάλως η αντιγραφή αυτούσιων των χαρακτηριστικών με εκπαίδευση μόνο του πλήρως συνδεδεμένου δικτύου είναι ικανή να μας παρέχει ένα δίκτυο με αυξημένη διακριτική ικανότητα.

6.4 Αποτελέσματα για VGG16#3

Για να περιγράψουμε εν συντομία την συγκεκριμένη τροποποίηση χρησιμοποιήσαμε τα ίδια βάρη με το αρχικό μας δίκτυο VGG16, αλλά με τη διαγραφή του τελευταίου επιπέδου (από pooling layer σε pooling layer) και των αντίστοιχων βαρών του για τα συνελκτικά επίπεδα. Επειδή το πλήρες συνδεδεμένο επίπεδο συνδέεται ουσιαστικά με νέα χαρακτηριστικά δεν υπήρχε νόημα να κρατηθούν τα βάρη στο πλήρως συνδεδεμένο επίπεδο και αρχικοποιήθηκαν τυχαία.

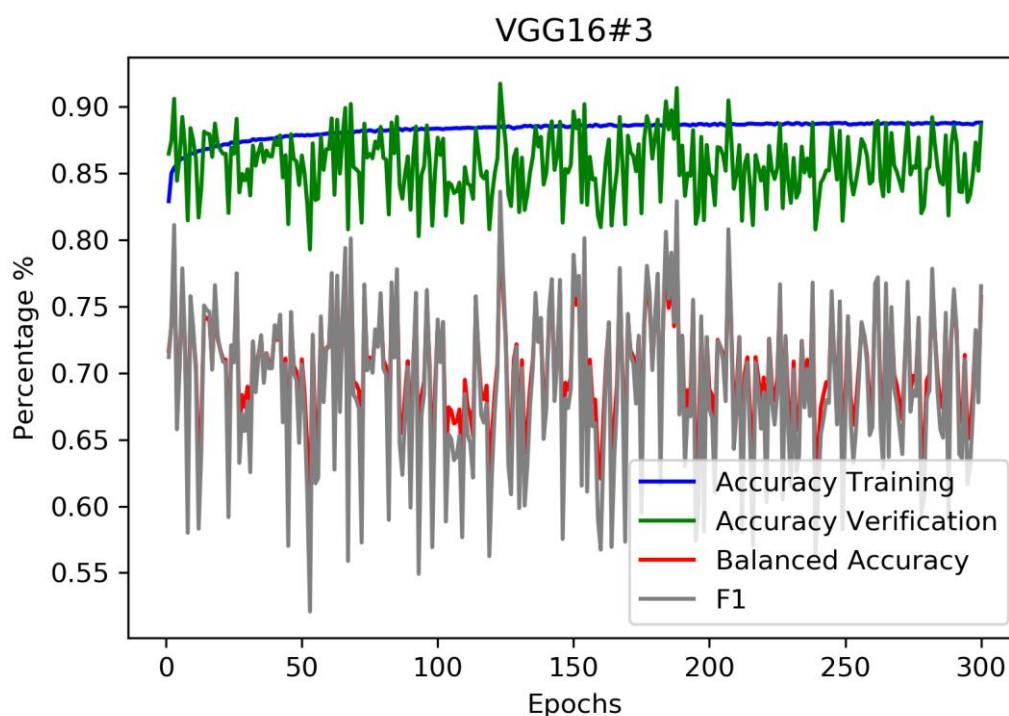
Οι παράμετροι που χρησιμοποιήθηκαν για τον αλγόριθμο ελαχιστοποίησης φαίνονται στον παρακάτω πίνακα.

Πίνακας 12. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#3

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
-----------------------------------	-----------------------------

Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	300
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Το διάγραμμα που λάβαμε για την ακρίβεια και τις υπόλοιπες μετρικές κατά την εκπαίδευση του δικτύου, ελεγμένο και στο training set αλλά και στο validation set παρατίθεται παρακάτω.

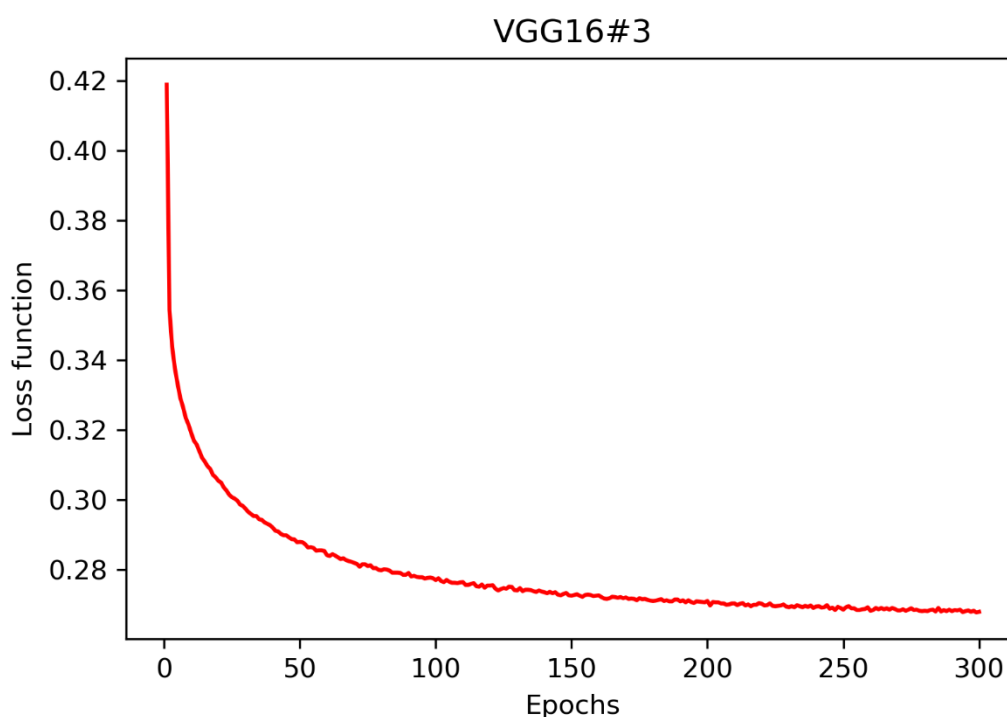


Εικόνα 6-5. Accuracy και άλλες μετρικές για το δίκτυο VGG16#3

Καταρχάς να διευκρινίσουμε ότι λόγω του γεγονότος ότι το πλήθος των εποχών εκπαίδευσης είναι πλέον αυξημένο στις 300, το γράφημα απεικονίζεται να έχει

μεγαλύτερες οι διακυμάνσεις από ότι στα γραφήματα για τα προηγούμενα μοντέλα που εξετάσαμε.

Αρχικά παρατηρούμε ότι στο training set το accuracy αρχίζει από 81% το οποίο συγκριτικά με την αρχή του VGG16#1 είναι υψηλότερο. Επίσης, μετά από τις πρώτες 60 εποχές εκπαίδευσης το training accuracy έχει φτάσει στο 88%, ενώ μετά από 300 εποχές το 88.82%, με πολύ ομαλή σύγκλιση που υποδεικνύει τοπικό ελάχιστο. Ενδιαφέρον παρουσιάζει και το γράφημα για το accuracy στο validation set, το οποίο αμφιταλαντεύεται γύρω από μια κεντρική τιμή στο 86% καθ' όλη τη διάρκεια εκπαίδευσης. Επίσης οι μετρικές BAC και F1 κυμαίνονται στο 70% χωρίς ιδιαίτερες αυξομειώσεις. Τα παραπάνω γεγονότα καταδεικνύουν ότι στο τελικό αποτέλεσμα, δεν αναμένουμε η παρούσα τροποποίηση να παρουσιάζει εντυπωσιακή βελτίωση στην απόδοση του δικτύου μας.



Εικόνα 6-6. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#3

Κατά την εκπαίδευση του δικτύου μας παρατηρούμε ότι είχαμε μια συνεχόμενη μείωση της συνάρτησης με πολύ ομαλό τρόπο που καταδεικνύει το γεγονός ότι οι παράμετροι εκπαίδευσης για το συγκεκριμένο δίκτυο μας οδήγησαν με ομαλό τρόπο σε τοπικό ελάχιστο.

Πίνακας 13. Μετρικές αποτελεσμάτων για την περίπτωση VGG#3 , μετά την 300η εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0.9464	0.6423	0,7653	0,7572	0,8866

Παρατηρούμε ότι σε όλες τις μετρήσεις έχουμε μείωση των επιδόσεων του δικτύου από σε σχέση με το VGG#1 που χρησιμοποίησε τα ίδια χαρακτηριστικά αλλά είχε και το τελευταίο επίπεδο συνέλιξης για τα θεωρούμενα ως τα πλέον περίπλοκα χαρακτηριστικά. Το παραπάνω μας κάνει να πιστεύουμε με μεγαλύτερη σιγουριά, ότι για ιατρικά προβλήματα με ανάλυση εικόνας που χρειάζονται διακριτική ικανότητα σε χαρακτηριστικά που βρίσκονται σε όλη την εικόνα, με ανάλυση υφής, πλήθους εμφανίσεων καθώς επίσης και έντασης στο σήμα που λαμβάνουμε η ύπαρξη πολλών επιπέδων είναι ωφέλιμη για το αποτέλεσμα που λαμβάνουμε από το δίκτυο.

Παρακάτω παραθέτουμε το confusion matrix για το validation set που λάβαμε μετά από 300 εποχές εκπαίδευσης. Συγκρίνοντας το με τα προηγούμενα αποτελέσματα που έχουμε λάβει ως τώρα, μπορούμε με ασφάλεια να δούμε ότι έχει μεγάλη ομοιότητα με το αποτέλεσμα που λάβαμε στο VGG16#2, με 10πλάσια εμφάνιση type II λαθών σε σχέση τα αντίστοιχα type I, με ελαφρώς χειρότερη απόδοση και στη θετική αλλά και στην αρνητική κατηγορία.

Πίνακας 14. Αποτελέσματα της τροποποίησης VGG16#3 σε μορφή πίνακα σύγχυσης, μετά την 300ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2137	1190
True Negative	121	8113

Συνοψίζοντας, το συγκεκριμένο δίκτυο εκπαιδεύτηκε ομαλά, και η απόδοση του είναι ελαφρώς υποδεέστερη του αντίστοιχου VGG16#1 μετά από την 44^η εποχή εκπαίδευσης του.

6.5 Αποτελέσματα για VGG16#4

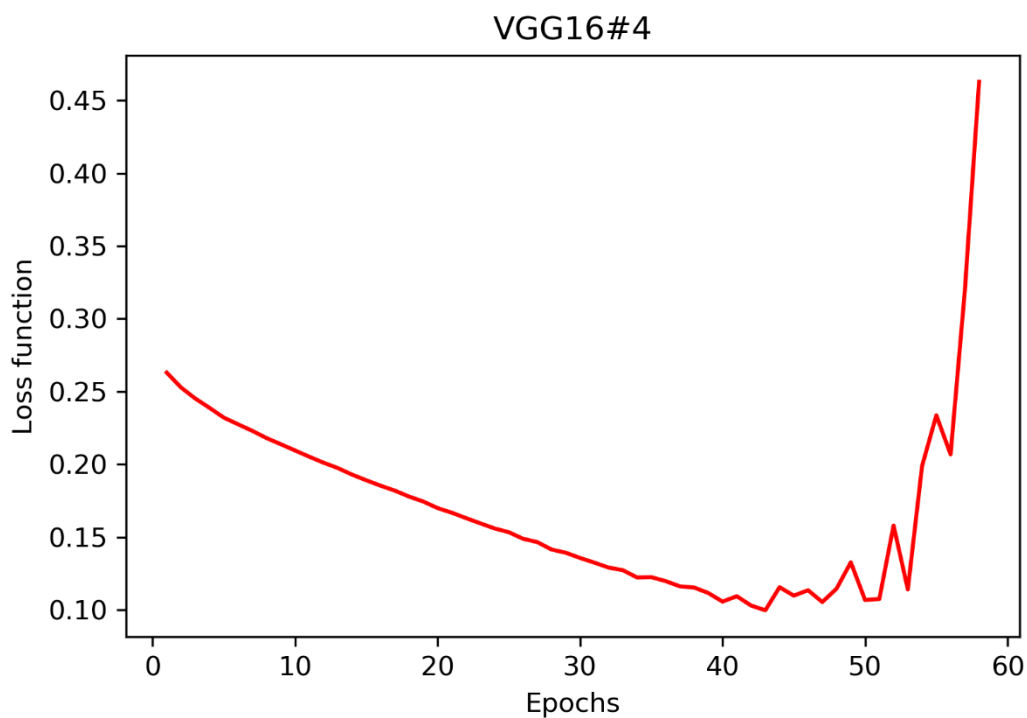
Περιγράφεται εν συντομία η συγκεκριμένη τροποποίηση. Χρησιμοποιήσαμε την ίδια ακριβώς αρχιτεκτονική με το δίκτυο VGG16#3, στην οποία έχουμε αφαιρέσει το τελευταίο συνελκτικό επίπεδο και έχουμε εκπαιδεύσει το πλήρως συνδεδεμένο επίπεδο. Στην παρούσα τροποποίηση εξετάζουμε κατά πόσο εάν εκπαιδεύσουμε πλέον μόνο τα «χαρακτηριστικά» του, δηλαδή το συνελκτικό επίπεδο θα μας παράσχει βελτίωση.

Πίνακας 15. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#4

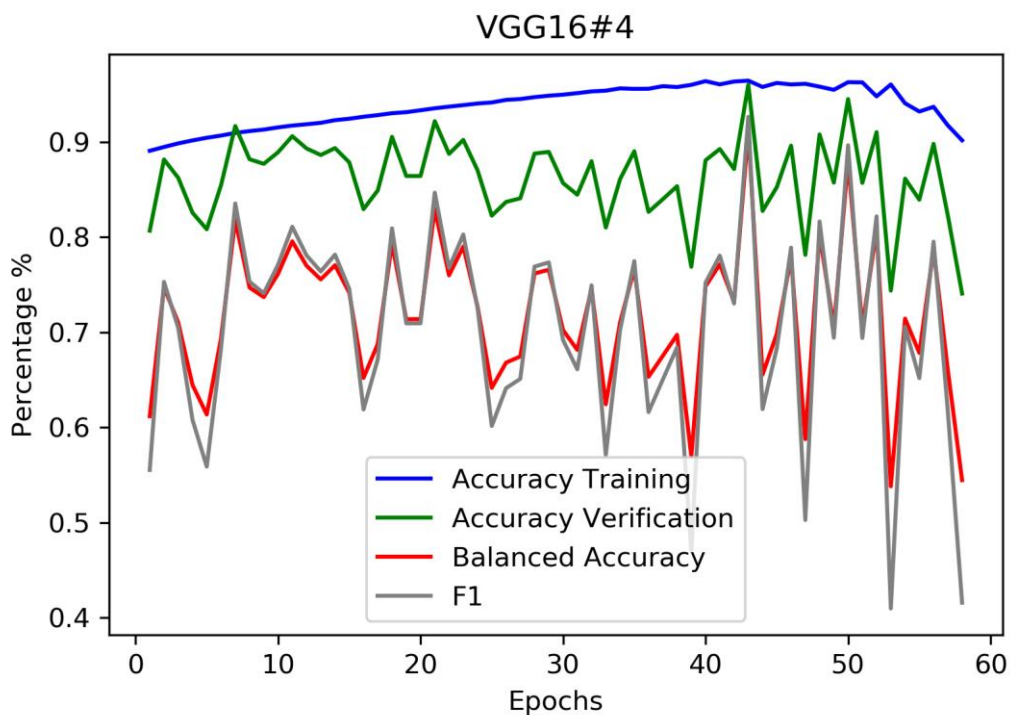
Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	58
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Αρχικά ας εξετάσουμε τη loss function για τη συγκεκριμένη τροποποίηση η οποία δεν έχει το αναμενόμενο και επιθυμητό σχήμα. Παρατηρούμε ότι μέχρι την 43 εποχή έχουμε μια συνεχόμενη μείωση της τιμής της συνάρτησης, η οποία αρχίζει από χαμηλά επίπεδα, πράγμα αναμενόμενο μιας και το πλήρως συνδεδεμένο επίπεδο αλλά και το συνελκτικό επίπεδο δεν έχουν τυχαία βάρη. Στη συνέχεια όμως παρατηρούμε ότι η loss function αντί να μειώνεται αυξάνεται ραγδαία. Το γεγονός αυτό μας οδηγεί στο να λάβουμε ως τελευταίο εκπαιδευμένο μοντέλο, αυτό που παρατηρήθηκε μετά την 43^η εποχή εκπαίδευσης. Ο λόγος για τον οποίο παρατηρείται το παραπάνω μπορεί να έχει να κάνει με τις παραμέτρους εκπαίδευσης όσο και με το ίδιο το σημείο στο χώρο παραμέτρων, κάτι που είναι λιγότερο πιθανό. Μια άλλη πιθανή εξήγηση μπορεί να είναι το γεγονός ότι ένα μέρος του δικτύου παρέμεινε με τα βάρη παγωμένα, γεγονός που μπορεί να προκαλέσει αστάθεια, αν λάβουμε υπόψιν ότι ο αλγόριθμος εκμάθησης μεταβάλλει μόνο

ότι μπορεί και μετά το πλήρες δίκτυο παράγει το αποτέλεσμα κάτι που μπορεί να αποτελέσει πηγή αστάθειας κατά τη διάρκεια της εκπαίδευσης.



Εικόνα 6-7. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#4



Εικόνα 6-8. Accuracy και άλλες μετρικές για το δίκτυο VGG16#4

Παραπάνω παραθέτουμε το διάγραμμα με τις μετρικές εκπαίδευσης που συλλέχθηκαν για το συγκεκριμένο μοντέλο.

Παρατηρούμε ότι το accuracy στο training set είναι ομοίως με τη συνάρτηση απώλειας ομαλό μέχρι τη 43^η εποχή εκπαίδευσης. Στο validation set παρατηρούμε ότι αρχίζουμε από επίπεδα του 80% και με μικρή σχετικά διακύμανση μέχρι την εποχή 43 βελτιώνεται και αγγίζει και το 96% το οποίο είναι ένα πολύ καλό αποτέλεσμα. Σε ότι αφορά στις μετρικές F1 και BAC λόγω του ότι εξαρτώνται από το accuracy αλλά και από το σε ποια κατηγορία παρατηρούμε τα σφάλματα, έχουν πολύ μεγαλύτερη διακύμανση. Ένα επιπλέον χαρακτηριστικό των μετρικών είναι ότι η μετρική F1 παίρνει πιο ακραίες τιμές από το BAC όταν έχουμε είτε πολύ καλές ή πολύ κακές εκτιμήσεις.

Πίνακας 16. Μετρικές αποτελεσμάτων για την περίπτωση VGG#4 , μετά την 43η εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0.9867	0.8729	0,9263	0,9119	0,9600

Τα παραπάνω αποτελέσματα είναι πλέον εντυπωσιακά για όλες τις μετρικές, ίσως εξαιρουμένης του recall. Το μόνο ψεγάδι που μπορούμε να αναφέρουμε είναι το γεγονός της μη ομαλής άφιξης στο παραπάνω μοντέλο, λόγω της «εκτροπής» της loss function μετά το εξεταζόμενο μοντέλο.

Πίνακας 17. Αποτελέσματα της τροποποίησης VGG16#4 σε μορφή πίνακα σύγχυσης, μετά την 43ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2904	423
True Negative	39	8195

Το συμπέρασμα που μπορούμε να βγάλουμε είναι ότι η εκπαίδευση του δικτύου και στα χαρακτηριστικά του, βελτιώνει την διακριτική ικανότητα του δικτύου. Το παραπάνω παρατηρήθηκε και στο VGG16#2 σε σχέση με το αρχικό του, αλλά και στο παρόν μοντέλο σε σχέση με το αρχικό του. Ερμηνευμένο το παραπάνω, σημαίνει ότι το να λάβουμε τα αρχικά βάρη για τα χαρακτηριστικά και να εκπαιδεύσουμε το δίκτυο είναι σίγουρα πολύ βοηθητικό, όμως θα πρέπει να επανέλθουμε και να εκπαιδεύσουμε το δίκτυο και στο

συνελκτικό μέρος που παράγει τα χαρακτηριστικά, ώστε να επιτρέψουμε στο δίκτυο να εντοπίσει τέτοια χαρακτηριστικά που να είναι χρήσιμα για το ιατρικό πρόβλημα που αντιμετωπίζουμε και όχι για τις γενικές κλάσεις με τις οποίες εκπαιδεύτηκε. Επίσης ένα επιπλέον συμπέρασμα που μπορούμε να εξάγουμε είναι ότι το βάθος δεν είναι ο μόνος σημαντικός παράγοντας, αλλά εξαρτάται από το πλήθος των κλάσεων που πρέπει να διαχωριστούν, μιας και τα αποτελέσματα που λάβαμε είναι καλύτερα από την περίπτωση που χρησιμοποιήθηκε το σύνολο του βάθους του αρχικού δικτύου VGG16.

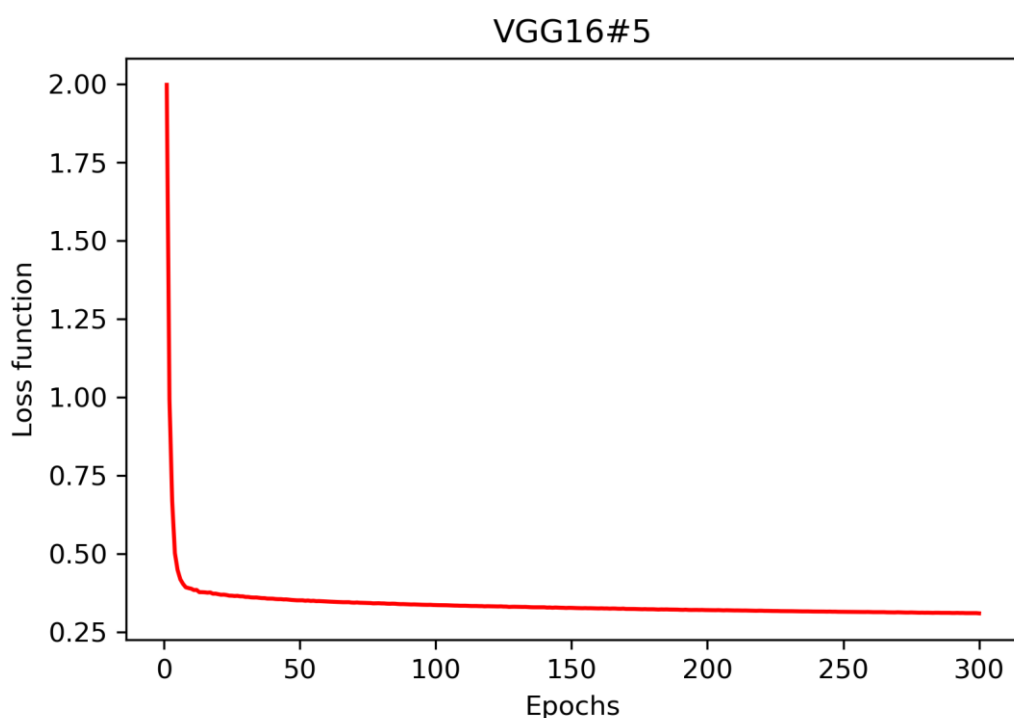
6.6 Αποτελέσματα για VGG16#5

Περιγράφουμε εν συντομία την συγκεκριμένη τροποποίηση. Σκοπός της παρούσας τροποποίησης είναι να εξεταστεί η περίπτωση όπου αφαιρούνται τα 2 τελευταία συνελκτικά επίπεδα του VGG16 δικτύου το οποίο πλέον μεταβάλλει σημαντικά το βάθος του δικτύου και μειώνει το χρόνο εκπαίδευσης του. Σκοπός είναι να δούμε τι επίδραση θα είχε στη διαχωριστική ικανότητα του δικτύου. Στην παρούσα τροποποίηση για να ελέγξουμε την ικανότητα του δικτύου πρέπει να εκπαιδεύσουμε το πλήρως συνδεδεμένο μέρος του δικτύου μας. Για αυτό το λόγο κρατήθηκαν τα βάρη του VGG16 ίδια, και αρχικοποιήθηκαν με τυχαίες τιμές το πλήρως συνδεδεμένο μέρος, το οποίο ήταν και το μόνο κομμάτι του δικτύου που επιτράπηκε να εκπαιδευτεί.

Πίνακας 18. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#5

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	300
Ταχύτητα εκμάθησης (learning rate)	0.00001
Χρήση ορμής εκμάθησης (momentum)	0.1
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Ακολουθεί το διάγραμμα για την συνάρτηση απώλειας, στην οποία παρατηρούμε ότι αρχίζει από πολύ υψηλές τιμές, το οποίο είναι αναμενόμενο μιας και το πλήρως συνδεδεμένο επίπεδο έχει αρχικοποιηθεί με τυχαίες τιμές. Επίσης παρατηρούμε πολύ γρήγορη σύγκλιση στο τοπικό ελάχιστο που εντοπίστηκε. Μετά από 30 εποχές εκπαίδευσης είναι πολύ κοντά στο τελικό αποτέλεσμα που λάβαμε μετά από τις 300 εποχές εκπαίδευσης. Επίσης παρατηρούμε ότι είναι πολύ ομαλή, γεγονός που μπορεί και να οφείλεται στη μείωση της ορμής που επιλέχθηκε για την εκπαίδευση του συγκεκριμένου δικτύου.

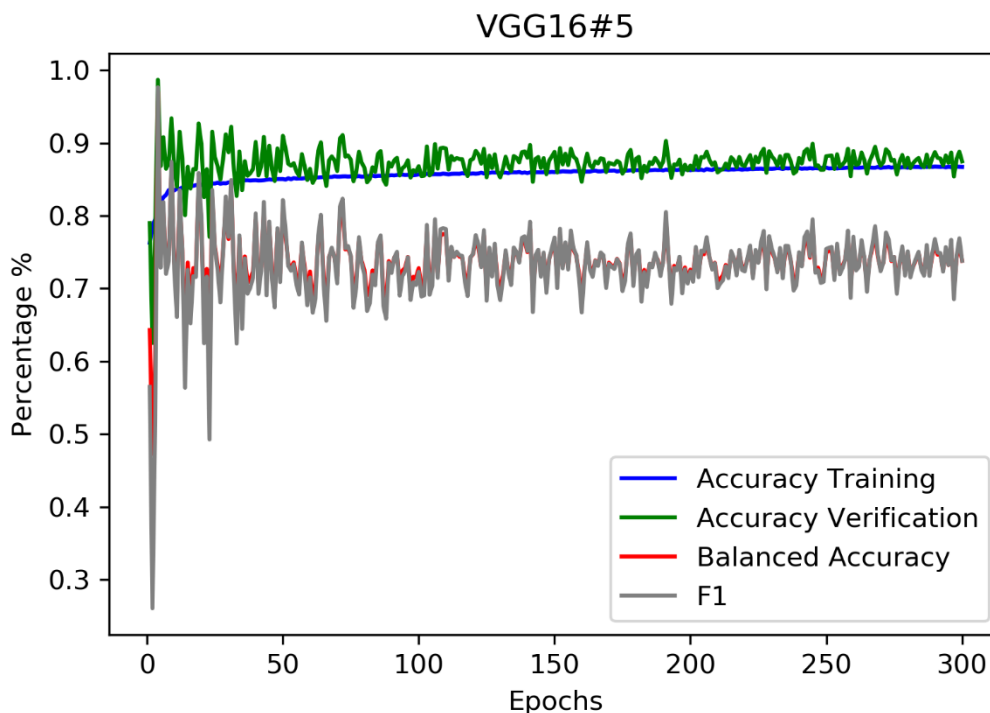


Εικόνα 6-9. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#5

Η εικόνα που παρατηρούμε στις μετρικές accuracy ακολουθούν την ίδια ομαλότητα που παρατηρείται και στη συνάρτηση απώλειας. Στις πρώτες 40 εποχές υπάρχει μεγαλύτερη διακύμανση στο validation accuracy και μετά καθώς έχει κατασταλάξει στο τοπικό ελάχιστο παρατηρούμε πολύ μικρές αλλαγές. Σε ότι αφορά το training accuracy παρατηρούμε μια διαρκή αύξηση με μειούμενο ρυθμό, όπως θα περιμέναμε βλέποντας το διάγραμμα της loss function. Οι μετρικές BAC και F1 έχουν ελαφρώς μεγαλύτερη διακύμανση από το validation accuracy, πράγμα αναμενόμενο, μιας και επηρεάζονται και από το σε ποια κλάση γίνεται το κάθε σφάλμα κατηγοριοποίησης. Επειδή η εκπαίδευση έγινε ομαλά θεωρούμε ως τελικό μοντέλο για το παρόν δίκτυο το μοντέλο που λάβαμε μετά την 300^ή εποχή εκπαίδευσης.

Πίνακας 19. Μετρικές αποτελεσμάτων για την περίπτωση VGG#5 , μετά την 300ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0.9265	0.6135	0,7382	0,7380	0,8748



Εικόνα 6-10. Accuracy και άλλες μετρικές για το δίκτυο VGG16#5

Συγκρινόμενο με όλες τις προηγούμενες υλοποιήσεις παρατηρούμε ότι τα αποτελέσματα που λαμβάνουμε είναι ελαφρώς υποδεέστερο σε σχέση με τις αντίστοιχες (ως προς το ότι τα χαρακτηριστικά του συνελκτικού μέρους δεν έχουν εκπαιδευτεί) τροποποιήσεων που έχουν δοκιμαστεί έως αυτό το σημείο. Σε σχέση με το VGG16#1 υπολείπεται κατά 1% και σε σχέση με το VGG16#3 που αφαιρέσαμε μόνο ένα συνελκτικό επίπεδο υπολείπεται κατά 2% στο validation accuracy. Επίσης παρατηρούμε από τον πίνακα σύγκυσης που ακολουθεί ότι τα περισσότερα σφάλματα που κάνει το δίκτυο αυτό είναι type II, το οποίο δεν είναι επιθυμητό.

Πίνακας 20. Αποτελέσματα της τροποποίησης VGG16#5 σε μορφή πίνακα σύγκυσης, μετά την 300ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2041	1286

True Negative	162	8072
----------------------	-----	------

Συμπεραίνουμε ότι η εκπαίδευση αυτού του δικτύου μπορεί να δώσει συγκρίσιμα με τις προηγούμενες τροποποιήσεις αποτελέσματα, με ελαφρώς μικρότερη απόδοση στο accuracy, αλλά λόγω μικρού μεγέθους του δικτύου είναι πολύ πιο γρήγορο στην εκπαίδευση, καθώς επίσης και εκπαιδεύεται πολύ πιο ομαλά λόγω λιγότερων επιπέδων μη γραμμικότητας (μείωση του βάθους του δικτύου). Μια τέτοιου είδους αντιμετώπιση είναι πιθανό επιθυμητή για προβλήματα όπου χρειαζόμαστε άμεσα αποτελέσματα, αλλά πιθανότατα όχι για ιατρικά προβλήματα όπου η μέγιστη ακρίβεια είναι αυτοσκοπός.

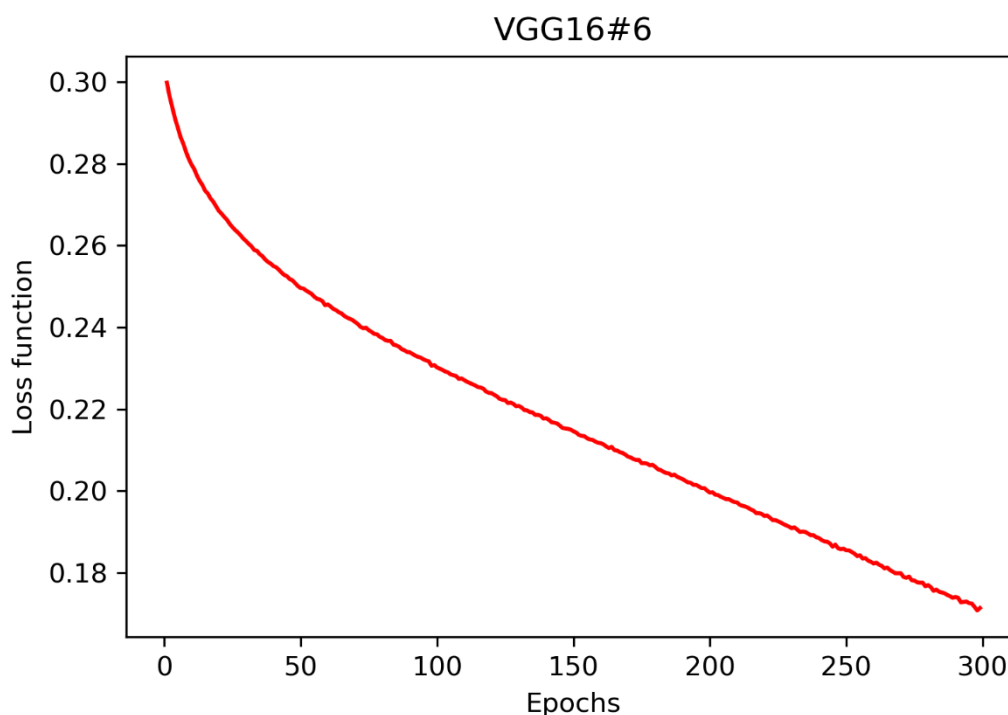
6.7 Αποτελέσματα για VGG16#6

Περιγράφεται εν συντομία η συγκεκριμένη τροποποίηση. Σκοπός της παρούσας τροποποίησης είναι να εξετάσουμε αφού έχουμε αφαιρέσει τα 2 τελευταία συνελκτικά επίπεδα του VGG16 δικτύου και έχουμε εκπαιδεύσει και το πλήρως συνδεδεμένο επίπεδο, εάν θα υπάρξει βελτίωση με την εκπαίδευση μόνο των χαρακτηριστικών του δικτύου. Να σημειωθεί ότι το δίκτυο αρχικοποιείται από το VGG16#5 σε όλα τα βάρη του δικτύου, όμως εκπαιδεύεται μόνο το συνελκτικό μέρος του δικτύου.

Πίνακας 21. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#6

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	300
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Ακολουθεί το διάγραμμα για την συνάρτηση απώλειας, στην οποία παρατηρούμε ότι αρχίζει από σχετικά χαμηλές τιμές, πράγμα αναμενόμενο μιας και το δίκτυο ουσιαστικά αρχικοποιείται από το VGG16#5. Παρατηρούμε ότι στις πρώτες 40 εποχές έχουμε μια σχετικά ταχεία μείωση της συνάρτησης απώλειας και στη συνέχεια μειώνετε με σχεδόν σταθερό ρυθμό, μέχρι και την 300ή εποχή που είχαμε θέσει ως όριο για την εκπαίδευση.

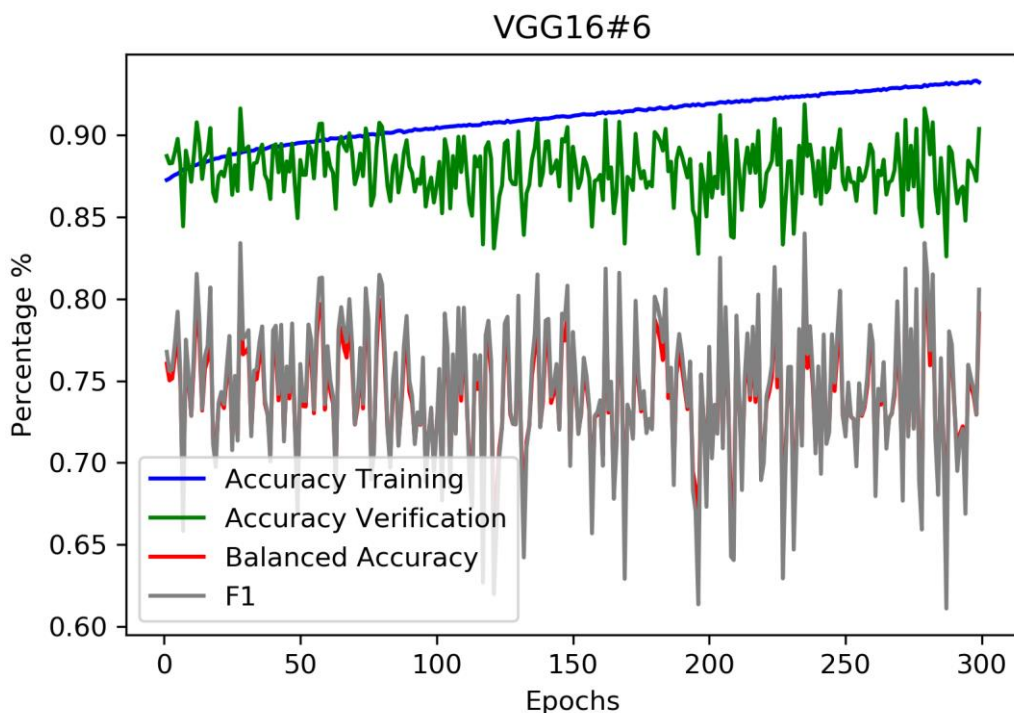


Εικόνα 6-11. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#6

Η εικόνα που παρατηρούμε στη μετρική accuracy για το training είναι εντυπωσιακά ομαλές. Αυτό έχει να κάνει πιθανότατα και με το γεγονός ότι πλέον το δίκτυο εκπαιδεύει μόνο τρία συνελκτικά επίπεδα, οπότε οι μεταβολές στο δίκτυο είναι μικρότερες. Παρατηρούμε ότι παρά την ξεκάθαρη και συνεχή βελτίωση στη συνάρτηση απώλειας στο validation set δεν έχουμε την αναμενόμενη βελτίωση.

Ενδεικτικά στην 5^η εποχή το validation accuracy είναι 89.76% και μετά από ακόμα 285 εποχές εκπαίδευσης το συναντάμε στο 90.37%, χωρίς ενδείξεις overfitting με μικρή βελτίωση όμως βελτίωση παρόλα αυτά. Τα παραπάνω σημαίνουν ότι με λιγότερα συνελκτικά επίπεδα χρειαζόμαστε λιγότερο χρόνο για εκμάθηση και το τοπικό ελάχιστο εντοπίζεται πολύ πιο γρήγορα. Επίσης όμως είναι εξίσου πιθανόν αυτό το τοπικό ελάχιστο να είναι λιγότερο επιθυμητό από αυτό που προκύπτει από ένα δίκτυο περισσότερων συνελκτικών επιπέδων με λιγότερες εποχές εκπαίδευσης.

Αντίστοιχα αποτελέσματα παρατηρούμε και στις μετρικές BAC και F1 όπου επίσης παρατηρούμε μεγαλύτερη διακύμανση από την ακρίβεια, όπως και στα προηγούμενα μοντέλα, όμως δίχως να παρατηρούμε αύξηση στις μετρικές με την πάροδο των εποχών. Επειδή η συνάρτηση απώλειας συνεχίζει να μειώνεται μέχρι και την τελευταία εποχή, θεωρούμε ως τελευταία εποχή την 300ή.



Εικόνα 6-12. Accuracy και άλλες μετρικές για το δίκτυο VGG16#6

Ακολουθούν τα αποτελέσματα όπως μετρήθηκαν στο τελευταίο μοντέλο, μετά το πέρας 300 εποχών εκπαίδευσης.

Πίνακας 22. Μετρικές αποτελεσμάτων για την περίπτωση VGG#6 , μετά την 416ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0.9613	0.6934	0.8056	0.7910	0.9037

Χαρακτηριστικό είναι ότι σε σχέση με το μοντέλο VGG16#5 με το οποίο αρχικοποιήθηκε έχουμε μια βελτίωση στο accuracy 2.89% και στη μετρική BAC 5.3%, όπως και στην μετρική F1 κατά 6.74%. Συνεπώς η εκπαίδευση των χαρακτηριστικών (συνελκτικά επίπεδα) απέδωσε καρπούς, και προτείνεται για την βελτίωση της κατηγοριοποίησης των δικτύων. Επίσης παρατηρούμε ότι όσο μικρότερο το μέγεθος του δικτύου που

καταλήγουμε τόσο σημαντικότερο είναι να εκπαιδευτούν τα χαρακτηριστικά, αφού καλούνται να μεταβληθούν ώστε να μάθουν χαρακτηριστικά που μαθαίνει το δίκτυο φυσιολογικά σε μεταγενέστερα επίπεδα. Ακολουθεί ο πίνακας σύγκρισης για το τελικό μοντέλο, στο validation set.

Πίνακας 23. Αποτελέσματα της τροποποίησης VGG16#6 σε μορφή πίνακα σύγκρισης, μετά την 200ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	2307	1020
True Negative	93	8141

6.8 Αποτελέσματα για VGG16#7

Περιγράφεται εν συντομία την συγκεκριμένη τροποποίηση. Σκοπός της παρούσας τροποποίησης είναι να εξετάσουμε εάν αφού έχουμε ένα δίκτυο που έχει εκπαιδευτεί και στο πλήρως συνδεδεμένο μέρος του, εάν έχει νόημα το να μην εκπαιδεύσουμε το πλήρες σετ από τα «χαρακτηριστικά» που παράγονται από τα πρωταρχικά επίπεδα αλλά να εξετάσουμε μόνο τα πλέον περίπλοκα και συγκεκριμένα για το εκάστοτε πρόβλημα. Για να κάνουμε το παραπάνω ουσιαστικά πήραμε τα βάρη και την αρχιτεκτονική που χρησιμοποιήθηκε στο VGG16#1 και εκπαιδεύσαμε μόνο τα 2 τελευταία συνελκτικά επίπεδα, κρατώντας και το πλήρες συνδεδεμένο μέρος του δικτύου, αλλά και τα 3 πρώτα συνελκτικά επίπεδα με τα βάρη τους παγωμένα.

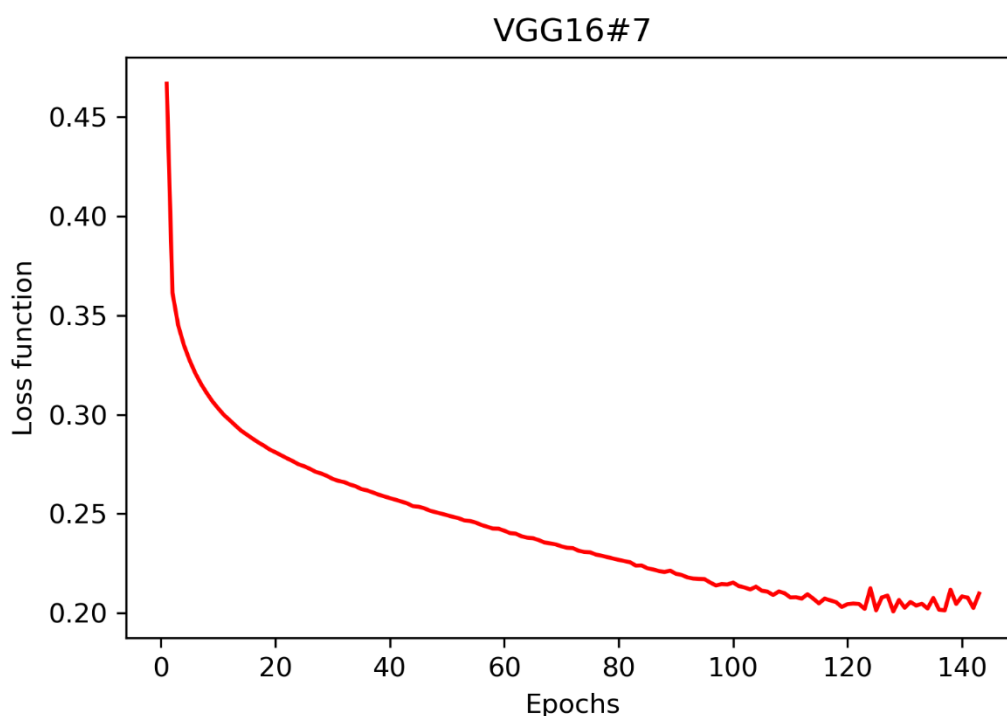
Πίνακας 24. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#7

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	143
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι

Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Ακολουθεί το διάγραμμα για την συνάρτηση απώλειας, στην οποία παρατηρούμε ότι αρχίζει από σχετικά ψηλές τιμές, αλλά πολύ χαμηλότερα από δίκτυα που δεν είναι εκπαιδευμένο το πλήρες συνδεδεμένο μέρος του δικτύου. Παρατηρούμε μια αρχικά ταχεία πτώση της συνάρτησης απώλειας, όμως στη συνέχεια μειώνετε σχεδόν γραμμικά από την 15ή εποχή μέχρι και την 120ή. Στη συνέχεια παρατηρούμε ότι επιπλέον εκπαίδευση παύει να μειώνει τη συνάρτηση απώλειας και απλά αυξομειώνεται γύρω από την ίδια τιμή. Το παραπάνω σημαίνει ότι πιθανόν βρέθηκε τοπικό ελάχιστο και συνεπώς ολοκληρώθηκε η διαδικασία της εκπαίδευσης.

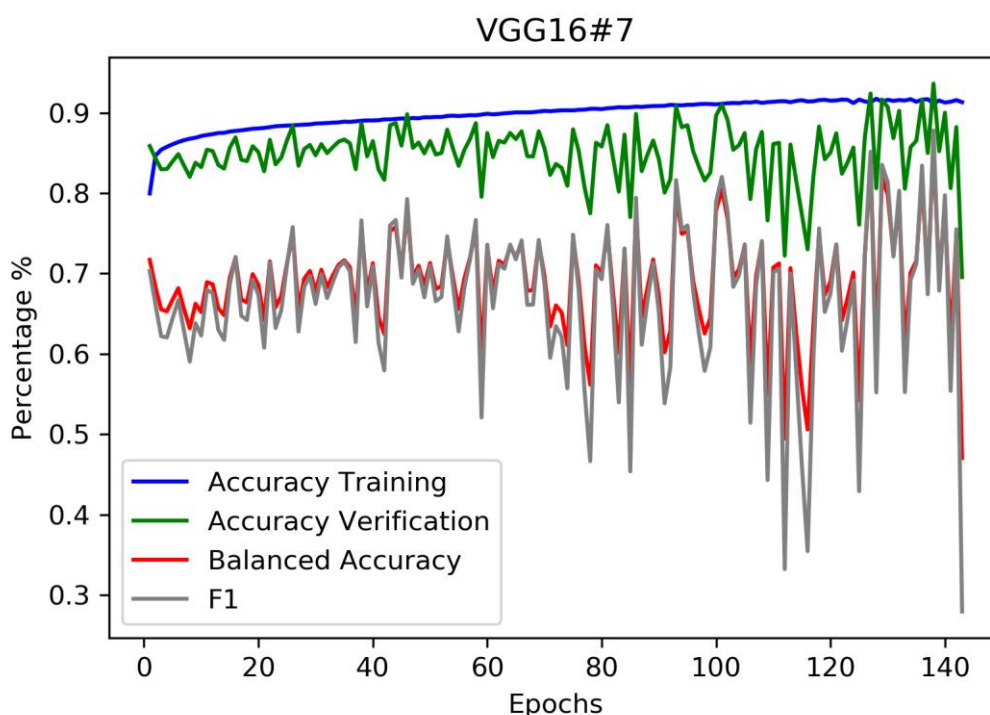
Η εποχή που ορίσαμε ως τελευταία εκπαιδευμένη είναι αυτή στην οποία πετύχαμε την ελάχιστη τιμή στη συνάρτηση απώλειας, η οποία στη συγκεκριμένη τροποποίηση είναι η 137ή.



Εικόνα 6-13. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#7

Στις μετρικές accuracy παρατηρούμε αρχικά ότι σε ότι αφορά το training accuracy ακολουθεί της συνάρτηση απώλειας αντεστραμμένη. Στις πρώτες 10 εποχές ραγδαία

αύξηση, στη συνέχεια σχεδόν γραμμική μικρή βελτίωση μέχρι την εποχή 120 και τέλος μια στασιμότητα. Σε αντίθεση, στο validation accuracy παρατηρούμε ότι μέχρι την 70ή εποχή έχουμε μια αύξηση σχεδόν γραμμική με μικρές διακυμάνσεις. Στη συνέχεια παρατηρούμε μια πτώση μέχρι και την 120ή εποχή, με παράλληλη αύξηση της διακύμανσης. Μετά την 120ή εποχή παρατηρούμε αύξηση του validation accuracy σε νέα μέγιστα, με μία φαινομενική σταθερότητα σε ψηλά επίπεδα. Οι μετρικές BAC και F1 έχουν πολύ μεγαλύτερη διακύμανση από το validation accuracy και από προηγούμενα μοντέλα, πράγμα που σημαίνει ότι κατά την εκπαίδευση του παρόντος μοντέλου αρκετά παραδείγματα άλλαζαν κλάση σε κάθε εποχή εκπαίδευσης που μπορεί να ερμηνευτεί ως μειονέκτημα της παρούσας διαδικασίας εκπαίδευσης.



Εικόνα 6-14. Accuracy και άλλες μετρικές για το δίκτυο VGG16#7

Πίνακας 25. Μετρικές αποτελεσμάτων για την περίπτωση VGG#7 , μετά την 137ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0,8984	0,5396	0,6742	0,6897	0,8499

Από τα παραπάνω αποτελέσματα παρατηρούμε ότι τελικό αποτέλεσμα είναι χειρότερο κατά 5% στο συνολικό accuracy από το VGG16#2, 12% χειρότερο στην μετρική F1 και 9% χειρότερο στο BAC. Συνεπώς μπορούμε με σχετική ασφάλεια να πούμε ότι παρόλο που τα επίπεδα που είναι πιο βαθιά στο συνελκτικό επίπεδο είναι πιο χρήσιμα στο να βγάλουν συμπεράσματα συγκεκριμένα για το πρόβλημα. Στην παρούσα αρχιτεκτονική φαίνεται ότι η εκπαίδευση του συνόλου του δικτύου παράγει σημαντικά καλύτερα αποτελέσματα. Ενδιαφέρον παρουσιάζει και η σύγκριση του δικτύου με το δίκτυο με το οποίο το αρχικοποιήσαμε, το VGG16#1. Παρατηρούμε ότι ενώ τα δύο δίκτυα έχουν το ίδιο precision, το recall είναι 12% χειρότερο, κάτι που οδηγεί το accuracy 3% χαμηλότερα. Δηλαδή εκπαιδεύσαμε τα τελευταία 2 επίπεδα και λάβαμε δεδομένα χειρότερα από αυτά που αρχίσαμε, κάτι που καταδεικνύει ότι δεν είναι επιτυχημένη η τεχνική του να εκπαιδεύονται απομονωμένα ενδιάμεσα κομμάτια του δικτύου όπως φαίνεται από τη συγκεκριμένη υλοποίηση.

Πίνακας 26. Αποτελέσματα της τροποποίησης VGG16#7 σε μορφή πίνακα σύγχυσης, μετά την 137ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	1795	1532
True Negative	203	8031

Από τον παραπάνω πίνακα σύγχυσης παρατηρούμε ότι έχουν γίνει πολλά περισσότερα σφάλματα και type I αλλά και type II το οποίο δεν αφήνει περιθώρια αμφισβήτησης για τη υποδεέστερη επίδοση που λαμβάνουμε με την παρούσα λογική του δικτύου.

6.9 Αποτελέσματα για VGG16#8

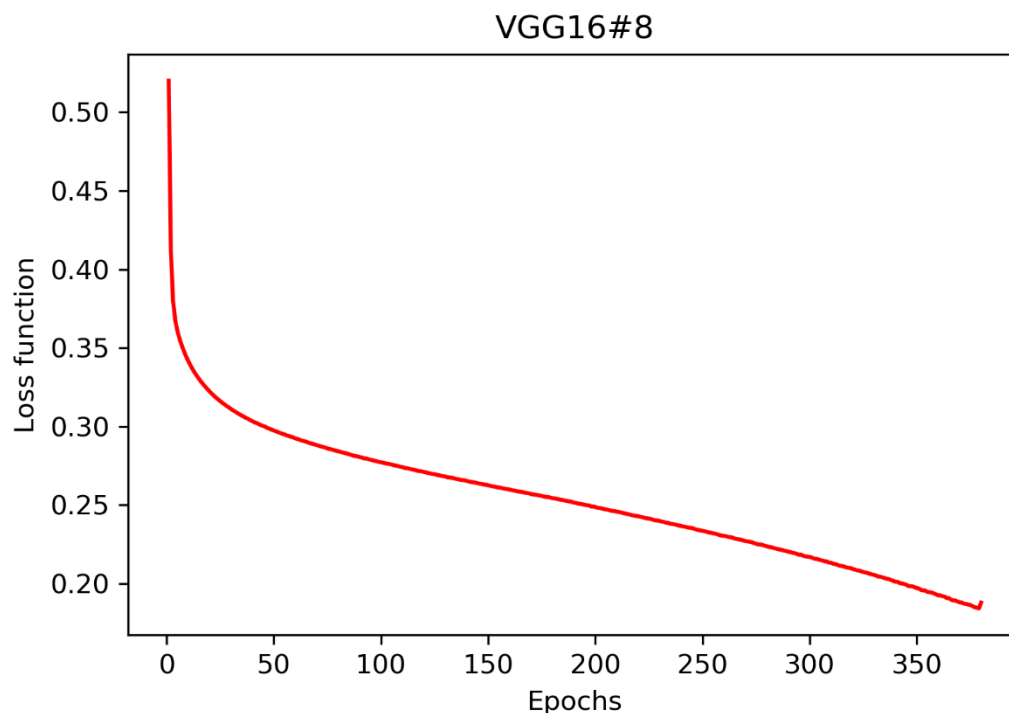
Περιγράφεται εν συντομία την συγκεκριμένη τροποποίηση. Σκοπός της παρούσας τροποποίησης είναι να εξετάσουμε εάν αφού έχουμε ένα δίκτυο που έχει εκπαιδευτεί και στο πλήρως συνδεδεμένο μέρος του υπάρχει νόημα στο να μην εκπαιδεύσουμε το πλήρες σετ από τα «χαρακτηριστικά» που παράγονται από τα πρωταρχικά συνελκτικά επίπεδα αλλά να εξετάσουμε μόνο τα πλέον περίπλοκα και συγκεκριμένα για το εκάστοτε πρόβλημα. Για να κάνουμε το παραπάνω ουσιαστικά πήραμε τα βάρη και την αρχιτεκτονική που χρησιμοποιήθηκε στο VGG16#1 και εκπαιδεύσαμε μόνο το τελευταίο

συνελκτικό επίπεδο, κρατώντας και το πλήρες συνδεδεμένο κομμάτι αλλά και τα 4 αρχικά συνελκτικά επίπεδα με τα βάρη τους παγωμένα. Σκοπός μας είναι να το συγκρίνουμε και με το VGG16#2 αλλά και με το VGG16#7.

Πίνακας 27. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#8

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	380
Ταχύτητα εκμάθησης (learning rate)	0.0002
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesteron	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

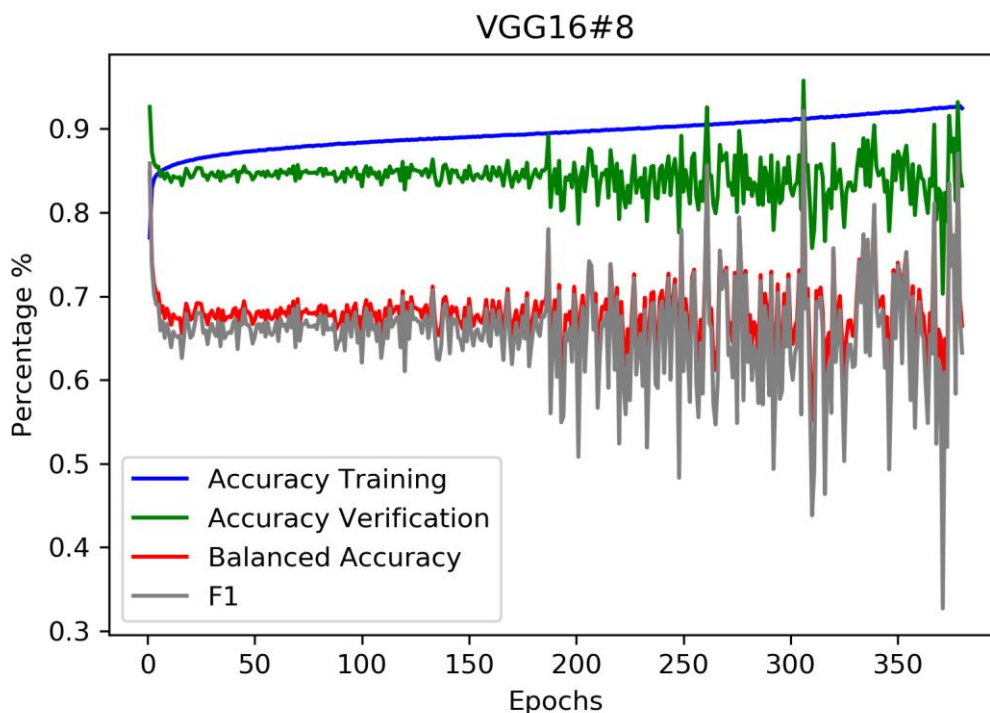
Ακολουθεί το διάγραμμα για την συνάρτηση απώλειας, στην οποία παρατηρούμε ότι αρχίζει από σχετικά ψηλές τιμές. Παρατηρούμε μια αρχικά ταχεία πτώση της συνάρτησης απώλειας μέχρι την 25η εποχή, όμως στη συνέχεια μειώνετε σχεδόν γραμμικά από την 25η εποχή μέχρι και την 379η. Ενδιαφέρον παρουσιάζει το γεγονός ότι η μόνη εποχή που η συνάρτηση απώλειας ήταν αύξουσα ήταν και η τελευταία εποχή εκπαίδευσης. Το παραπάνω συνιστά ένδειξη ότι η εκπαίδευση είναι κοντά στο τοπικό ελάχιστο και δεν σταμάτησε η διαδικασία εκπαίδευσης πρόωρα.



Εικόνα 6-15. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#8

Η εποχή που ορίσαμε ως τελευταία εκπαιδευμένη είναι αυτή στην οποία πετύχαμε την ελάχιστη τιμή στη συνάρτηση απώλειας, η οποία στη συγκεκριμένη τροποποίηση είναι η 379η.

Η εικόνα που παρατηρούμε στις μετρικές accuracy είναι αρκετά ενδιαφέρουσα. Σε ότι αφορά το training accuracy παρατηρούμε μια διαρκή αύξηση με μειούμενο ρυθμό, όπως θα περιμέναμε βλέποντας το διάγραμμα της loss function καθ' όλη τη διάρκεια εκπαίδευσης εκτός από την τελευταία εποχή. Το validation accuracy παρουσιάζει πολύ μεγαλύτερο ενδιαφέρον. Στις πρώτες 180 εποχές υπάρχει μικρή διακύμανση γύρω από μία κεντρική τιμή στο 84%. Στη συνέχεια παρατηρούμε μια σχετικά απότομη αύξηση στη διακύμανση αλλά και πάλι γύρω από την ίδια τιμή, με τις αντίστοιχες μετρικές BAC και F1 όπως έχουμε παρατηρήσει και στα προηγούμενα διαγράμματα να έχουν ακόμα μεγαλύτερη διακύμανση. Από την εποχή 330 και έπειτα παρατηρούμε ότι το δίκτυο πλέον έχει βελτιώσει το validation accuracy.



Εικόνα 6-16. Accuracy και άλλες μετρικές για το δίκτυο VGG16#8

Ένας λόγος που πιθανόν να εξηγείται το μικρότερο variance στην εκπαίδευση σε σχέση με την τροποποίηση VGG16#7 είναι ότι πλέον μπορούμε να αλλάξουμε βάρη μόνο σε ένα επίπεδο, οπότε μπορούμε να αλλάξουμε μόνο ένα επίπεδο μη γραμμικότητας, που οδηγεί σε μικρότερες δυνατές μεταβολές σε κάθε εποχή εκπαίδευσης. Όπως και στην αρχιτεκτονική VGG16#7 θεωρούμε τελικό επίπεδο εκπαίδευσης το επίπεδο που σημειώθηκε η μικρότερη τιμή στη συνάρτηση απώλειας, η οποία για την παρούσα αρχιτεκτονική είναι η 379ή.

Πίνακας 28. Μετρικές αποτελεσμάτων για την περίπτωση VGG#8 , μετά την 379ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0,8842	0,5485	0,6770	0,6952	0,8494

Συγκρίνοντας το τελικό μοντέλο που λάβαμε εκπαιδύοντας μόνο το τελευταίο συνελκτικό επίπεδο παρατηρούμε από τον παραπάνω πίνακα, ότι τα αποτελέσματα που πήραμε είναι παρόμοια με το VGG16#7. Είναι 1% χειρότερα στο precision, και στα υπόλοιπα απέχουν λιγότερο από 1%. Το παραπάνω είναι μη αναμενόμενο μιας και περιμέναμε ότι το γεγονός ότι δεν μπορούμε να εκπαιδεύσουμε τα χαρακτηριστικά ακόμα

ενός επιπέδου θα ανάγκαζε το μοντέλο να παράγει χειρότερα αποτελέσματα κάτι που δεν επιβεβαιώνεται πειραματικά. Το γεγονός ότι τα αποτελέσματα είναι αντίστοιχα με το δίκτυο VGG16#7 σημαίνει επίσης ότι και το παρόν μοντέλο είναι χειρότερο από το αρχικό δίκτυο VGG16#1 με το οποίο αρχικοποιήθηκε το παρόν δίκτυο! Συνεπώς η φιλοσοφία της εκπαίδευσης μόνο των τελευταίων συνελκτικών επιπέδων δεν απέδωσε θετικά αποτελέσματα σε καμία από τις 2 υλοποιήσεις που δοκιμάστηκαν με την αρχιτεκτονική VGG16 ως αφετηρία. Ακολουθεί ο πίνακας σύγκυσης για την τροποποίηση VGG16#8.

Πίνακας 29. Αποτελέσματα της τροποποίησης VGG16#8 σε μορφή πίνακα σύγκυσης, μετά την 379ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	1825	1502
True Negative	239	7995

6.10 Αποτελέσματα για VGG16#9

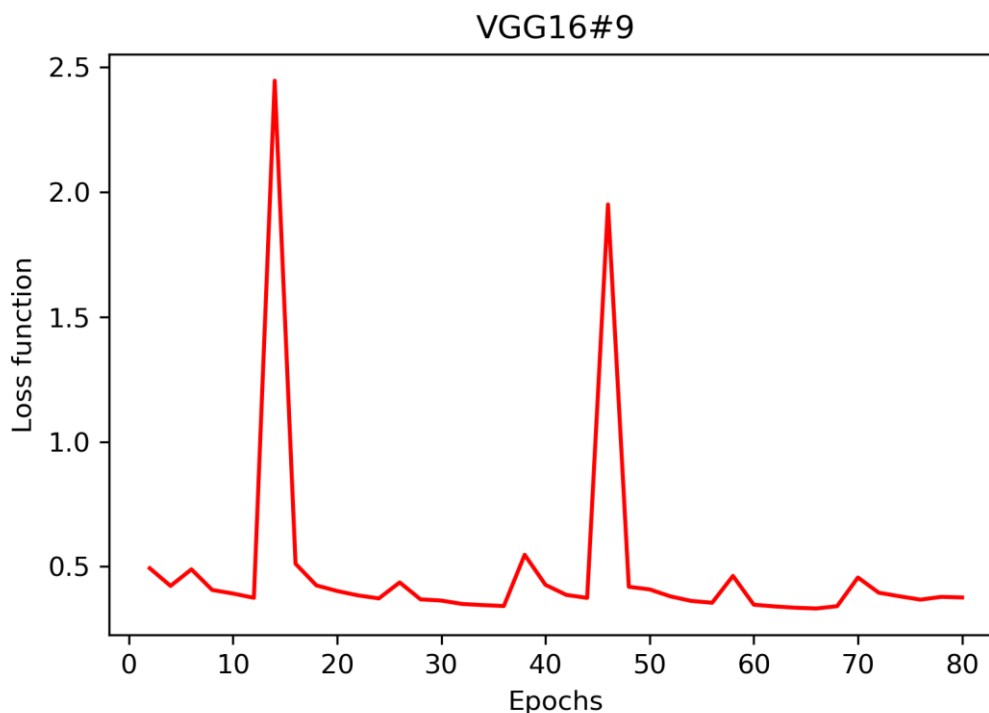
Περιγράφουμε εν συντομία την συγκεκριμένη τροποποίηση. Σκοπός της παρούσας τροποποίησης είναι να εξετάσουμε τη συμπεριφορά της εκπαίδευσης του δικτύου, εάν το αρχικοποιήσουμε με τα βάρη του VGG16 στο συνελκτικό μέρος και τυχαία βάρη στο πλήρες συνδεδεμένο επίπεδο. Σε αυτή την τροποποίηση όμως, σε αντίθεση με την VGG16#1 που εκπαίδευσαμε μόνο το πλήρως συνδεδεμένο μέρος του δικτύου, αφήνουμε όλο το δίκτυο να εκπαιδευτεί. Συνεπώς σκοπός μας είναι να δούμε εάν υπάρχει όφελος στην εκπαίδευση του συνολικού δικτύου σε σχέση με την εκπαίδευση σε βήματα συνελκτικού και πλήρως συνδεδεμένου μέρους του δικτύου.

Πίνακας 30. Παράμετροι εκτέλεσης για την τροποποίηση VGG16#9

Αλγόριθμος ελαχιστοποίησης	Stochastic Gradient Descent
Μέγεθος δεσμίδας εκπαίδευσης	150
Πλήθος εποχών	380
Ταχύτητα εκμάθησης (learning rate)	0.0002

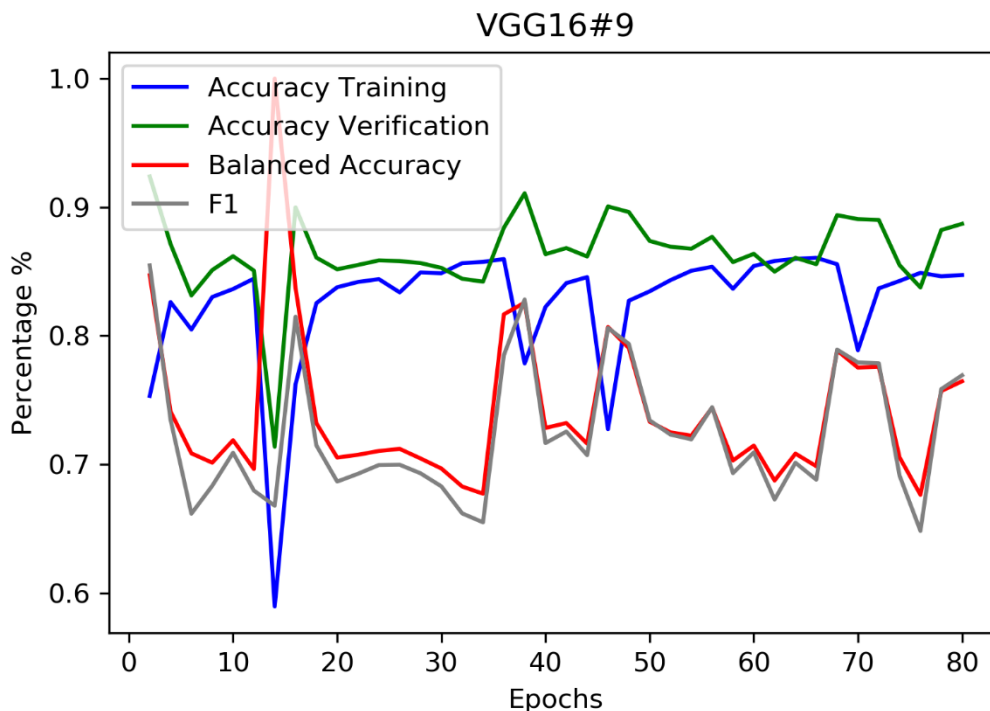
Χρήση ορμής εκμάθησης (momentum)	0.3
Χρήση κανόνα Nesterov	Ναι
Ρυθμός μείωσης ταχύτητας εκμάθησης (decay)	0.000001
Loss Function	Binary Cross Entropy

Ακολουθεί το διάγραμμα για την συνάρτηση απώλειας, στην οποία παρατηρούμε ότι αρχίζει από χαμηλές σχετικά τιμές, το οποίο είναι μη αναμενόμενο λόγω του ότι τα πλήρως συνδεδεμένα επίπεδα είναι αρχικοποιημένα με τυχαίες τιμές. Παρατηρούμε επίσης ότι ούτε στα αρχικά επίπεδα αλλά και ούτε στη συνέχεια η συνάρτηση δείχνει την μονότονη φύση που θα ήταν επιθυμητό. Υπάρχουν δύο χαρακτηριστικές αιχμές της συνάρτησης στις εποχές 14 και 46, που εκτοξεύουν τη συνάρτηση σε τετραπλάσιες τιμές από τις «γειτονικές» εποχές εκπαίδευσης. Το παραπάνω μαρτυρά ότι η εκπαίδευση όλου του δικτύου μονομιάς ενδέχεται να καθυστερεί αρκετά και μπορεί και να μη βρει τοπικό ελάχιστο. Επίσης άξιο προσοχής είναι ότι στην δεύτερη εποχή εκπαίδευσης έχει τιμή 0,49 ενώ στη χαμηλότερη τιμή της φτάνει στην 66ή εποχή με τιμή 0,33. Το παραπάνω μας δείχνει ότι είχαμε καλή αρχικοποίηση στο δίκτυο ή δεν κατάφερε να εκπαιδευτεί επαρκώς.



Εικόνα 6-17. Απεικόνιση της συνάρτησης απώλειας (loss function) για το δίκτυο VGG16#9

Σε ότι αφορά τις μετρικές που έχουμε αναλύσει παρατηρούμε ότι το training accuracy αρχίζει από 75.33% και πετυχαίνει την υψηλότερη τιμή του στην εποχή 66 με τιμή 86.07%, το οποίο είναι αρκετά χαμηλό συγκρίνοντάς το με προηγούμενα μοντέλα που εξετάστηκαν. Σε ότι αφορά το validation accuracy, βλέπουμε ότι αρχίζει από πολύ υψηλή τιμή η οποία όσο εκπαιδεύεται μειώνεται μέχρι την 14^η εποχή και στη συνέχεια ξανανεβαίνει όμως δίχως να ξαναφτάσει τα αρχικά επίπεδα. Η εκπαίδευση σύμφωνα με το γράφημα είναι αρκετά ασταθής, με το δίκτυο να μη βελτιώνει αισθητά τις επιδόσεις με την πάροδο των εποχών. Οι μετρικές BAC και F1, είναι χαμηλότερα από το validation accuracy όπως έχουμε παρατηρήσει σε όλες τις παραστάσεις. Ενδιαφέρον παρουσιάζει η μεγάλη απόκλιση που εμφανίζουν στην 6^η εποχή που ενώ σε όλο το γράφημα κινούνται παράλληλα, στην 6^η εποχή η μετρική F1 μειώνεται αισθητά περισσότερο από την BAC. Ως εκπαιδευμένο δίκτυο θεωρήθηκε το μοντέλο μετά την 66^η εποχή εκπαίδευσης, το οποίο παρουσίασε και τη μικρότερη τιμή στη συνάρτηση απώλειας.



Εικόνα 6-18. Accuracy και άλλες μετρικές για το δίκτυο VGG16#9

Πίνακας 31. Μετρικές αποτελεσμάτων για την περίπτωση VGG#9 , μετά την 66ή εποχή εκπαίδευσης

Precision	Recall	F1	BAC	Accuracy
0,9104	0,5531	0,6881	0,6986	0,8557

Σε ότι αφορά τις επιδόσεις του δικτύου μετά την εκπαίδευση, παρατηρούμε ότι η accuracy είναι κατά 2.5% χειρότερη από το VGG16#1, δίκτυο το οποίο εκπαιδευσε μόνο το πλήρως συνδεδεμένο μέρος και 4% χειρότερη από το VGG16#2 το οποίο εκπαιδεύτηκε και στο συνελκτικό κομμάτι του δικτύου. Το παραπάνω μας κάνει να πιστεύουμε, πλέον με απτά επιχειρήματα ότι η εκπαίδευση του πλήρως συνδεδεμένου δικτύου πρέπει να προηγείται της εκπαίδευσης του συνελκτικού μέρους του δικτύου. Ένα επιπρόσθετο στοιχείο είναι ο χρόνος εκπαίδευσης. Στην περίπτωση που όλο το δίκτυο πρέπει να εκπαιδευτεί το κάθε βήμα εκπαίδευσης παίρνει πολλαπλάσιο χρόνο από ότι στην περίπτωση που «παγώνουμε» την εκπαίδευση ενός μέρους του δικτύου.

Ακολουθεί ο πίνακας σύγκρισης για το τελικό μοντέλο που εκπαιδεύτηκε στο σύνολο του βάθους της αρχιτεκτονικής του ταυτόχρονα.

Πίνακας 32. Αποτελέσματα της τροποποίησης VGG16#9 σε μορφή πίνακα σύγκυσης, μετά την 66ή εποχή εκπαίδευσης

	Predicted Positive	Predicted Negative
True Positive	1840	1487
True Negative	181	8053

7. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο σκοπός της παρούσας ερευνητικής εργασίας ήταν εξ' αρχής διττός. Αρχικά θέλαμε να αξιολογήσουμε την επίδοση των Νευρωνικών Δικτύων και πιο συγκεκριμένα των Συνελκτικών Νευρικών Δικτύων CNN στην ταξινόμηση εικόνων πορώδους διηθητικού καρκίνου καθώς επίσης και να συγκρίνουμε τα αποτελέσματα μας με την παρούσα βιβλιογραφία. Ο δεύτερος στόχος της εργασίας ήταν να συγκρίνουμε διάφορες τροποποιήσεις στην αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου, καθώς επίσης και να εξετάσουμε την επίδραση που έχει η αρχικοποίηση του δικτύου στην εκπαίδευση του, καθώς επίσης και τον τρόπο εκπαίδευσης του δικτύου.

Σε αυτή την ενότητα δεν θέλουμε απλά να αναπαράγουμε τα αποτελέσματα που εξηγούνται στην προηγούμενη ενότητα, αλλά θεωρούμε πιο σημαντικό να αναφέρουμε τη γνώση που μπορεί να αποκομιστεί από τα παραπάνω αποτελέσματα.

Αρχικά παρατηρήσαμε ότι σε όλες τις τροποποιήσεις μας όπου το πλήρως συνδεδεμένο επίπεδο του δικτύου εκπαιδεύτηκε χωρίς να τροποποιήσουμε τα αρχικά χαρακτηριστικά του δικτύου (VGG16#1, VGG16#3, VGG16#5) ήταν υποδεέστερα ως προς το accuracy των αντίστοιχων δικτύων όπου εκπαιδεύτηκαν μόνο τα χαρακτηριστικά τους για το συγκεκριμένο πρόβλημα (VGG16#2, VGG16#4, VGG16#6) αντίστοιχα. Αυτό σημαίνει ότι παρόλο που το αρχικό δίκτυο έχει εκπαιδευτεί πάνω σε δεκάδες χιλιάδες εικόνες, το να εκπαιδεύσουμε το τμήμα «παραγωγής χαρακτηριστικών» (συνελκτικό μέρος του δικτύου) για το συγκεκριμένο πρόβλημα βελτιώνει την απόδοση του δικτύου. Κάτι όμως που δεν ήταν αναμενόμενο είναι το ποσοστό της βελτίωσης της ακρίβειας (accuracy). Στις τροποποιήσεις VGG16#2 και VGG16#6 η βελτίωση ήταν σχετικά μικρή, της τάξης του 1% και 2% αντίστοιχα, ενώ για την τροποποίηση VGG16#4 η βελτίωση ήταν σημαντική και της τάξης του 7%. Το παραπάνω γεγονός μας δείχνει ότι το βήμα της εκπαίδευσης των χαρακτηριστικών του δικτύου για το συγκεκριμένο πρόβλημα είναι εξαιρετικά χρήσιμο να ολοκληρώνεται, ειδικά αν αφορά σε θέματα πρόγνωσης υγείας ασθενών όπου και η στοιχειώδης βελτίωση στην ακρίβεια είναι σημαντική.

Ένας επιπλέον παράγοντας που αξίζει να θιχτεί και λόγω των αποτελεσμάτων που πήραμε στην τροποποίηση VGG16#1 είναι η αρχικοποίηση στην εκπαίδευση συνελκτικών δικτύων. Η τυχαιότητα που παρέχεται από την αρχικοποίηση των δικτύων με τυχαίες τιμές συνήθως στο πλήρως συνδεδεμένο επίπεδο που είναι συγκεκριμένο για το πρόβλημα που θέλουμε να λύσουμε είναι ένας παράγοντας που πρέπει να ληφθεί

υπόψιν. Εάν υπάρχει ο χρόνος για την εκπαίδευση ενός δικτύου είτε για ερευνητικούς σκοπούς αλλά κυρίως για τις περιπτώσεις ιατρικές όπου έχουν εφαρμογή σε ζώες ασθενών, είναι επιθυμητό να γίνει η εκπαίδευση για κάποιο πλήθος αρχικοποιήσεων με διαφορετικά βάρη, ώστε να μειωθεί η επίδραση του παράγοντα τύχη στα αποτελέσματα που λαμβάνουμε για την εκπαίδευση του δικτύου.

Επίσης διαπιστώσαμε ότι το βάθος του δικτύου παίζει σημαντικό ρόλο στο αποτέλεσμα που θα λάβουμε από την εκπαίδευση του δικτύου. Όταν το δίκτυο VGG16#2 αποτελούνταν από 5 συνελκτικά επίπεδα, κατάφερε να αγγίξει το 98% ακρίβεια. Αντίστοιχα το δίκτυο VGG16#4 με 4 συνελκτικά επίπεδα άγγιξε το 96% στην ακρίβεια, ενώ για το δίκτυο VGG16#6 είχαμε μόλις 90.3% ακρίβεια. Οπότε το βάθος στα συνελκτικά δίκτυα αποδεικνύεται ότι είναι κρίσιμος παράγοντας, ειδικότερα όσο το πρόβλημα μας αποτελείται από αναγνώριση λεπτομερειών και υψής σε ολόκληρη την εικόνα ή σημαντικού πλήθους κλάσεων. Προφανώς η αύξηση του βάθους του δικτύου και συνάμα του μεγέθους του δικτύου αυξάνει το χρόνο εκπαίδευσης, καθώς και το πλήθος των δειγμάτων εκπαίδευσης που απαιτούνται, οπότε η εύρεση του κατάλληλου βάθους που αντιστοιχεί στο πλήθος των διαθέσιμων εικόνων είναι μία επιπλέον παράμετρος που πρέπει να ληφθεί υπόψιν.

Μια επιπλέον εκδοχή που δοκιμάστηκε στις τροποποιήσεις VGG16#7 και VGG16#8 είναι η εκπαίδευση μόνο ενός ή δύο επιπέδων του συνελκτικού δικτύου. Τα επίπεδα που δοκιμάστηκαν ήταν εκείνα που εξετάζουν τα πλέον «πολύπλοκα» χαρακτηριστικά, που βρίσκονται και σε μεγαλύτερο βάθος στο συνελκτικό μέρος του νευρωνικού δικτύου. Τα αποτελέσματα που λάβαμε μετά την εκπαίδευση μόνο των συγκεκριμένων επιπέδων ήταν απογοητευτικά, καθώς και στις δύο περιπτώσεις ήταν χειρότερα του δικτύου με το οποίο αρχικοποιήθηκαν. Συνεπώς η εκπαίδευση απομονωμένων επιπέδων του συνελκτικού νευρωνικού δικτύου δεν απέδωσε καρπούς στην παρούσα υλοποίηση και δεν προτείνεται ως μέθοδος για εκπαίδευση δικτύων σε αντίστοιχα προβλήματα.

Τέλος προτείνεται σε περιπτώσεις που υπάρχει ένα προ-εκπαιδευμένο δίκτυο που έχει την αρχιτεκτονική που θέλουμε να χρησιμοποιήσουμε και είναι εκπαιδευμένο σε ένα αξιόλογο σύνολο εικόνων είναι πιθανόν να ληφθούν καλύτερα αποτελέσματα εάν εκπαιδευτούν με τη σειρά πρώτα μόνο το πλήρως συνδεδεμένο κομμάτι του δικτύου και στη συνέχεια μόνο το συνελκτικό μέρος του δικτύου με τα βάρη παγωμένα στο πλήρως συνδεδεμένο μέρος του δικτύου.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
haptics	απτική
artificial intelligence	τεχνητή νοημοσύνη
machine learning	μηχανική μάθηση
supervised learning	επιτηρούμενη μάθηση
semi-supervised learning	ημι-επιτηρούμενη μάθηση
unsupervised learning	μη επιτηρούμενη μάθηση
neural networks	νευρωνικά δίκτυα
convolution neural networks	συνελκτικά νευρωνικά δίκτυα
fully connected neural network	πλήρως συνδεδεμένο συνελκτικό δίκτυο
linear regression	γραμμικής παλινδρόμησης
bias-variance dilemma	διλήμματος προκατάληψης διακύμανσης
cost function	συνάρτηση κόστους
minimum square error	ελαχιστοποίηση τετραγώνων σφάλματος
supervised learning	επιβλεπόμενη μάθηση
threshold	κατώφλι
layer	επίπεδο
input layer	επίπεδο εισόδου
output layer	επίπεδο εξόδου
hidden layers	κρυφά επίπεδα
convolutional layer	συνελκτικό επίπεδο
fully layer connected	πλήρως συνδεδεμένο επίπεδο
bias terms	κατώφλια
Activations	ενεργοποιήσεις

hypothesis function	διαχωριστική συνάρτηση
forward propagation	προς τα εμπρός διάδοση του σήματος
Texture	υφής
activation function	συνάρτηση ενεργοποίησης
rectified linear function	βελτιωμένη γραμμική συνάρτηση
rectified linear unit	διορθωμένη γραμμική μονάδα
backpropagation algorithm	αλγόριθμο οπισθοδιάδοσης σφάλματος
batch gradient descent	ελαχιστοποίηση με βάση παράγωγο συνόλου
gradient descent	ελαχιστοποίηση με βάση παράγωγο
non-convex	μη-κυρτή
forward pass	πέρασμα προς εμπρός
batch optimization	βελτιστοποίησης κατά πακέτα
batch methods	μεθόδους υπολογισμού με πακέτα
pixel	εικονοστοιχείο
receptive field	πεδίο αντίληψης
pooling	μηχανισμούς δειγματοληψίας
pooling layer	επίπεδο δειγματοληψίας
transparency	διαφάνεια
kernels	φίλτρα
feature maps	χάρτες χαρακτηριστικών
bias units	κατώφλια
pixels	εικονοστοιχεία
perceptron	είδος αισθητήρα TN
forward pass	πέρασμα προς εμπρός

stride	διασκελισμός
dilation convolutions	διαστελλόμενες συνελίξεις
balanced accuracy	ισοροπημένη ακρίβεια
invasive ductal carcinoma	πορογενές διηθητικό καρκίνωμα
whole slide imaging	λήψη φωτογραφιών πλήρους μεγέθους
overfitting	υπέρ-προσαρμογή
softmax	softmax

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

ΕΚΠΑ	Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
TN	τεχνητή νοημοσύνη
ΝΔ	νευρωνικό δίκτυο
AI	artificial intelligence
ΒΝΔ	βιολογικά νευρωνικά δίκτυα
SGD	stochastic gradient descent
RELU	rectified linear unit
RELF	rectified linear function
CNN	convolutional neural networks
SVM	support vector machines
RGB	red green blue
BGR	blue green red
BAC	balanced accuracy ή
IDC	invasive ductal carcinoma
WSI	Whole Slide Imaging
VGG16	Visual Geometry Group 16
TN	τεχνητή νοημοσύνη

ΠΑΡΑΡΤΗΜΑ Ι

Ο κώδικας για όλες τις τροποποιήσεις που παρουσιάζονται στην παρούσα εργασία βρίσκεται στην παρακάτω διεύθυνση :

<https://github.com/aktorion/Master-Thesis-Code-Convolution-Neural-Networks->

ΑΝΑΦΟΡΕΣ

- [1] J. McCarthy, “Programs with common sense,” *Proc. Symp. Mech. Thought Process.*, pp. 1–15, 1963.
- [2] A. Newel and H. A. Simon, “[5.1] Computer Science as Empirical Inquiry Symbols and Search,” vol. 19, no. 3, 1976.
- [3] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Orlando, FL, USA: Academic Press, Inc., 2008.
- [4] W. Sun, T.-L. B. Tseng, J. Zhang, and W. Qian, “Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data,” *Comput. Med. Imaging Graph.*, vol. 57, pp. 4–9, 2017.
- [5] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [6] D. A. Drachman, “Do we have brain to spare?,” *Neurology*, vol. 64, no. 12, pp. 2004–2005, 2005.
- [7] F. Rosenblatt and C. Nonr-, “Frosenblatt,” vol. 65, no. 6, pp. 1–23, 2005.
- [8] B. G. Farley and W. A. Clark, “Simulation of self-organized systems by digital computer,” *IRE Trans. Inf. Theory*, vol. 136, no. 4, pp. 76–84, 1954.
- [9] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [10] S. Geman, R. Doursat, and E. Bienenstock, “Neural Networks and the BiasVariance Dilemma,” *Neural Comput.*, vol. 4, pp. 1–58, 1992.
- [11] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.
- [12] J. A. Belward, I. W. Turner, and M. Ilić, “On derivative estimation and the solution of least squares problems,” *J. Comput. Appl. Math.*, vol. 222, no. 2, pp. 511–523, 2008.
- [13] K. Chen, X. Yu, and H. Chi, “Combining linear discriminant functions with neural networks for supervised learning,” *Neural Comput. Appl.*, vol. 6, no. 1, pp. 19–41, 1997.
- [14] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991.
- [15] S. Hochreiter and P. Frasconi, “Gradient Flow in Recurrent Nets : the Difficulty of Learning Long-Term Dependencies.”
- [16] L. A. Gatys, A. S. Ecker, and M. Bethge, “Texture Synthesis Using Convolutional Neural Networks,” pp. 1–9, 2015.
- [17] S. Ruder, “An overview of gradient descent optimization algorithms,” pp. 1–14, 2016.
- [18] M. Alber, I. Bello, B. Zoph, P.-J. Kindermans, P. Ramachandran, and Q. Le, “Backprop Evolution,”

- 2018.
- [19] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015.
- [20] A. Bordes, L. Bottou, and P. Gallinari, "SGD-QN: Careful Quasi-Newton Stochastic Gradient Descent," *J. Mach. Learn. Res.*, vol. 10, pp. 1737–1754, Jul. 2009.
- [21] "ΣΥΝΕΛΙΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΣΤΗΝ ΥΠΟΛΟΓΙΣΤΙΚΗ ΟΡΑΣΗ," 2016.
- [22] N. Loizou and P. Richtárik, "Momentum and Stochastic Momentum for Stochastic Gradient, Newton, Proximal Point and Subspace Descent Methods," 2017.
- [23] Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe, "Deep Convolutional Ranking for Multilabel Image Annotation," *CV*, pp. 1–9, 2013.
- [24] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable CNNs."
- [25] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [26] S. R. Tarando, C. Fetita, Y.-W. Kim, H. Cho, and P.-Y. Brillet, "Boosting CNN performance for lung texture classification using connected filtering," 2018, vol. 10575, no., pp. 1057505–1057513.
- [27] C. H. Hsieh and B. R. Barmish, "On Drawdown-Modulated Feedback Control in Stock Trading," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 952–958, 2017.
- [28] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture and art with deep neural networks," *Curr. Opin. Neurobiol.*, vol. 46, pp. 178–186, 2017.
- [29] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.* 25, pp. 1–9, 2012.
- [30] M. Lin, Q. Chen, and S. Yan, "Network In Network," pp. 1–10, 2013.
- [31] C. Szegedy *et al.*, "Going Deeper with Convolutions," *arXiv1409.4842 [cs]*, pp. 1–9, 2014.
- [32] G. Huang, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks."
- [33] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact Bilinear Pooling," no. 2, pp. 317–326, 2015.
- [34] C. Lee, P. W. Gallagher, and Z. Tu, "Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree," *PMLR*, vol. 51, pp. 464–472, 2016.
- [35] D. M. W. D. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," *J. Mach. Learn. Technol.*, vol. 2, no. December, p. 24, 2007.
- [36] Y. Sasaki, "The truth of the F-measure." pp. 1–5, 2007.
- [37] V. Garcia, R. A. Mollineda, and J. S. Sanchez, "Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions," pp. 73–87, 2015.
- [38] A. Cruz-Roa *et al.*, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," vol. 9041, no. 216, p. 904103, 2014.

- [39] G. Litjens *et al.*, “A Survey on Deep Learning in Medical Image Analysis,” *arXiv*, vol. 1702.05747, no. 1995, pp. 1–34, 2017.
- [40] A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *J. Pathol. Inform.*, vol. 7, no. 1, p. 29, 2016.
- [41] A. A. Cruz-Roa, J. E. A. O. A. Madabhushi, and F. A. G. Osorio, “A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection,” *MICCAI*, vol. Lecture No, no. 8150, pp. 403–410, 2013.
- [42] W. Yu *et al.*, “Visualizing and Comparing AlexNet and VGG using Deconvolutional Layers,” *Icml*, vol. 48, pp. 1–7, 2016.
- [43] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in *International Conference on Learning Representations*, 2015.
- [44] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, “Differential data augmentation techniques for medical imaging classification tasks,” in *AMIA Annual Symposium Proceedings*, 2017, vol. 2017, p. 979.
- [45] D. Pedamonti, “Comparison of non-linear activation functions for deep neural networks on MNIST classification task,” no. 3, 2018.
- [46] J. Brandt, “Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks,” *CVPR*, p. 8, 2014.