

Πανεπιστήμιο Πειραιώς
Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης
Μεταπτυχιακό Πρόγραμμα Στην Εφαρμοσμένη
Στατιστική

Μηχανική Μάθηση και Ανάλυση Εικόνων στον Καρκίνο του Εγκεφάλου

Γαλιατσάτου Ανατολή

Διπλωματική Εργασία που υποβλήθηκε στο Τμήμα Στατιστικής και
Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος
των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος
Ειδίκευσης στην Εφαρμοσμένη Στατιστική.

Πειραιάς
Μάρτιος 2024

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν:

Σωτήριος Τασουλής (Επίκουρος καθηγητής)(Επιβλέπων)

Σωτήριος Μπερσίμης (Καθηγητής)

Γεώργιος Τζαβελάς (Αναπληρωτής καθηγητής)

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

University of Piraeus

School of Finance and Statistics



Department Of Statistics and Insurance Science

Postgraduate Program In Applied Statistics

Machine Learning and Image Analysis in Brain Cancer

Galiatsatou Anatoli

MSc Dissertation submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics .

Piraeus, Greece
March 2024

Περίληψη

Όγκος είναι μία ανώμαλη μάζα ιστού, η οποία σχηματίζεται όταν τα κύτταρα αναπτύσσονται και διαιρούνται ανεξέλεγκτα ή όταν δεν πεθαίνουν όταν πρέπει. Περισσότεροι από 150 διαφορετικοί τύποι εγκεφαλικών όγκων έχουν βρεθεί από τους ερευνητές, οι οποίοι διακρίνονται σε καλοήθεις και κακοήθεις(καρκίνος). Τόσο η ομοιότητα μεταξύ των διαφορετικών τύπων εγκεφαλικού όγκου, όσο και η ομοιότητα μεταξύ των καρκινικών και των υγιών κυττάρων που τα περιβάλλουν καθιστούν τη διάγνωση των ασθενών με εγκεφαλικό όγκο ιδιαίτερα δύσκολη. Από την άλλη μεριά τα υψηλά ποσοστά θνησιμότητας των ασθενών με κακοήθεις εγκεφαλικούς όγκους, καθιστούν αναγκαία την ανίχνευση και ταξινόμηση αυτών σε πρώιμο στάδιο. Μία τεχνική ιατρικής απεικόνισης, η μαγνητική τομογραφία MRI βοηθά τους γιατρούς να εντοπίσουν και να αναγνωρίσουν τους εγκεφαλικούς όγκους, ωστόσο λόγω όσων περιγράφηκαν παραπάνω, αυτή είναι μία ιδιαίτερα απαιτητική διαδικασία που απαιτεί χρόνο και εξειδίκευση. Η πρόοδος της μηχανικής και βαθιάς μάθησης, επιτρέπει στους ειδικούς γιατρούς την ευκολότερη και πιο αξιόπιστη διάγνωση. Οι παραδοσιακές μέθοδοι μηχανικής μάθησης, χρησιμοποιούν ορισμένες χειροκίνητες μεθόδους εξαγωγής χαρακτηριστικών για την πραγματοποίηση της ταξινόμησης των εγκεφαλικών όγκων. Αντίθετα, το πλεονέκτημα των μεθόδων βαθιάς μάθησης, όπως αυτές που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία, είναι ότι δεν απαιτούν καμία χειροκίνητη εξαγωγή χαρακτηριστικών. Η εργασία αυτή προτείνει μία σειρά μεθόδων για την ανίχνευση και δυαδική ταξινόμηση: παρουσία και απουσία εγκεφαλικού όγκου, από ένα δημόσιο σύνολο δεδομένων, αποτελούμενο συνολικά από 3762 εικόνες MRI. Προτείνονται διάφορες προσεγγίσεις και τροποποιήσεις των μοντέλων InceptionV3, ResNet50 και MobileNetV2, με τη χρήση της τεχνικής Transfer Learning, και δύο ειδικά κατασκευασμένα συνελκτικά νευρωνικά δίκτυα(CNN) 8 και 10 επιπέδων αντίστοιχα.

Abstract

A tumour is an abnormal mass of tissue, which forms when cells grow and divide uncontrollably or in case they do not die when they should. More than 150 types of brain tumours have been identified by researchers, which are divided into benign and malignant (cancer). Both the similarity between the different types of brain tumour and the similarity between the cancerous and healthy cells surrounding them make the diagnosis of brain tumour patients particularly difficult. On the other hand, the high mortality rates of patients with malignant brain tumours make it necessary to detect and classify them at an early stage. One medical imaging technique, MRI helps doctors to detect and identify brain tumours, however, due to what has been described previously, this is a very demanding procedure that requires time and expertise. Progress in machine and deep learning allows medical specialists to provide easier and more reliable diagnosis. Traditional machine learning methods, use certain manual feature extraction techniques to perform classification of brain tumors. In contrast, the advantage of deep learning methods, such as those used in this thesis, is that they do not require any manual feature extraction. This project proposes a set of methods for detection and binary classification: tumour and no tumour brain images, from a public dataset consisting of a total of 3762 MRI images. Several approaches and modifications of InceptionV3, ResNet50 and MobileNetV2 models are proposed, using Transfer Learning. Additionally two custom made convolutional neural networks (CNNs) consisting of 8 and 10 layers respectively are used.

Περιεχόμενα

1	Εισαγωγή	7
1.1	Σκοπός της Εργασίας	7
1.2	Ο Εγκεφαλικός Όγκος	7
1.3	Επιδημιολογία	8
1.4	Διάγνωση	9
1.5	Η συνεισφορά της Μηχανικής και Βαθιάς Μάθησης στη Διάγνωση των Εγκεφαλικών Όγκων	10
2	Σχετική Βιβλιογραφία	13
3	Νευρωνικά Δίκτυα	23
3.1	Feedforward Νευρωνικά Δίκτυα	24
3.1.1	Single layer Perceptron	24
3.1.2	Multilayer Perceptron	25
3.1.3	Συνελκτικά Νευρωνικά Δίκτυα(Convolutional Neural Networks-CNNs)	26
3.2	Συναρτήσεις Ενεργοποίησης	31
3.2.1	Σιγμοειδής Συνάρτηση Ενεργοποίησης(Sigmoid Activation Function)	31
3.2.2	Συνάρτηση Ενεργοποίησης Υπερβολικής Εφαπτομένης(Tanh Activation Function)	33
3.2.3	Συνάρτηση Ενεργοποίησης Hard Tanh	33
3.2.4	Συνάρτηση Ενεργοποίησης Softmax	34
3.2.5	Συνάρτηση Ενεργοποίησης ReLu(Rectified Linear Unit)	35
3.2.6	Συνάρτηση Ενεργοποίησης Leaky ReLu(Leaky Rectified Linear Unit)	36
3.2.7	Συνάρτηση Ενεργοποίησης Softplus	36
3.3	Συναρτήσεις Απώλειας	37
3.3.1	Binary Cross-Entropy Loss Function-Log Loss	37
3.3.2	Categorical Cross-Entropy Loss Function	38
3.3.3	Sparse Categorical Cross-Entropy	39
4	Transfer Learning	41
4.1	ResNets/ResNet50	42
4.2	InceptionV3	45
4.3	MobileNetV2	51

5	Μεθοδολογία και Αξιολόγηση των Μοντέλων Βαθιάς Μάθησης	55
5.1	Σύνολο Δεδομένων	55
5.2	Εισαγωγή και Προ-επεξεργασία των Δεδομένων	55
5.3	Μέτρα Αξιολόγησης της Απόδοσης	56
5.4	Μεθοδολογία	58
5.5	Αποτελέσματα	62
5.6	Καμπύλες Ακρίβειας κα Απώλειας	64
6	Συμπεράσματα	69
6.1	Προτάσεις μελλοντικής έρευνας και βελτίωσης	70
	Παράρτημα	76

Κατάλογος σχημάτων

3.1	Παράδειγμα συνέλιξης μίας εικόνας διαστάσεων 3x3 με ένα φίλτρο διαστάσεων 2x2 που δίνει σαν αποτέλεσμα έναν feature map διαστάσεων 2x2. Πηγή: Deep Learning: Foundations and Concepts	28
3.2	Απεικόνιση της διαδικασίας Max Pooling, στην οποία σε κάθε μπλοκ μεγέθους 2x2 ενός feature map επιλέγεται η μέγιστη τιμή για την παραγωγή ενός νέου feature map μικρότερων διαστάσεων. Πηγή: Deep Learning: Foundations and Concepts	30
3.3	Η Σιγμοειδής Συνάρτηση Ενεργοποίησης	32
3.4	Η Συνάρτηση Ενεργοποίησης <i>Tanh</i>	33
3.5	Η Συνάρτηση Ενεργοποίησης <i>Hard Tanh</i>	34
3.6	Η Συνάρτηση Ενεργοποίησης <i>ReLU</i>	35
3.7	Η Συνάρτηση Ενεργοποίησης <i>Leaky ReLU</i>	36
3.8	Η Συνάρτηση Ενεργοποίησης <i>Softplus</i>	37
4.1	Residual Learning: ένα residual μπλοκ 2 στρωμάτων Πηγή:Deep Residual Learning for Image Recognition	43
4.2	Μία συνάρτηση υπολοίπων F για το ImageNet. Αριστερά: ένα residual μπλοκ(σε χάρτες χαρακτηριστικών 56x56), για το ResNet34. Δεξιά: ένα bottleneck μπλοκ για τα ResNet50/101/152. Πηγή:Deep Residual Learning for Image Recognition	44
4.3	Αρχιτεκτονικές για το ImageNet. Τα δομικά μπλοκ παρουσιάζονται στις αγκύλες, μαζί με τον αριθμό των stacked μπλοκ. Πηγή:Deep Residual Learning for Image Recognition	45
4.4	Αριστερά: inception module naive μορφή. Δεξιά Inception module με μείωση διαστάσεων. Πηγή:Going Deeper with Convolutions	47
4.5	Το δίκτυο GoogLeNet. Πηγή:Going Deeper with Convolutions	48

4.6	Η διάταξη της αρχιτεκτονικής InceptionV2. Το μέγεθος εξόδου κάθε module είναι το μέγεθος εισόδου του επόμενου. Έχει χρησιμοποιηθεί η τεχνική grid reduction, για τη μείωση του μεγέθους των πλεγμάτων μεταξύ των Inception μπλοκ, όπου είναι εφικτό. Έχει χρησιμοποιηθεί συνέλιξη με προσθήκη μηδενικών(zero- padding), για τη διατήρηση του μεγέθους του πλέγματος. Zero-padding χρησιμοποιήθηκε, επιπλέον, μέσα στα Inception modules που δεν μειώνουν το μέγεθος του πλέγματος. Όλα τα υπόλοιπα στρώματα δεν χρησιμοποιούν zero-padding. Πηγή:Rethinking the Inception Architecture for Computer Vision	49
4.7	Inception module που μειώνει το μέγεθος του πλέγματος, ενώ αυξάνει το μέγεθος των φίλτρων. Στα δεξιά, το διάγραμμα υπό την προοπτική των μεγεθών των πλεγμάτων και όχι των λειτουργιών που εφαρμόζονται. Πηγή:Rethinking the Inception Architecture for Computer Vision	50
4.8	Inception modules όπου κάθε συνέλιξη 5x5, αντικαθίσταται από δύο συνελίξεις 3x3. Πηγή:Rethinking the Inception Architecture for Computer Vision	50
4.9	Inception modules μετά την παραγοντοποίηση(factorization) των nxn συνελίξεων. Στο InceptionV2, έχει επιλεγθεί $n = 7$ για το 17x17 πλέγμα. Πηγή:Rethinking the Inception Architecture for Computer Vision	50
4.10	Δομή ενός residual bottleneck block που μετασχηματίζει από k σε k' κανάλια, με stride s και expansion factor t . Πηγή:MobileNetV2: Inverted Residuals and Linear Bottlenecks	53
4.11	MobileNetV2: Κάθε γραμμή περιγράφει μία ακολουθία ενός ή περισσότερων ίδιων επιπέδων που επαναλαμβάνονται n φορές. Όλα τα επίπεδα, στην ίδια ακολουθία έχουν τον ίδιο αριθμό εξόδων c . Το πρώτο επίπεδο κάθε ακολουθίας έχει stride ίσο με s και όλα τα υπόλοιπα έχουν stride ίσο με 1. Πηγή:MobileNetV2: Inverted Residuals and Linear Bottlenecks	53
5.1	Confusion Matrix.	57
5.2	Το Συνελκτικό Νευρωνικό Δίκτυο 8 Επιπέδων	61
5.3	Το Συνελκτικό Νευρωνικό Δίκτυο 10 Επιπέδων	62
5.4	ResNet50:Καμπύλες Ακρίβειας.	65
5.5	ResNet50:Καμπύλες Απώλειας.	65
5.6	MobileNetV2:Καμπύλες Ακρίβειας.	65
5.7	MobileNetV2:Καμπύλες Απώλειας.	65
5.8	InceptionV3+CNN1: Καμπύλες Ακρίβειας.	66
5.9	InceptionV3+CNN1: Καμπύλες Απώλειας.	66
5.10	CNN 10 layers: Καμπύλες Ακρίβειας.	67
5.11	CNN 10 layers: Καμπύλες Απώλειας.	67

Κεφάλαιο 1

Εισαγωγή

1.1 Σκοπός της Εργασίας

Ο καρκίνος του εγκεφάλου αποτελεί μία σπάνια αλλά θανατηφόρα ασθένεια[11], [26], [31], η διάγνωση της οποίας είναι ιδιαίτερα απαιτητική για τους ειδικούς γιατρούς, ακόμη και όταν χρησιμοποιούνται εξειδικευμένες ιατρικές απεικονιστικές εξετάσεις, όπως η Μαγνητική Τομογραφία. Παρόλα αυτά η διάγνωση των ασθενών με καρκίνο του εγκεφάλου σε πρώιμο στάδιο, βοηθά τον κατάλληλο προγραμματισμό της θεραπείας, βελτιώνοντας το βιοτικό επίπεδο και αυξάνοντας τις πιθανότητες επιβίωσης αυτών. Με βάση τα παραπάνω, η αυτοματοποιημένη ανίχνευση καρκινικών εγκεφαλικών όγκων έχει καταστεί αναγκαία. Τα τελευταία χρόνια οι εφαρμογές τεχνητής νοημοσύνης στην ιατρική αυξάνονται και επηρεάζουν κάθε πτυχή της φροντίδας των ασθενών με καρκινικούς όγκους. Στην παρούσα διπλωματική εργασία, πραγματοποιήθηκε εκτενής μελέτη, με τη χρήση μεθόδων Βαθιάς Μάθησης, για την έγκαιρη και αξιόπιστη ανίχνευση και ταξινόμηση ενός συνόλου εικόνων MRI σε περιπτώσεις απουσίας και παρουσίας εγκεφαλικού όγκου. Σκοπός ήταν η σύγκριση των μεθόδων που χρησιμοποιήθηκαν και η αύξηση της απόδοσής τους, ώστε να καταστούν χρήσιμες σε πραγματικές κλινικές μελέτες.

1.2 Ο Εγκεφαλικός Όγκος

Ο όγκος του εγκεφάλου, γνωστός ως ενδοκρανιακός όγκος, είναι μία ανώμαλη μάζα ιστού στην οποία τα κύτταρα αναπτύσσονται και πολλαπλασιάζονται ανεξέλεγκτα, φαινομενικά χωρίς να ελέγχονται από τους μηχανισμούς που ελέγχουν τα φυσιολογικά κύτταρα. Τα κύτταρα είναι τα βασικά δομικά στοιχεία του σώματος, από τα οποία σχηματίζονται τα όργανα και οι ιστοί. Οι όγκοι μπορεί να είναι είτε καλοήθεις, είτε κακοήθεις(καρκίνος). Οι καλοήθεις όγκοι μπορούν να μεγαλώσουν, αλλά όχι να εξαπλωθούν σε γειτονικούς ιστούς και άλλα όργανα. Οι κακοήθεις όγκοι μπορούν να εξαπλωθούν και να εισβάλλουν σε γειτονικούς ιστούς και άλλα μέρη του σώματος, μέσω του αίματος και των λεμφικών συστημάτων. Οι όγκοι του κεντρικού νευρικού συστήματος και κυρίως του εγκεφάλου αποτελούν ένα ιδιαίτερο θέμα στη νευρολογική ιατρική. Η σημασία τους απορρέει από τη μεγάλη ποικιλία τους, τα πολυάριθμα νευρολογικά συμπτώματα που προκαλούν ανάλογα με το μέγεθός τους, το

σημείο στο οποίο βρίσκονται και τις διεισδυτικές τους ικανότητες, την καταστροφή και τη μετατόπιση των γειτονικών ιστών, την ενδοκράνια υπέρταση που προκαλούν και κατά κύριο λόγο τη θνησιμότητα τους. Είναι γνωστό, ότι υπάρχουν καταγεγραμμένοι περισσότεροι από 150 διαφορετικοί όγκοι του εγκεφάλου, αλλά οι δύο κύριες ομάδες αυτών είναι οι πρωτοπαθείς και οι μεταστατικοί όγκοι. Οι πρωτοπαθείς όγκοι του εγκεφάλου περιλαμβάνουν τους όγκους που προέρχονται από τους ιστούς του εγκεφάλου ή από το άμεσο γειτονικό περιβάλλον του εγκεφάλου. Οι πρωτοπαθείς όγκοι του εγκεφάλου κατηγοριοποιούνται ως γλοιακοί(δηλαδή αποτελούμενοι από γλοιακά κύτταρα) ή μη γλοιακοί(δηλαδή εκείνοι που αναπτύχθηκαν πάνω ή μέσα στις δομές του εγκεφάλου, συμπεριλαμβανομένων των νεύρων, των αιμοφόρων αγγείων και των αδένων) και ως καλοήθεις ή κακοήθεις. Οι πρωτοπαθείς καρκίνοι του εγκεφάλου δεν εξαπλώνονται έξω από τον εγκέφαλο και τον νωτιαίο μυελό. Οι μεταστατικοί όγκοι του εγκεφάλου, περιλαμβάνουν τους όγκους που εμφανίζονται σε κάποιο άλλο σημείο τους σώματος(για παράδειγμα στον μαστό ή στους πνεύμονες) και μεταναστεύουν στον εγκέφαλο μέσω της κυκλοφορίας του αίματος. Οι μεταστατικοί όγκοι θεωρούνται καρκίνοι και είναι κακοήθεις. Μία μετάσταση διατηρεί το όνομα του αρχικού καρκίνου. Για παράδειγμα, ο καρκίνος του εντέρου που εξαπλώθηκε στον εγκέφαλο εξακολουθεί να ονομάζεται μεταστατικός καρκίνος του εντέρου, παρόλο που ο ασθενής μπορεί να έχει συμπτώματα επειδή ο καρκίνος βρίσκεται στον εγκέφαλο. Παρουσιάζονται κάποιες γενικεύσεις που βοηθούν στην καλύτερη κατανόηση της ασθένειας αυτής. Καταρχάς, παρά το γεγονός πως αρκετοί τύποι όγκων μπορούν να εμφανιστούν στο κεντρικό νευρικό σύστημα, είναι γνωστό ότι συγκεκριμένοι είναι πολύ πιο συχνοί από άλλους. Ο καρκίνος του μαστού, του πνεύμονα, του γαστρεντερικού και το μελάνωμα ή λέμφωμα εμφανίζουν συχνά τάση για μετάσταση στους νευρικούς ιστούς. Δεύτερον, κάποιοι πρωταρχικοί ενδοκρανιακοί όγκοι και όγκοι του νωτιαίου μυελού τείνουν να αναπτύσσονται σε συγκεκριμένα τμήματα της ενδοκρανιακής κοιλότητας, προκαλώντας έτσι χαρακτηριστικά νευρολογικά σύνδρομα. Τέτοιοι όγκοι είναι το κραιοφαρυγγίωμα, το μηνιγγίωμα και το σβάννωμα. Τρίτον, η παρουσία μίας κατάστασης ανοσοκαταστολής όπως το AIDS ή η χημειοθεραπεία για καρκίνο, ειδικές κληρονομικές διαταραχές, όπως η νευρονωμάτωση, και η έκθεση σε ακτινοβολία προκαλούν προδιάθεση για την ανάπτυξη όγκων στο κεντρικό νευρικό σύστημα. Τέταρτον οι ρυθμοί ανάπτυξης και η διεισδυτικότητα των όγκων ποικίλουν, για παράδειγμα κάποιοι από αυτούς, όπως το γλοιοβλάστωμα είναι εξαιρετικά επιθετικοί και ταχέως αναπτυσσόμενοι, ενώ άλλοι όπως το μηνιγγίωμα είναι καλοήθεις, βραδέως αναπτυσσόμενοι και λιγότερο καταστροφικοί. Αυτές οι διαφορετικές ιδιότητες των καλοηθών και κακοηθών νεοπλασιών, έχουν ουσιαστικές κλινικές επιπτώσεις, που καθορίζουν τον ρυθμό εξέλιξης της κλινικής τους εικόνας, την θεραπευτική προσέγγιση και αποτελεσματικότητα αυτής και συνεπώς την πρόγνωση.

1.3 Επιδημιολογία

Η συχνότητα εμφάνισης των όγκων του εγκεφάλου συνολικά, δηλαδή για καλοήθεις και κακοήθεις όγκους μαζί, είναι 18,71 ανά 100.000 ανθρωποέτη. Πιο συγκεκριμένα, 11,52 ανά 100.000 ανθρωποέτη για καλοήθεις όγκους και 7,19 ανά 100.000 ανθρωποέτη για κακοήθεις όγκους. Οι κακοήθεις όγκοι του εγκεφάλου αντιπροσω-

πεύουν μόνο το 1% έως 2% όλων των καρκίνων στους ενήλικες και επομένως θεωρούνται σπάνιοι. Μεταξύ όλων των νόσων που συναντώνται στους ενήλικες, ο όγκος υπερβαίνει σε θνησιμότητα μόνο το εγκεφαλικό επεισόδιο. Επιπλέον η τρέχουσα υψηλή συχνότητα νεοπλασμάτων του εγκεφάλου που παρατηρείται σε σύγκριση με παλαιότερα είναι κατά πάσα πιθανότητα αποτέλεσμα της διαθεσιμότητας και χρήσης των απεικονιστικών μέσων (αξονική τομογραφία εγκεφάλου, μαγνητική τομογραφία εγκεφάλου κλπ). Έχει διαπιστωθεί ότι μεταστατικοί όγκοι είναι περισσότεροι από τους πρωταρχικούς στον εγκέφαλο. Ο πιο συχνά εμφανιζόμενος ενδοκρανιακός όγκος είναι το μηνιγγίωμα, το οποίο όμως αντιπροσωπεύει ένα πολύ μικρό ποσοστό των θανάτων. Ακολουθούν σε συχνότητα οι όγκοι της υπόφυσης, οι οποίοι παρά το γεγονός πως είναι ενδοκρανιακοί αποτελούν μία ειδική περίπτωση που στην πραγματικότητα δεν ανήκει στους όγκους του εγκεφάλου. Οι πιο σοβαροί πρωταρχικοί καρκίνοι του εγκεφάλου προέρχονται από τα γλοιακά κύτταρα. Αναφορικά με τα παιδιά και τους εφήβους, οι όγκοι του εγκεφάλου είναι οι πιο συνηθισμένοι συμπαγείς όγκοι που τους επηρεάζουν. Οι όγκοι του εγκεφάλου στα παιδιά συχνά σχηματίζονται σε διαφορετικά σημεία του εγκεφάλου από ότι στους ενήλικες και μπορεί να έχουν διαφορετικές θεραπείες και προοπτικές. Τα παιδιά είναι πιο πιθανό να αναπτύξουν όγκους στο κάτω μέρος του εγκεφάλου, το οποίο περιλαμβάνει τις περιοχές που ελέγχουν τις λειτουργίες ύπνου/αφύπνισης, την κίνηση και τον συντονισμό. Ορισμένοι τύποι όγκων είναι πιο συνηθισμένοι στα παιδιά από ότι στους ενήλικες. Οι πιο συνηθισμένοι τύποι παιδιατρικών όγκων είναι τα μυελοβλαστώματα, τα χαμηλού βαθμού αστροκυτώματα, τα επενδυώματα, τα κρανιοφαρυγγιώματα και τα γλοιώματα του εγκεφαλικού στελέχους. Λόγω της θέσης τους ορισμένοι παιδιατρικοί όγκοι, καθώς και οι απαιτούμενες θεραπείες τους, μπορούν να προκαλέσουν σημαντική μακροχρόνια βλάβη στην πνευματική και νοητική λειτουργία. Ο καρκίνος του εγκεφάλου κατά την παιδική ηλικία είναι σπάνια νόσος και συνιστά <1% όλων των περιπτώσεων καρκίνου. Η ετήσια συχνότητα για παιδιά από τη γέννηση μέχρι την ηλικία των 14 χρόνων είναι 14 ανά 100.000 και για άτομα από 15-19 χρόνων είναι 20 ανά 100.000.

1.4 Διάγνωση

Στους περισσότερους ασθενείς η διάγνωση του εγκεφαλικού όγκου γίνεται μετά την εμφάνιση των συμπτωμάτων. Συχνά η διάγνωση ενός εγκεφαλικού όγκου γίνεται ,αρχικά, από έναν ενδοκρινολόγο ή έναν νευρολόγο, που θα πραγματοποιήσει φυσική εξέταση. Ο όρος φυσική εξέταση, ουσιαστικά αναφέρεται, στην νευρολογική εξέταση που πραγματοποιεί ο γιατρός με σκοπό να αξιολογήσει το νευρικό σύστημα ,ελέγχοντας πως λειτουργούν διάφορα μέρη του εγκεφάλου και του σώματος, συμπεριλαμβανομένου του λόγου, της ακοής, της όρασης και της κίνησης. Μερικά παραδείγματα αποτελούν, η δοκιμή δύναμης των μυών των ποδιών και των χεριών, ασκήσεις για τον εγκέφαλο, όπως απλές αριθμητικές πράξεις ή τεστ μνήμης, δοκιμή των αισθήσεων, όπως η αντίδραση του ασθενή σε ένα ελαφρύ άγγιγμα ή τσίμπημα κ.α. Είναι πιθανό, ο ασθενής να υποβληθεί σε αιματολογικές εξετάσεις με σκοπό να ελεγχθεί εάν ο όγκος παράγει ασυνήθιστα επίπεδα ορμονών. Ο γιατρός, εκτός από τα παραπάνω, θα ζητήσει και την διενέργεια κάποιου τεστ για να προχωρήσει στην διάγνωση του εγκεφαλικού όγκου. Οι απεικονιστικές εξετάσεις

μπορούν να βοηθήσουν τους γιατρούς να διαπιστώσουν εάν ο όγκος στον εγκέφαλο είναι πρωτοπαθής ή είναι καρκίνος που έχει εξαπλωθεί στον εγκέφαλο από κάποιο άλλο σημείο του σώματος. Οι απεικονιστικές εξετάσεις, αποτελούν εικόνες από το εσωτερικό του σώματος. Οι παρακάτω παράγοντες μπορούν να ληφθούν υπόψη κατά την διαδικασία επιλογής διαγνωστικής εξέτασης:

- Ο τύπος του υποπτευόμενου όγκου
- Τα συμπτώματα του ασθενή
- Η ηλικία και η γενικότερη κατάσταση της υγείας του ασθενή
- Τα αποτελέσματα προηγούμενων ιατρικών εξετάσεων

Για τη διάγνωση ενός εγκεφαλικού όγκου, υπάρχουν πολλές και διαφορετικές εξετάσεις που μπορούν να προταθούν σε έναν ασθενή, όπως η Μαγνητική τομογραφία (MRI), η Ακτινογραφία CT, η Τομογραφία PET, η Αγγειογραφία εγκεφάλου, η Οσφυονωτιαία παρακέντηση, το Μυελόγραμμα, η Εξέταση βιοδεικτών του όγκου, Νευρολογικές εξετάσεις και εξετάσεις ακοής και όρασης, η Νευρογνωστική αξιολόγηση, το Ηλεκτροεγκεφαλογράφημα κλπ. Οι διάφορες μορφές απεικόνισης, γίνονται ολοένα και πιο δημοφιλείς και προτιμώνται από του ειδικούς γιατρούς, καθώς είναι πιο ακριβείς και εισάγουν λιγότερους κινδύνους για του ασθενείς. Η πιο διαδεδομένη μέθοδος λήψης ιατρικών δεδομένων απεικόνισης είναι η μαγνητική τομογραφία MRI, καθώς παρέχει λεπτομερείς εικόνες του εσωτερικού του σώματος, υψηλής ανάλυσης χωρίς καμία ακτινοβολία. Η μαγνητική τομογραφία χρησιμοποιείται τόσο για την ανίχνευση των όγκων, όσο και για την μέτρηση του μεγέθους αυτών. Για να πραγματοποιηθεί η μαγνητική τομογραφία, χρησιμοποιείται μία ειδική χρωστική ουσία που βοηθά στην παραγωγή μιας πιο καθαρής εικόνας. Αυτή η χρωστική ουσία χορηγείται στον ασθενή ενδοφλέβια. Σημειώνεται πως αν κάποια από τις παραπάνω εξετάσεις υποδείξει την παρουσία όγκου, ο πιο συνηθισμένος τρόπος για να προσδιοριστεί ο τύπος του εγκεφαλικού όγκου είναι να παρθεί και να εξεταστεί ένα τμήμα ιστού, μετά από βιοψία.

1.5 Η συνεισφορά της Μηχανικής και Βαθιάς Μάθησης στη Διάγνωση των Εγκεφαλικών Όγκων

Η βελτίωση της διαγνωστικής διαδικασίας μέσω της εφαρμογής τεχνικών Μηχανικής και Βαθιάς Μάθησης σηματοδοτεί μία σημαντική πρόοδο στον τομέα της ιατρικής απεικόνισης και ιδιαίτερα στην ανίχνευση και διάγνωση εγκεφαλικών όγκων. Μία σειρά επισμονικών ερευνών έχει συγκρίνει την ικανότητα διάκρισης των διαφορετικών τύπων εγκεφαλικών όγκων των αλγορίθμων Μηχανικής και Βαθιάς Μάθησης, με τις υψηλότερες αποδόσεις, και των ειδικών νευροακτινολόγων, με τα αποτελέσματα να καταδεικνύουν την υπεροχή των αλγορίθμων Μηχανικής και Βαθιάς Μάθησης στην εργασία αυτή [27, 28]. Το βασικό πλεονέκτημα των μεθόδων αυτών έγκειται στην ικανότητά τους να ενισχύουν την ακρίβεια της διάγνωσης, γεγονός κρίσιμο για την πρόιμη ανίχνευση των εγκεφαλικών όγκων αλλά και για τον σχεδιασμό της θεραπείας που θα ακολουθήσει ο ασθενής. Οι μέθοδοι μηχανικής μάθησης

που χρησιμοποιούνται κατά την διαδικασία της διάγνωσης επικεντρώνονται στην εξαγωγή χαρακτηριστικών, με βάση τα οποία θα πραγματοποιηθεί η ταξινόμηση. Η εξαγωγή των χαρακτηριστικών περιλαμβάνει τον εντοπισμό σημαντικών χαρακτηριστικών των εικόνων MRI, με βάση τα οποία μπορεί να γίνει η διάκριση μεταξύ των περιοχών των όγκων και αυτών που αποτελούνται από υγιή εγκεφαλικό ιστό. Τα χαρακτηριστικά αυτά μπορεί να βασίζονται σε διάφορες ιδιότητες των εικόνων, όπως η ένταση (με βάση μία τιμή κατωφλιού), η χωρική οργάνωση (με βάση τη συσταδοποίηση), η μορφή (με βάση το περίγραμμα) ή η υφή. Με την ολοκλήρωση της διαδικασίας εξαγωγής χαρακτηριστικών, ακολουθεί η διαδικασία επιλογής χαρακτηριστικών που επικεντρώνεται στον εντοπισμό των πιο σχετικών από αυτά, που θα συμβάλλουν σημαντικά στην ακρίβεια της ταξινόμησης που θα ακολουθήσει. Το βήμα αυτό, είναι κρίσιμο καθώς διασφαλίζει ότι το μοντέλο θα επικεντρωθεί στις πτυχές των δεδομένων που παρέχουν την περισσότερη πληροφορία, ενισχύοντας έτσι την προβλεπτική του απόδοση. Ένας σημαντικός περιορισμός των παραδοσιακών προσεγγίσεων Μηχανικής Μάθησης, έγκειται στην πιθανή απώλεια πολύτιμης πληροφορίας κατά τη διαδικασία εξαγωγής και επιλογής χαρακτηριστικών. Αυτό συμβαίνει επειδή αυτές οι μέθοδοι βασίζονται σε προκαθορισμένα κριτήρια για τον εντοπισμό χαρακτηριστικών, τα οποία μπορεί να παραβλέψουν πολύπλοκα μοτίβα ή μικρές διαφοροποιήσεις που παρουσιάζονται στις αρχικές εικόνες. Με άλλα λόγια, ακόμη και αν ένας όγκος παρουσιάζει κάποια μοναδικά χαρακτηριστικά, αν αυτά δεν έχουν προηγουμένως καθοριστεί ως σημαντικά για την ανάλυση, τότε ο αλγόριθμος μπορεί να τα αγνοήσει ή να μην τα ερμηνεύσει σωστά. Οι περιορισμοί που τίθενται από τους αλγορίθμους Μηχανικής Μάθησης αντιμετωπίζονται από τους αλγορίθμους Βαθιάς Μάθησης και ιδιαίτερα από τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs), τα οποία εργάζονται απευθείας με τις ακατέργαστες εικόνες που δέχονται ως είσοδο. Τα Συνελκτικά Νευρωνικά Δίκτυα μαθαίνουν αυτόματα να αναγνωρίζουν τα σημαντικά χαρακτηριστικά μέσω των πολλαπλών επιπέδων τους, εξαλείφοντας την ανάγκη για χειροκίνητη εξαγωγή και επιλογή χαρακτηριστικών. Η ικανότητα αυτή, επιτρέπει στα Συνελκτικά Νευρωνικά Δίκτυα να ανιχνεύουν περίπλοκα μοτίβα στα ακατέργαστα δεδομένα εισόδου, οδηγώντας σε βελτιωμένη απόδοση των μοντέλων. Τα Συνελκτικά Νευρωνικά Δίκτυα έχουν τη δυνατότητα να προσαρμόζονται και να βελτιώνονται συνεχώς με την εκπαίδευση σε νέα δεδομένα. Γίνονται έτσι πιο αποδοτικά στη διάκριση μεταξύ των διαφορετικών τύπων όγκων και στην αναγνώριση αυτών σε πρώιμο στάδιο, ακόμη και όταν δεν είναι εμφανείς στο ανθρώπινο μάτι. Παρόλα αυτά τα μοντέλα Βαθιάς Μάθησης, και ειδικά τα CNN, απαιτούν μεγάλες ποσότητες δεδομένων για την αποτελεσματική εκπαίδευση τους, γεγονός που μπορεί να αποτελέσει σημαντική πρόκληση στο ιατρικό τομέα λόγω ζητημάτων ιδιωτικότητας και διαθεσιμότητας. Τη λύση στο πρόβλημα αυτό δίνουν τα προ εκπαιδευμένα συνελκτικά νευρωνικά δίκτυα. Τα μοντέλα αυτά έχουν εκπαιδευτεί σε μεγάλες βάσεις δεδομένων εικόνας και μπορούν να προσαρμοστούν και να αποδώσουν καλά, στο εκάστοτε πρόβλημα χρησιμοποιώντας σχετικά μικρά σύνολα δεδομένων. Η προσέγγιση αυτή επιτρέπει την αξιοποίηση της ικανότητας αναγνώρισης μοτίβων των συνελκτικών νευρωνικών δικτύων, χωρίς την απαίτηση τεράστιων ποσοτήτων νέων δεδομένων, κάνοντας την τεχνολογία πιο προσιτή και εφαρμόσιμη στον ιατρικό τομέα.

Κεφάλαιο 2

Σχετική Βιβλιογραφία

1η Μελέτη

Στην παρούσα μελέτη, οι ερευνητές πρότειναν μία αποτελεσματική λύση για την ταξινόμηση των εγκεφαλικών όγκων από εικόνες MRI, χρησιμοποιώντας μία προεκπαιδευμένη αρχιτεκτονική συνελκτικών νευρωνικών δικτύων. Στόχος ήταν η επίτευξη υψηλής ακρίβειας, διατηρώντας ταυτόχρονα την πολυπλοκότητα χαμηλή. Για την πραγματοποίηση της μελέτης, συγκεντρώθηκε μία εκτενής συλλογή 3303 εικόνων εγκεφάλου από MRI τομογράφους. Οι 3253 εικόνες συλλέχθηκαν από δύο διαδικτυακές βάσεις δεδομένων και οι υπόλοιπες 50 αποκτήθηκαν από MRI τομογραφίες που πραγματοποιήθηκαν σε δύο νοσοκομεία στο Μπαγκλαντές. Το σύνολο δεδομένων, αποτελείται από 1685 εικόνες MRI ατόμων που διαγνώστηκαν με καρκίνο στον εγκέφαλο και 1618 εικόνες MRI υγιών ατόμων. Επιπλέον, οι ερευνητές πραγματοποίησαν image augmentation, υποβάλλοντας τις αρχικές εικόνες σε μετασχηματισμούς όπως κάθετες, οριζόντιες και κατά ορισμένη γωνία μετατοπίσεις. Για κάθε, αρχική, εισερχόμενη στο δίκτυο εικόνα δημιούργησαν εννέα επιπλέον εικόνες, πιο συγκεκριμένα 15165 εικόνες από ασθενείς με όγκο στον εγκέφαλο και 14562 εικόνες από υγιείς ασθενείς. Ο συνδυασμός όλων των συνόλων που αναφέρθηκαν, οδήγησε στο τελικό σύνολο δεδομένων, που για την διεξαγωγή των πειραμάτων, χωρίστηκε σε τρία υποσύνολα, το σύνολο εκπαίδευσης (80% των εικόνων χρησιμοποιήθηκε κατά την εκπαίδευση), το σύνολο επικύρωσης (10% των εικόνων χρησιμοποιήθηκε κατά την επικύρωση) και το σύνολο δοκιμής (10% των εικόνων χρησιμοποιήθηκε κατά τη δοκιμή). Όλες οι εικόνες υπέστησαν μείωση μεγέθους, στα 224x224 pixels και μετατράπηκαν σε αριθμητικές αναπαραστάσεις, καθώς τα μοντέλα μηχανικής και βαθιάς μάθησης λειτουργούν με αριθμητικές τιμές. Η προτεινόμενη αρχιτεκτονική, βασίστηκε στο Visual Geometry Group 19 (VGG19), ένα προ-εκπαιδευμένο μοντέλο που αποτελείται συνολικά από 19 επίπεδα, εκ των οποίων τα 16 είναι συνελκτικά επίπεδα και τα υπόλοιπα τρία είναι πλήρως συνδεδεμένα επίπεδα. Η αρχιτεκτονική αυτή χρησιμοποιεί 2D συνελκτικά επίπεδα, ακολουθούμενα από max pooling επίπεδα, μετά από κάθε τρία ή τέσσερα συνελκτικά επίπεδα. Οι ερευνητές, προσέθεσαν τρία επιπλέον επίπεδα Dropout και τρία Dense επίπεδα στην έξοδο της αρχιτεκτονικής VGG19, για την ενίσχυση της απόδοσης της, τη μείωση της μαθηματικής πολυπλοκότητας και την αποφυγή εμφάνισης υπερπροσαρμογής. Το προτεινόμενο μοντέλο σημείωσε ακρίβεια εκπαίδευσης

99,62%, ακρίβεια επικύρωσης 96,42% και ακρίβεια δοκιμής 96,42%. Το μοντέλο, στο σύνολο δοκιμής, αξιολογήθηκε επιπλέον με τα εξής μέτρα απόδοσης: Precision 91,49%, Recall 99,44%, F1-score 95,3%, Specificity 97,8%. Το μοντέλο δεν εμφάνισε σημάδια υπερπροσαρμογής. Η προτεινόμενη αρχιτεκτονική σημείωσε υψηλότερη ακρίβεια, σε σύγκριση με πολλά state-of-the-art μοντέλα.

2η Μελέτη

Στην μελέτη που ακολουθεί, οι ερευνητές, με στόχο την πρόβλεψη της κατάστασης του βιοδείκτη MGMT στο γλοιοβλάστωμα, που αποτελεί την πιο συχνή και επιθετική μορφή καρκίνου του εγκεφάλου στους ενήλικες, χρησιμοποίησαν μία σειρά μοντέλων βαθιάς μάθησης. Πιο συγκεκριμένα χρησιμοποιήθηκε η 3D του Vision Transformer (ViT3D), το ResNet50, το Xception και το EfficientNet-B3. Όλα τα μοντέλα που χρησιμοποιήθηκαν, πραγματοποίησαν μία δυαδική ταξινόμηση, με τις δύο κλάσεις να αντιπροσωπεύουν την ανίχνευση και τη μη ανίχνευση του βιοδείκτη MGMT αντίστοιχα. Το σύνολο των δεδομένων που χρησιμοποιήθηκε, έχει συλλεχθεί από ιδρύματα σε όλο τον κόσμο, ως μέρος ενός δεκαετούς έργου για την προώθηση της τεχνητής νοημοσύνης στη διάγνωση και θεραπεία των εγκεφαλικών όγκων, το διαγωνισμό Brain Tumor Segmentation (BraTS). Για την πραγματοποίηση των πειραμάτων, χωρίστηκε σε τρία υποσύνολα: το σύνολο εκπαίδευσης, που αποτελείται από 468 εικόνες mpMRI, το σύνολο επικύρωσης, που αποτελείται από 117 εικόνες mpMRI και το σύνολο δοκιμών, που αποτελείται από 87 εικόνες mpMRI. Ως βήμα επεξεργασίας, για την εκπαίδευση του ViT3D, οι εικόνες που βρίσκονταν σε μορφή DICOM, μετατράπηκαν σε ένα τρισδιάστατο πίνακα Numpy. Έγινε στη συνέχεια μείωση μεγέθους στα 256x256 pixels και εφαρμόστηκε η τεχνική VOI LUT για την ενίσχυση της αντίθεσης των εικόνων. Τέλος εφαρμόστηκαν οι εξής τεχνικές data augmentation, περιστροφή 90 μοιρών δεξιόστροφα, περιστροφή 90 μοιρών αριστερόστροφα και περιστροφή 180 μοιρών. Για την εκπαίδευση των μοντέλων EfficientNet-B3, ResNet50 και Xception έγινε αλλαγή μεγέθους των εικόνων στα 300x300 pixels για το EfficientNet-B3 και στα 512x512 pixels για τα ResNet50 και Xception. Εφαρμόστηκαν οι εξής τεχνικές data augmentation, οριζόντια αναστροφή, τυχαία περιστροφή κατά γωνία 36 μοιρών, τυχαία οριζόντια και κάθετη translation (+- 10% της αρχικής διάστασης). Στην παρούσα μελέτη, όσον αφορά την αρχιτεκτονική ViT-3D, οι ερευνητές εκπαίδευσαν 4 μοντέλα ViT-3D ένα για κάθε τύπο εικόνων mpMRI (T1w, T1WCE, T2w, FLAIR). Συγκρίθηκαν δύο διαφορετικές αρχιτεκτονικές του ViT-3D. Η πρώτη χρησιμοποιεί διάσταση εισόδου 16x16x16 και αποτελείται από δύο μπλοκ κωδικοποιητών Transformer, ο καθένας με 16 attention heads και ποσοστό dropout ίσο με 10% τόσο για τα γενικά όσο και για τα embedding στρώματα. Η δεύτερη αρχιτεκτονική διατηρεί τις ίδιες υπερπαραμέτρους (Δύο μπλοκ κωδικοποιητών Transformer, 16 attention heads και ποσοστά dropout 10%), άλλα υιοθετεί μεγαλύτερη διάσταση εισόδου, συγκεκριμένα 32x32x32. Η προσαρμογή αυτή έγινε για τη διερεύνηση της επίδρασης του αυξημένου χωρικού πλαισίου στην ικανότητα του μοντέλου να ανιχνεύει μοτίβα στις mpMRI εικόνες. Στη συνέχεια εφαρμόστηκε είτε απλός μέσος όρος, είτε stacking ensembles με λογιστική παλινδρόμηση για να ληφθούν οι τελικές πιθανότητες πρόβλεψης. Αναφορικά με τα μοντέλα EfficientNet-B3, ResNet50

και Xception, χρησιμοποιήθηκαν τα προεκπαιδευμένα στο ImageNet με την προσθήκη των εξής στρώματων: ένα στρώμα Global Average Pooling, δύο στρώματα Batch Normalization, δύο στρώματα Dropout και δύο Dense στρώματα. Αυτά τα επιπλέον στρώματα εκπαιδεύτηκαν για 20 εποχές. Για κάθε τύπο εικόνας mpMRI εκπαιδεύτηκε ένα μοντέλο και στο τέλος χρησιμοποιήθηκε η μέθοδος της μέσης τιμής των πιθανοτικών βαθμολογιών κάθε μοντέλου για τον υπολογισμό της τελικής πρόβλεψης. Για την αξιολόγηση των μοντέλων ViT-3D χρησιμοποιήθηκε η καμπύλη AUC και καλύτερη απόδοση είχε το μοντέλο με την τεχνική του απλού μέσου όρου και patch size 32x32x32. Για την αξιολόγηση των μοντέλων EfficientNet-B3, ResNet50 και Xception χρησιμοποιήθηκαν τα εξής μέτρα απόδοσης: καμπύλη AUC, precision, ευαισθησία, ειδικότητα και ακρίβεια. Το EfficientNet-B3 εμφάνισε μέτρια απόδοση, πιο συγκεκριμένα είχε AUC στο σύνολο επικύρωσης 0.37407 και AUC στο σύνολο δοκιμής 0.55817, precision 0.40, ευαισθησία 0.27, ειδικότητα 0.56 και accuracy 0.40. Το ResNet50 παρουσίασε βελτιωμένη απόδοση σε σύγκριση με το EfficientNet-B3, με AUC στο σύνολο επικύρωσης 0.42099 και AUC στο σύνολο δοκιμής 0.58078, precision 0.50, ευαισθησία 0.30, ειδικότητα 0.67 και accuracy 0.47. Τέλος, τα αποτελέσματα έδειξαν ότι το Xception υπερέβη τα προηγούμενα δύο μοντέλα όσον αφορά την ακριβή πρόβλεψη του βιοδείκτη MGMT από δεδομένα mpMRI, με AUC στο σύνολο επικύρωσης 0.63827, AUC στο σύνολο δοκιμής 0.61745, precision 0.72, ευαισθησία 0.43, ειδικότητα 0.814 και accuracy 0.61.

3η Μελέτη

Στη μελέτη που ακολουθεί, οι ερευνητές στην προσπάθεια βελτίωσης της ταξινόμησης των εγκεφαλικών όγκων, χρησιμοποίησαν και συνέκριναν την απόδοση τεσσάρων προεκπαιδευμένων μοντέλων. Οι εγκεφαλικοί όγκοι, που απεικονίζονταν στις εικόνες MRI που χρησιμοποιήθηκαν, κατατάχθηκαν σε τρεις τύπους όγκων: το μηνιγγίωμα, το γλοιώμα και τον όγκο της υπόφυσης. Το σύνολο των δεδομένων που χρησιμοποιήθηκε για τα πειράματα ήταν το Figshare. Κατά την προεπεξεργασία των εικόνων δεν χρησιμοποιήθηκε κάποια τεχνική data augmentation, καθώς σκοπός ήταν να τονιστεί η σημασία ενός ισορροπημένου συνόλου δεδομένων χωρίς τη χρήση περίπλοκων τεχνικών προεπεξεργασίας. Στα δεδομένα που συλλέχθηκαν βρέθηκαν 708 περιπτώσεις μηνιγγιώματος, 1426 περιπτώσεις γλοιώματος και 930 περιπτώσεις όγκων της υπόφυσης. Για την επίτευξη της ισορροπίας μεταξύ των τριών κλάσεων των δεδομένων εφαρμόστηκε η class weight τεχνική, τα βάρη που δόθηκαν σε κάθε κατηγορία ήταν τα εξής: στον «όγκο της υπόφυσης» δόθηκε βάρος 0.716, στο «μηνιγγίωμα» δόθηκε βάρος 1.443 και στο «γλοιώμα» δόθηκε βάρος 1.098. Τα μοντέλα που χρησιμοποιήθηκαν για την ταξινόμηση ήταν το ResNet50, το VGG16, το InceptionV3 και το MobileNetV2. Τροποποιήθηκαν, αλλάζοντας τον αριθμό των νευρώνων στο πλήρως συνδεδεμένο στρώμα ώστε να αντιστοιχούν στις τρεις κλάσεις του συνόλου δεδομένων. Τα μοντέλα αξιολογήθηκαν με βάση τα εξής μέτρα απόδοσης: accuracy, precision, recall και F1-score. Κάθε ένα από τα δίκτυα εκπαιδεύτηκε για 10 εποχές, χρησιμοποιήθηκε ο Adam optimizer και η συνάρτηση απώλειας Sparse Categorical Cross Entropy. Καταγράφηκαν τα αποτελέσματα της ταξινόμησης τόσο πριν, όσο και μετά την εφαρμογή της class weight τεχνικής για την εξισορρόπηση του συνόλου δεδομένων. Πριν την εφαρμογή της εξισορρόπησης

τα αποτελέσματα είχαν ως εξής: Το ResNet εμφάνισε accuracy 94% , το VGG16 95%, το InceptionV3 92% και το MobileNetV2 96%. Τα μοντέλα ResNet50 και VGG16 εμφάνισαν συγκρίσιμα ποσοστά accuracy και αποδείχθηκαν αποτελεσματικά στην διάκριση μεταξύ των γλοιωμάτων και των όγκων της υπόφυσης, ωστόσο το ResNet50 αντιμετώπισε δυσκολίες στην ακριβή ταξινόμηση των μηνιγγιωμάτων. Το InceptionV3 εμφάνισε χαμηλό ποσοστό recall για την ανίχνευση του μηνιγγιώματος και το MobileNetV2 εμφάνισε υψηλή απόδοση σε όλες τις κατηγορίες. Τα αποτελέσματα με την εξισορρόπηση του συνόλου των δεδομένων είχαν ως εξής: Το ResNet εμφάνισε accuracy 99.06% και σημαντική βελτίωση στην απόδοση σε όλες τις κλάσεις σε σύγκριση με τα προηγούμενα αποτελέσματα. Το VGG16 έδωσε accuracy 97% και εμφάνισε μικρές βελτιώσεις των μετρήσεων απόδοσης σε σύγκριση με τα προηγούμενα αποτελέσματα. Το InceptionV3 έδωσε accuracy 96% και σημαντική βελτίωση των τιμών των μέτρων απόδοσης για όλες τις κλάσεις. Τέλος το MobileNetV2 έδωσε accuracy 99% και εξαιρετικές τιμές precision και recall σε όλες τις κλάσεις. Συνοψίζοντας η εφαρμογή της class weight τεχνικής, για την επίτευξη της ισορροπίας μεταξύ των κλάσεων του συνόλου δεδομένων, βοήθησε στην βελτίωση της απόδοσης των μοντέλων και στην ενίσχυση της αξιοπιστίας και της ακρίβειας των προβλέψεων.

4η Μελέτη

Στην εργασία που ακολουθεί, οι ερευνητές χρησιμοποίησαν το EfficientNetB0, το ResNet50, το Xception, το MobileNetV2 και το VGG16 με την τεχνική του Transfer Learning για την ανίχνευση και ταξινόμηση τριών τύπων εγκεφαλικών όγκων: το γλοίωμα, το μηνιγγίωμα και τον όγκο της υπόφυσης. Το σύνολο των δεδομένων που χρησιμοποιήθηκε ήταν από το Kaggle και αποτελείται από 3264 εικόνες. Στο σύνολο αυτό των δεδομένων περιλαμβάνονται επίσης και εικόνες MRI χωρίς εγκεφαλικό όγκο. Αποτελείται από 500 εικόνες MRI χωρίς εγκεφαλικό όγκο, 901 εικόνες MRI με όγκο της υπόφυσης, 937 εικόνες MRI με μηνιγγίωμα και 926 με γλοίωμα. Σε όλες τις εικόνες έγινε αλλαγή των διαστάσεων στα 150x150 pixels. Με δεδομένο ότι ο αριθμός των εικόνων χωρίς όγκο ήταν αρκετά μικρότερος από αυτούς των υπόλοιπων κατηγοριών, αυτός αυξήθηκε ξεχωριστά. Χρησιμοποιήθηκαν διάφορες τεχνικές προ επεξεργασίας των δεδομένων, όπως MR Bias Correction και skull stripping. Για τη μείωση του θορύβου που υπήρχε στις εικόνες MRI χρησιμοποιήθηκαν τα εξής: Gaussian blur filter, BM3D denoised και Total Variation smoothing. Επιπλέον για την αύξηση της απόδοσης των δικτύων και την αποφυγή εμφάνισης υπερπροσαρμογής οι ερευνητές χρησιμοποίησαν τις εξής τεχνικές data augmentation: περιστροφή, οριζόντια αναστροφή, κατακόρυφη αναστροφή, μετατόπιση κατά πλάτος και μετατόπιση κατά ύψος. Τα τελευταία επίπεδα όλων των μοντέλων αφαιρέθηκαν για να προσαρμοστούν στα δεδομένα της εργασίας. Το τελευταίο πλήρως συνδεδεμένο στρώμα αφαιρέθηκε και προστέθηκε ένα πλήρως συνδεδεμένο στρώμα με 4 νευρώνες, κατάλληλο για την ταξινόμηση στις 4 κλάσεις του προβλήματος. Ως συνάρτηση απώλειας χρησιμοποιήθηκε η Cross Categorical Cross Entropy. Στη συνέχεια προστέθηκε επιπλέον ένα Global Average Pooling στρώμα ακολουθούμενο από ένα στρώμα Dropout. Τα μοντέλα εκπαιδεύτηκαν για 12 εποχές, με τον Adam optimizer, ρυθμό μάθησης ίσο με 0.0003 και batch size

32. Κατά την εκπαίδευση χρησιμοποιήθηκαν οι λειτουργίες `callback TensorBoard`, `ModelCheckpoint ReduceLROnPlateau`, για την ταχύτερη επίλυση σφαλμάτων, την ανάπτυξη καλύτερων μοντέλων και την αποφυγή εμφάνισης υπερπροσαρμογής. Τα τροποποιημένα μοντέλα αξιολογήθηκαν με βάση τα εξής μέτρα απόδοσης: `accuracy`, `precision`, `recall` και `F1-score`. Το πιο αποδοτικό μοντέλο, με βάση το `accuracy` ήταν το `EfficientNetB0` που κατέγραψε ποσοστό 97.61%. Τα `ResNet50`, `Xception` και `MobileNetV2`, κατέγραψαν επίσης πολύ υψηλό ποσοστό `accuracy` στο 96%. Το `VGG16` δεν κατάφερε να ανιχνεύσει σωστά τους εγκεφαλικούς όγκους, με `accuracy` 72%, παρά την επανειλημμένη ρύθμιση των παραμέτρων του από τον ερευνητές.

5η Μελέτη

Η παρούσα εργασία παρουσιάζει μία εφαρμογή της μηχανικής μάθησης στον τομέα της ιατρικής απεικόνισης για την ανίχνευση και ταξινόμηση όγκων του εγκεφάλου. Πιο συγκεκριμένα οι ερευνητές, με σκοπό τη σωστή διάγνωση και ταξινόμηση των όγκων του εγκεφάλου από εικόνες MRI σε τέσσερις κατηγορίες: το γλοίωμα, το μηνιγγίωμα, τον όγκο της υπόφυσης και εικόνες χωρίς εγκεφαλικό όγκο, χρησιμοποίησαν προεκπαιδευμένα συνελκτικά νευρωνικά δίκτυα. Τα προεκπαιδευμένα αυτά μοντέλα, που χρησιμοποιήθηκαν και στη συνέχεια τα αποτελέσματα τους συγκρίθηκαν, ήταν τα εξής: `EfficientNetB1`, `EfficientNetB7`, `EfficientNetV2B1` και `ResNet50`. Το σύνολο των δεδομένων που χρησιμοποιήθηκε για την πραγματοποίηση των πειραμάτων της παρούσας έρευνας προήλθε από τη βάση δεδομένων του Kaggle. Το σύνολο αυτό αποτελεί έναν συνδυασμό τριών άλλων βάσεων δεδομένων, `figshare`, `SARTAJ` και `Br25H`. Συνολικά χρησιμοποιήθηκαν 7022 εικόνες MRI του ανθρώπινου εγκεφάλου, οι οποίες χωρίστηκαν σε σύνολο εκπαίδευσης και δοκιμής με αναλογία 80% και 20% αντίστοιχα. Κατά την προεπεξεργασία των δεδομένων, με σκοπό την αύξηση του αριθμού των εικόνων, που το μοντέλο διακρίνει ως ξεχωριστές εικόνες, εφαρμόστηκαν οι εξής τεχνικές `data augmentation`: περιστροφή κατά σταθερή γωνία 90 μοιρών, οριζόντια αναστροφή και κάθετη αναστροφή. Το μέγεθος των εικόνων διατηρήθηκε στην αρχική του μορφή, στα 512x512 pixels. Οι ερευνητές τροποποίησαν τα προεκπαιδευμένα μοντέλα, αφαιρώντας το τελευταίο πλήρως συνδεδεμένο στρώμα, προσθέτοντας μία σειρά νέων στρωμάτων ορίζοντας έτσι τη νέα αρχιτεκτονική. Πιο συγκεκριμένα, η νέα αυτή αρχιτεκτονική περιλάμβανε ένα `Average Pooling` στρώμα, ακολούθως ένα `Flatten` στρώμα και έπειτα δύο πλήρως συνδεδεμένα στρώματα, κάθε ένα από τα οποία ακολουθούσε ένα `Dropout` στρώμα. Τέλος, ένα πλήρως συνδεδεμένο στρώμα, με 4 νευρώνες για την τελική ταξινόμηση και ως συνάρτηση απώλειας την `Categorical Cross Entropy`. Κατά την εκπαίδευση χρησιμοποιήθηκε ο `Adam optimizer`, ο ρυθμός μάθησης ήταν 0.001 και το `batch size` ίσο με 8. Επιπλέον εφαρμόστηκε το `Callback Early Stopping`. Τα βάρη του μοντέλου αποθηκεύονταν κάθε φορά που επιτυγχάνονταν η ελάχιστη τιμή της απώλειας επικύρωσης. Ο μέγιστος αριθμός των εποχών ορίστηκε να είναι 50 και στην περίπτωση που δεν υπήρχε μείωση της απώλειας για 9 συνεχόμενες εποχές η εκπαίδευση του μοντέλου σταματούσε. Τα μοντέλα αξιολογήθηκαν με βάση το `training accuracy`, το `training loss`, το `validation accuracy` και το `validation loss`. Το `EfficientNetB1` έδειξε την καλύτερη απόδοση με την υψηλότερη ακρίβεια για τα σύνολα εκπαίδευσης και επικύρωσης, καθώς και την χαμηλότερη απώλεια για

το σύνολο επικύρωσης. Συνολικά όλα τα μοντέλα EfficientNet που χρησιμοποιήθηκαν έδειξαν μία σταθερή αύξηση στην ακρίβεια και μία σταθερή μείωση στην απώλεια του συνόλου εκπαίδευσης. Το ResNet50 είχε την χαμηλότερη απόδοση και επιπλέον από την καμπύλη ακρίβειας του συνόλου εκπαίδευσης, παρατηρήθηκε αρχική μείωση και στη συνέχεια αύξηση αυτής.

6η Μελέτη

Η μελέτη που ακολουθεί επικεντρώνεται κυρίως στην αύξηση της ακρίβειας ταξινόμησης εικόνων με εγκεφαλικούς όγκους, χρησιμοποιώντας την τεχνική Transfer Learning, βασισμένη σε βαθιά νευρωνικά δίκτυα. Το σύνολο των δεδομένων που χρησιμοποιήθηκε προέρχεται από 233 ασθενείς και αποτελείται από 3064 MRI εικόνες όγκων του εγκεφάλου. Πιο συγκεκριμένα, το σύνολο των δεδομένων διαμορφώνεται με 1426 εικόνες με μηνιγγίωμα, 708 εικόνες με γλοιώμα και 930 εικόνες όγκου της υπόφυσης. Κάθε μία από τις εικόνες είχε αρχικό μέγεθος 512x512 pixels, όπου για τη μείωση του χρόνου υπολογισμού μειώθηκε στα 150x150 pixels. Στη συνέχεια το σύνολο των δεδομένων χωρίστηκε σε σύνολο εκπαίδευσης, που αποτελούσε το 80% του συνολικού συνόλου δεδομένων, και σύνολο επικύρωσης που αποτελούσε το 20% του συνολικού συνόλου δεδομένων. Κατά την προεπεξεργασία των εικόνων, με σκοπό την αύξηση της ποικιλομορφίας του συνόλου των δεδομένων, πραγματοποιήθηκε μία σειρά τεχνικών image augmentation που περιλάμβανε τις εξής: ζουμ, οριζόντια αναστροφή, μετατόπιση κατά πλάτος, μετατόπιση κατά ύψος και shear. Η προτεινόμενη αρχιτεκτονική αποτελείται από το InceptionV3 και ένα προσαρμοσμένο βαθύ νευρωνικό δίκτυο. Για την ενίσχυση της αποτελεσματικότητας του ταξινομητή, η έξοδος του InceptionV3 συγχρονίζεται με το βαθύ νευρωνικό δίκτυο που κατασκεύασαν οι ερευνητές, γεγονός που οδηγεί στο fine-tuning του transfer learning μοντέλου. Το βαθύ αυτό νευρωνικό δίκτυο, που κατασκευάστηκε, αποτελείται από τέσσερα επίπεδα. Πιο συγκεκριμένα χρησιμοποιήθηκε ένα στρώμα Flatten, στη συνέχεια ένα πλήρως συνδεδεμένο στρώμα αποτελούμενο από 1024 νευρώνες, ακολουθούμενο από ένα στρώμα Dropout. Τέλος, το τελικό πλήρως συνδεδεμένο στρώμα, που σχεδιάστηκε για να παράγει την έξοδο, με 3 νευρώνες που αντιπροσωπεύουν τους τύπους όγκων εγκεφάλου του γλοιώματος, του μηνιγγιώματος και του όγκου της υπόφυσης. Για την εκπαίδευση του δικτύου χρησιμοποιήθηκε ο Adam optimizer, ο ρυθμός μάθησης ήταν ίσος με 0,00003, το batch size ίσο με 32 και ως συνάρτηση απώλειας χρησιμοποιήθηκε η Categorical Cross Entropy. Το μοντέλο εκπαιδεύτηκε για 19 εποχές και η διαδικασία αυτή διήρκεσε περίπου 6 λεπτά. Η απώλεια εκπαίδευσης και επικύρωσης καταγράφηκαν 0.0614 και 0.1468 αντίστοιχα και οι καμπύλες του εμφάνισαν φθίνουσα πορεία κατά τη διάρκεια των εποχών. Η ακρίβεια που επιτεύχθηκε για τα σύνολα εκπαίδευσης και επικύρωσης ήταν 97,80% και 96,25% αντίστοιχα, γεγονός που σε συνδυασμό με την αύξουσα πορεία που παρουσίασαν οι καμπύλες ακρίβειας και για τα δύο σύνολα κατά τη διάρκεια των εποχών εκπαίδευσης, εξασφαλίζουν την καλή προσαρμογή του προτεινόμενου μοντέλου. Επιπλέον για την αξιολόγηση της απόδοσης του μοντέλου χρησιμοποιήθηκαν τα εξής μέτρα: accuracy, precision, recall και F1-score. Το προτεινόμενο, μοντέλο εμφάνισε ικανοποιητικές τιμές σε όλα τα μέτρα και πιο συγκεκριμένα το precision και η ευαισθησία του συνόλου εκπαίδευσης καταγράφη-

καν στο 98,16% και 96,39% αντίστοιχα και για το σύνολο επικύρωσης στο 97,55% και 95,92% αντίστοιχα. Τέλος οι τιμές του F1-score για τα σύνολα εκπαίδευσης και επικύρωσης υπολογίστηκαν σε 97,85% και 96,15% αντίστοιχα.

7η Μελέτη

Στην εργασία που παρουσιάζεται, οι ερευνητές επικεντρώθηκαν στην ανάπτυξη ενός συνελκτικού νευρωνικού δικτύου, το οποίο θα ταξινομεί εικόνες MRI εγκεφαλικών όγκων μεταξύ τριών κλάσεων: το γλοιώμα, το μηνιγγίωμα και τον όγκο της υπόφυσης, με σκοπό τον έγκαιρο εντοπισμό τους και την έναρξη των θεραπειών των ασθενών σε πρώιμο στάδιο. Το σύνολο των δεδομένων που χρησιμοποιήθηκε, περιλαμβάνει 3064 εικόνες MRI και έχει συλλεχθεί από 233 ασθενείς. Πιο συγκεκριμένα, αποτελείται από 708 εικόνες με μηνιγγίωμα, 1426 εικόνες με γλοιώμα και 930 εικόνες με όγκο της υπόφυσης. Κατά την προ επεξεργασία των εικόνων, αυτές μειώθηκαν από το αρχικό τους μέγεθος, που ήταν 512x512 pixels, στα 112x112 pixel. Στη συνέχεια εφαρμόστηκε το Gaussian φίλτρο με μέγεθος πυρήνα 5x5, που ομαλοποιεί την εικόνα. Ακολούθως χρησιμοποιήθηκε η τεχνική histogram equalization σε κάθε μία από τις εικόνες, μία τεχνική επεξεργασίας εικόνων που βελτιώνει την αντίθεση αυτών. Το συνολικό σύνολο δεδομένων χωρίστηκε σε τρία υποσύνολα. Το σύνολο εκπαίδευσης, που αποτελείται από 2051 εικόνες, το σύνολο επικύρωσης αποτελούμενο από 513 εικόνες και το σύνολο δοκιμών, αποτελούμενο από 500 εικόνες. Οι ερευνητές χρησιμοποίησαν συνέλιξη σε κάθε εισερχόμενη εικόνα, με ένα πυρήνα 3x3 ως φίλτρο συνέλιξης, για τη δημιουργία ενός feature map. Το προτεινόμενο συνελκτικό δίκτυο αποτελείται από 3 στρώματα συνέλιξης, με μέγεθος πυρήνα 5x5 και 2 downsampling στρώματα με pooling window 2x2. Σε κάθε συνελκτικό στρώμα χρησιμοποιείται η συνάρτηση ενεργοποίησης ReLu. Σε κάθε downsampling στρώμα χρησιμοποιείται max pooling, για την μείωση του αριθμού των παραμέτρων και συνεπώς του χρόνου που απαιτείται για την εκπαίδευση του δικτύου. Χρησιμοποιήθηκαν, επίσης, 4 dropout στρώματα και 2 πλήρως συνδεδεμένα στρώματα. Στο τελευταίο πλήρως συνδεδεμένο στρώμα χρησιμοποιήθηκαν 3 νευρώνες, για την ταξινόμηση των εικόνων στις 3 κλάσεις εγκεφαλικών όγκων, και ως συνάρτηση ενεργοποίησης η Softmax. Το μοντέλο εκπαιδεύτηκε για 100 εποχές, χρησιμοποιήθηκε ο Adam optimizer και το batch size ήταν ίσο με 256. Για την αξιολόγηση του προτεινόμενου μοντέλου χρησιμοποιήθηκαν τα εξής μέτρα απόδοσης: confusion matrix, precision, recall, F1-score, support και η καμπύλη ROC. Το μοντέλο, στο σύνολο δοκιμής, κατέγραψε απώλεια 28.16% και accuracy 94,39%. Η περιοχή κάτω από την καμπύλη ROC κυμαινόταν από 0.99 έως 1, για τις διάφορες κλάσεις. Το μοντέλο πέτυχε precision 88%, 94% και 98% για τις κλάσεις του γλοιώματος, του μηνιγγιώματος και του όγκου της υπόφυσης αντίστοιχα, γεγονός που οδηγεί σε average precision 93.33%. Έδειξε υψηλότερη ικανότητα αναγνώρισης των περιπτώσεων του όγκου της υπόφυσης συγκριτικά με τις άλλες δύο κατηγορίες.

8η Μελέτη

Στην εργασία που ακολουθεί με σκοπό την ταξινόμηση εικόνων MRI με εγκεφαλικούς όγκους, σε καρκινικούς και μη καρκινικούς, οι ερευνητές πρότειναν ένα απλό συνελκτικό νευρωνικό δίκτυο που μειώνει την υπολογιστική πολυπλοκότητα και την απαίτηση μεγάλου όγκου δεδομένων για την εκπαίδευσή του. Επιπλέον, έγινε σύγκριση του προτεινόμενου μοντέλου με τα εξής προ εκπαιδευμένα μοντέλα: VGG-16, ResNet50 και InceptionV3. Το σύνολο των δεδομένων που χρησιμοποιήθηκε αποτελείται από 253 εικόνες MRI του εγκεφάλου. Πιο συγκεκριμένα περιλαμβάνει 155 εικόνες καρκινικών όγκων και 98 εικόνες με καλοήθεις όγκους. Το σύνολο αυτό, χωρίστηκε σε τρία υποσύνολα για τη διενέργεια των πειραμάτων: το σύνολο εκπαίδευσης, αποτελούμενο από 185 εικόνες, το σύνολο επικύρωσης, αποτελούμενο από 48 εικόνες, και το σύνολο δοκιμών, αποτελούμενο από 20 εικόνες. Κατά την προ επεξεργασία των εικόνων χρησιμοποιήθηκε ο αλγόριθμος Canny Edge Detection και στη συνέχεια έγινε περικοπή και κρατήθηκε μόνο το τμήμα της εικόνας στο οποίο παρουσιάζεται ο εγκέφαλος. Καθώς το σύνολο των δεδομένων που χρησιμοποιήθηκε ήταν αρκετά μικρό, εφαρμόστηκαν, στο σύνολο εκπαίδευσης, διάφορες τεχνικές image augmentation, όπως περιστροφή, αναστροφή και φωτεινότητα. Το μέγεθος των εικόνων εισόδου ορίστηκε στα 224x224 pixels και το batch size ίσο με 32. Το προτεινόμενο μοντέλο αποτελείται από 8 συνελκτικά στρώματα, με φίλτρα μεγέθους 3x3, και 4 max pooling στρώματα. Σε κάθε στρώμα συνέλκξης χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLu. Τέλος χρησιμοποιήθηκε ένα πλήρως συνδεδεμένο στρώμα, 256 νευρώνων, μαζί με το επίπεδο εξόδου Softmax για τον υπολογισμό της πιθανότητας κάθε κατηγορίας και την υλοποίηση της τελικής ταξινόμησης. Κατά τη διαδικασία της εκπαίδευσης χρησιμοποιήθηκε ο Adam optimizer. Πραγματοποιήθηκε σύγκριση της απόδοσης του προτεινόμενου μοντέλου, με τα προ εκπαιδευμένα μοντέλα VGG-16, ResNet και InceptionV3, διατηρώντας παγωμένα κάποια από τα επίπεδά τους για την αποφυγή της υπερπροσαρμογής. Όλα τα μοντέλα εκπαιδεύτηκαν για 15 εποχές και με batch size ίσο με 32. Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν τα εξής μέτρα απόδοσης: accuracy, precision, recall, F1-score και ROC-AUC score. Το προτεινόμενο μοντέλο έδειξε υπεροχή έναντι των προ εκπαιδευμένων, έχοντας καταγράψει accuracy 100% με τα VGG-16, ResNet50 και InceptionV3 να καταγράφουν 96%, 89% και 75% αντίστοιχα. Το προτεινόμενο μοντέλο έδειξε υπεροχή και στις τιμές των υπόλοιπων μέτρων αξιολόγησης και κατέγραψε τον μικρότερο μέσο χρόνο εκπαίδευσης ανά εποχή, στα 205 δευτερόλεπτα, έναντι των 456, 606 και 375 δευτερολέπτων που κατέγραψαν τα VGG-16, ResNet50 και InceptionV3 αντίστοιχα.

9η Μελέτη

Στην εργασία που ακολουθεί, προτείνεται ένα σύστημα για την αυτόματη ταξινόμηση καρκίνων του εγκεφάλου βασισμένο σε δύο μοντέλα βαθιάς μάθησης. Πιο συγκεκριμένα, προτείνεται μία αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου 23 στρωμάτων για την ταξινόμηση τριών τύπων εγκεφαλικών όγκων(μηνιγγίωμα, γλοίωμα, όγκος της υπόφυσης) και ένα τροποποιημένο VGG-16 δίκτυο, που ενσωματώνει ορισμένα κομμάτια της αρχιτεκτονικής του συνελκτικού νευρωνικού

δικτύου των 23 στρωμάτων, για τη δυαδική ταξινόμηση εικόνων με βάση την ανίχνευση ή μη ανίχνευση εγκεφαλικών όγκων. Για την πραγματοποίηση των πειραμάτων χρησιμοποιήθηκαν δύο διαφορετικά σύνολα δεδομένων. Το πρώτο(αναφερόμενο ως σύνολο δεδομένων 1) είναι το σύνολο δεδομένων CE-MRIFigshare, που αποτελείται από 3064 εικόνες MRI από 233 ασθενείς που διαγνώστηκαν με έναν από του τρεις τύπους εγκεφαλικών όγκων, δηλαδή το μηνιγγίωμα, το γλοίωμα και τον όγκο της υπόφυσης. Το δεύτερο(αναφερόμενο ως σύνολο δεδομένων 2) συλλέχθηκε από το αρχείο του Harvard και περιλαμβάνει 152 εικόνες MRI, από τις οποίες οι 71 δεν περιέχουν κανέναν όγκο και οι 81 περιέχουν εγκεφαλικό όγκο. Τα δύο σύνολα δεδομένων χωρίστηκαν σε υποσύνολα ως εξής: 70% των δεδομένων χρησιμοποιήθηκε στο σύνολο εκπαίδευσης και 30% των δεδομένων στα σύνολα επικύρωσης και δοκιμής. Αναφορικά με το σύνολο δεδομένων 2, ορισμένες τεχνικές image augmentation εφαρμόστηκαν για την αποφυγή της υπερπροσαρμογής εξαιτίας του μικρού αριθμού των δεδομένων. Το συνελκτικό νευρωνικό δίκτυο που πρότειναν οι ερευνητές αποτελείται από τέσσερα στρώματα συνέλιξης με διαφορετικά μεγέθη πυρήνων και συνάρτηση ενεργοποίησης τη ReLu. Χρησιμοποιήθηκαν 3 max pooling στρώματα και 7 batch normalization στρώματα. Στη συνέχεια χρησιμοποιήθηκε ένα global average pooling στρώμα, ακολουθούμενο από τέσσερα πλήρως συνδεδεμένα στρώματα. Η ταξινόμηση λαμβάνει χώρα στο τελευταίο πλήρως συνδεδεμένο στρώμα και ως συνάρτηση ενεργοποίησης, στο στρώμα αυτό, χρησιμοποιήθηκε η συνάρτηση Softmax. Για την αποφυγή της υπερπροσαρμογής ένα στρώμα Dropout έχει τοποθετηθεί πριν από το επίπεδο της τελικής ταξινόμησης. Η παραπάνω αρχιτεκτονική χρησιμοποιήθηκε για την ταξινόμηση των εικόνων του συνόλου δεδομένων 1 και κατά την εκπαίδευση χρησιμοποιήθηκε ο Adam optimizer και η συνάρτηση απώλειας Sparse Categorical Cross Entropy. Αναφορικά με το σύνολο δεδομένων 2, καθώς, λόγω του μικρού αριθμού των διαθέσιμων δεδομένων θα εμφανίζονταν το φαινόμενο της υπερπροσαρμογής με τη χρήση του προτεινόμενου συνελκτικού νευρωνικού δικτύου, οι ερευνητές χρησιμοποίησαν την τεχνική Transfer Learning, με βάση το VGG-16. Αυτό συνδυάστηκε με το συνελκτικό νευρωνικό δίκτυο των 23 επιπέδων. Κατά την εκπαίδευση χρησιμοποιήθηκε ο Adam optimizer και η συνάρτηση απώλειας Categorical Cross Entropy. Τα προτεινόμενα μοντέλα αξιολογήθηκαν με μία σειρά μέτρων που είναι τα εξής: accuracy, precision, recall, f1-score, True negative rate και false positive rate. Οι τιμές των μέτρων απόδοσης των δύο μεθόδων ήταν εξαιρετικά ικανοποιητικές. Το συνελκτικό νευρωνικό δίκτυο των 23 στρωμάτων πέτυχε accuracy, στο σύνολο δοκιμής, 97.8% στο πρώτο σύνολο δεδομένων και ο συνδυασμός του προ εκπαιδευμένου VGG-16 με το συνελκτικό νευρωνικό δίκτυο των 23 στρωμάτων κατέγραψε accuracy, στο σύνολο δοκιμής, 100% για το δεύτερο σύνολο δεδομένων.

Κεφάλαιο 3

Νευρωνικά Δίκτυα

Με τον όρο νευρωνικά δίκτυα, αναφερόμαστε σε μία ευρύτερη οικογένεια μοντέλων που χαρακτηρίζεται από μεγάλο χώρο παραμέτρων και ευέλικτη δομή. Η λειτουργία του κεντρικού νευρικού συστήματος και του ανθρώπινου εγκεφάλου αποτελεί την έμπνευση πίσω από την αρχιτεκτονική των νευρωνικών δικτύων. Τα κύτταρα του ανθρώπινου εγκεφάλου που ονομάζονται νευρώνες αποτελούν ένα περίπλοκο, ιδιαίτερα διασυνδεδεμένο δίκτυο. Ένας νευρώνας είναι ένα εξειδικευμένο κύτταρο, με την δυνατότητα μετάδοσης ηλεκτροχημικού σήματος. Οι νευρώνες στέλνουν ηλεκτροχημικά σήματα, ο ένας στον άλλο, με σκοπό να βοηθήσουν τους ανθρώπους να επεξεργαστούν τις διάφορες πληροφορίες που δέχονται. Ομοίως, ένα τεχνητό νευρωνικό δίκτυο αποτελείται από πολλά διασυνδεδεμένα στοιχεία γνωστά ως «νευρώνες» ή «κόμβοι» που συνεργάζονται για την επεξεργασία και την ανάλυση πληροφοριών. Όπως συμβαίνει με τον ανθρώπινο εγκέφαλο έτσι και στην περίπτωση ενός νευρωνικού δικτύου, η γνώση αποκτάται από το δίκτυο μέσω μιας μαθησιακής διαδικασίας, ενώ για την αποθήκευση της γνώσης χρησιμοποιείται η ισχύς της σύνδεσης των νευρώνων, που είναι γνωστή και ως «συναπτικά βάρη». Έτσι μπορούμε να θεωρήσουμε ότι ένα νευρωνικό δίκτυο είναι ένας μαζικός και παράλληλα κατανεμημένος επεξεργαστής, που έχει μία φυσική τάση για την αποθήκευση της βιωματικής γνώσης και της διάθεσής της για χρήση[35].

Τα νευρωνικά δίκτυα προτάθηκαν για πρώτη φορά, το 1944 από τους Warren McCullough και Walter Pitts. Οι ερευνητές πρότειναν ένα μοντέλο νευρωνικού δικτύου, στο οποίο διασυνδεδεμένοι νευρώνες μπορούσαν να ανταποκριθούν σε εξωτερικά ερεθίσματα και να εκτελούν απλούς υπολογισμούς. Το 1950 τα νευρωνικά δίκτυα υλοποιήθηκαν για πρώτη φορά σε υπολογιστές και το 1958 ο Frank Rosenblatt δημιούργησε ένα πρώιμο μοντέλο νευρωνικού δικτύου το Perceptron. Η αρχιτεκτονική του Perceptron, αρκετά παρόμοια με αυτή των σύγχρονων νευρωνικών δικτύων, αποτελούνταν από ένα επίπεδο εισόδου, ένα επίπεδο με προσαρμόσιμα βάρη και κατώφλια και ένα επίπεδο εξόδου. Από τη δεκαετία του 1980 και μετά τα νευρωνικά δίκτυα γνώρισαν μεγάλη ανάπτυξη, ιδιαίτερα με την εμφάνιση των αλγορίθμων βαθιάς μάθησης και της αυξημένης δυνατότητας επεξεργασίας των σύγχρονων υπολογιστών. Η χρήση νευρωνικών δικτύων για την επεξεργασία και ταξινόμηση εικόνων, ξεκίνησε τη δεκαετία του 1980, αναδείχθηκε όμως μετά τα τέλη του 2000 με την ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων.

3.1 Feedforward Νευρωνικά Δίκτυα

3.1.1 Single layer Perceptron

Το Single layer Perceptron είναι ένας γραμμικός ταξινομητής δυαδικής ταξινόμησης με μία απλή σχέση εισόδου-εξόδου. Ο αλγόριθμος δεν έχει καμία εκ των προτέρων γνώση και για αυτό στην αρχή της εκπαίδευσης αρχικοποιεί το διάνισμα των βαρών με μικρές τυχαίες τιμές ή με τιμές ίσες με το 0. Στη συνέχεια υπολογίζει το άθροισμα του γινομένου των εισόδων επί τα αντίστοιχα βάρη συν τον όρο bias, που μαθηματικά μπορεί να διατυπωθεί ως εξής:

$$z = \sum_{i=1}^n w_i \cdot x_i + b, \quad (3.1)$$

όπου, w_i είναι τα βάρη που συνδέονται με κάθε είσοδο και x_i είναι οι τιμές της εισόδου. Κάθε x_i αντιπροσωπεύει μια διαφορετική είσοδο στο νευρωνικό δίκτυο. Το b είναι ο όρος bias. Ο όρος bias είναι μία τιμή (δεν εξαρτάται από τις εισόδους), που προστίθεται στο συνολικό άθροισμα και επηρεάζει το επίπεδο στο οποίο ο νευρώνας θα ενεργοποιηθεί ή όχι.

Η τιμή z δίνεται ως όρισμα στην συνάρτηση ενεργοποίησης, που στην περίπτωση αυτή είναι συνήθως μία συνάρτηση βήματος Heaviside με κατώφλι τιμής 0.5. Αν η τιμή του z υπερβαίνει την τιμή του καθορισμένου κατωφλιού ο νευρώνας ενεργοποιείται (έξοδος=1), διαφορετικά παραμένει ανενεργός (έξοδος=0). Αφού υπολογιστεί η προβλεπόμενη ταξινόμηση κάθε εισόδου, αυτή συγκρίνεται με την πραγματική ταξινόμηση. Εάν η προβλεπόμενη έξοδος ταυτίζεται με την πραγματική έξοδο τότε η απόδοση θεωρείται ικανοποιητική και δεν γίνονται αλλαγές στα βάρη. Στην περίπτωση που η πρόβλεψη είναι λανθασμένη τα βάρη και ο όρος bias ανανεώνονται, με σκοπό να μειωθεί το σφάλμα δηλαδή η απόκλιση της προβλεπόμενης από την πραγματική έξοδο ως εξής:

$$w_i(\text{new}) = w_i(\text{old}) + \Delta w_i \quad (3.2)$$

$$b(\text{new}) = b(\text{old}) + \Delta b \quad (3.3)$$

Όπου:

$$\Delta w_i = \eta \cdot (t - y) \cdot x_i$$

$$\Delta b = \eta \cdot (t - y)$$

η : ο ρυθμός μάθησης

t : η πραγματική ετικέτα

y : η προβλεπόμενη έξοδος

Ο αλγόριθμος εκμάθησης Perceptron αλλάζει τα βάρη και τα biases στο μοντέλο μέχρι όλες οι εγγραφές εισόδου να ταξινομηθούν σωστά. Ο αλγόριθμος δεν θα τερματιστεί αν η είσοδος εκμάθησης δεν είναι γραμμικά διαχωρίσιμη. Ένα γραμμικά διαχωρίσιμο σετ δεδομένων είναι ένα για το οποίο μπορούμε να βρούμε τις τιμές ενός υπερεπιπέδου που θα διαχωρίσει καθαρά τις δύο κατηγορίες του σετ δεδομένων.

3.1.2 Multilayer Perceptron

Τα feedforward νευρωνικά δίκτυα ή αλλιώς πολυεπίπεδα Perceptrons, αποτελούν τα πιο βασικά και ευρέως χρησιμοποιούμενα μοντέλα στην τεχνητή νοημοσύνη και τη μηχανική μάθηση. Η ονομασία τους οφείλεται στο γεγονός πως η πληροφορία, δηλαδή τα δεδομένα, ρέουν προς μία μόνο κατεύθυνση, δηλαδή από το επίπεδο εισόδου προς το επίπεδο εξόδου μέσω των ενδιάμεσων κρυφών επιπέδων. Δεν υπάρχει καμία αναδρομή ή επαναχρησιμοποίηση των εξόδων του δικτύου στο ίδιο. Στόχος ενός feedforward νευρωνικού δικτύου είναι η να προσεγγίσει κάποια συνάρτηση f^* . Αυτό σημαίνει πως το νευρωνικό δίκτυο προσπαθεί να μάθει ένα σύνολο σχέσεων μεταξύ των εισόδων και των εξόδων, ώστε να παράγει ακριβείς προβλέψεις ή ταξινομήσεις. Για παράδειγμα, για έναν ταξινομητή, η σχέση $y = f^*(x)$ αντιστοιχεί μία είσοδο x σε μία κατηγορία y . Το δίκτυο, ορίζει μία συνάρτηση $y = f(x; \theta)$ και μαθαίνει τις τιμές των παραμέτρων θ (π.χ βάρη και biases), που οδηγούν στην καλύτερη προσέγγιση της συνάρτησης f^* . Η εκπαίδευση ενός feedforward νευρωνικού δικτύου περιλαμβάνει την προσαρμογή των βαρών και των biases του δικτύου με στόχο την ελαχιστοποίηση της διαφοράς μεταξύ των προβλεπόμενων και των πραγματικών εξόδων (σφάλματα πρόβλεψης ή ταξινόμησης). Κατά τη διαδικασία της εκπαίδευσης ενός feedforward νευρωνικού δικτύου, εκτελείται ο αλγόριθμος forward propagation. Τα δεδομένα εισάγονται στο επίπεδο εισόδου, όπου οι νευρώνες εισόδου θα τα μεταφέρουν στα εσωτερικά επίπεδα χωρίς να εκτελέσουν κάποιο υπολογισμό. Τα δεδομένα εισόδου συμβολίζονται ως x . Στη συνέχεια οι υπολογιστικοί νευρώνες, που βρίσκονται στα κρυφά επίπεδα, πραγματοποιούν έναν υπολογισμό βασισμένο στα βάρη και τα biases που συνδέονται με τον καθένα. Ο υπολογισμός είναι το σταθμισμένο άθροισμα S , δηλαδή ένα άθροισμα των εισόδων επί τα συναπτικά βάρη, συν το bias ακολουθούμενο από την εφαρμογή μίας συνάρτησης ενεργοποίησης. Μαθηματικά μπορεί να διατυπωθεί ως εξής:

$$\sum_{i=1}^N (x_{ki} w_{ki} + w_{k0}), \quad (3.4)$$

όπου, x_{ki} είναι η i -οστή είσοδος στον νευρώνα k , που αντιπροσωπεύει τις διάφορες πληροφορίες που λαμβάνει ο νευρώνας, w_{ki} είναι το συναπτικό βάρος που αντιστοιχεί στην i -οστή είσοδο του νευρώνα k , και καθορίζει πόσο ισχυρά επηρεάζει κάθε είσοδος την τελική έξοδο, w_{k0} είναι η πόλωση ή το κατώφλι (bias) για τον νευρώνα k , που λειτουργεί ως ένας ρυθμιστικός παράγοντας για το επίπεδο ενεργοποίησης του νευρώνα.

Με βάση την τιμή που προκύπτει από την συνάρτηση ενεργοποίησης αποφασίζεται αν θα ενεργοποιηθεί ή όχι ένας νευρώνας. Στην περίπτωση, όπου η τιμή αυτή υπερβαίνει μία ορισμένη τιμή κατωφλιού ο νευρώνας ενεργοποιείται και παράγει ένα σήμα, διαφορετικά ο νευρώνας παραμένει ανενεργός. Η διαδικασία συνεχίζεται μέχρι η πληροφορία να φτάσει στο επίπεδο εξόδου, όπου οι νευρώνες διαμορφώνουν την τελική έξοδο \hat{y} . Στη συνέχεια, το δίκτυο υπολογίζει πόσο κοντά είναι οι προβλέψεις του στις πραγματικές τιμές και χρησιμοποιώντας συνήθως τον αλγόριθμο backpropagation υπολογίζει τα gradients της συνάρτησης απώλειας σε σχέση με τα βάρη, δηλαδή υπολογίζει πως κάθε βάρος επηρεάζει το σφάλμα του νευρωνικού δικτύου. Με τη χρήση μεθόδων όπως η Stochastic Gradient Descent (ή και παραλλαγές

της) τα βάρη ενημερώνονται με σκοπό της ελαχιστοποίηση της συνάρτησης απώλειας. Με τον τρόπο αυτό, τα νευρωνικά δίκτυα βελτιώνουν σταδιακά την ακρίβεια των προβλέψεών τους.

3.1.3 Συνελικτικά Νευρωνικά Δίκτυα(Convolutional Neural Networks-CNNs)

Τα συνελικτικά νευρωνικά δίκτυα αποτελούν μία εξειδικευμένη κατηγορία νευρωνικών δικτύων για την επεξεργασία δεδομένων, που έχουν μία γνωστή τοπολογία μορφής πλέγματος. Παραδείγματα περιλαμβάνουν τα δεδομένα χρονοσειρών που μπορούν να θεωρηθούν ως ένα 1D πλέγμα δεδομένων που λαμβάνει δείγματα σε τακτά χρονικά διαστήματα και τα δεδομένα εικόνων που μπορούν να θεωρηθούν ως ένα 2D πλέγμα εικονοστοιχείων. Τα συνελικτικά νευρωνικά δίκτυα έχουν σημειώσει τεράστια επιτυχία σε μία σειρά εφαρμογών όπως για παράδειγμα η αναγνώριση εικόνων, η αναγνώριση ομιλίας, η επεξεργασία φυσικής γλώσσας και άλλες περιοχές της τεχνητής νοημοσύνης. Το όνομα συνελικτικά νευρωνικά δίκτυα δηλώνει ότι το δίκτυο χρησιμοποιεί μία μαθηματική λειτουργία που ονομάζεται συνέλιξη. Τα συνελικτικά νευρωνικά δίκτυα, είναι απλώς νευρωνικά δίκτυα που χρησιμοποιούν τη συνέλιξη, αντί του πλήρη πολλαπλασιασμού μητρώων, σε τουλάχιστον ένα από τα επίπεδά τους. Τα συνελικτικά νευρωνικά δίκτυα αναπτύχθηκαν λόγω της ανάγκης για εξειδικευμένη εξαγωγή χαρακτηριστικών από δεδομένα εικόνας. Τα συνελικτικά νευρωνικά δίκτυα μετασχηματίζουν τα δεδομένα εισόδου, από το επίπεδο εισόδου, μέσω όλων των συνδεδεμένων επιπέδων σε ένα σύνολο από βαθμολογίες κλάσεων που παρέχονται από το επίπεδο εξόδου. Υπάρχουν αρκετές παραλλαγές της αρχιτεκτονικής των CNN, αλλά βασίζονται στο εξής μοτίβο επιπέδων:

- Επίπεδο Εισόδου
- Επίπεδο Εξαγωγής Χαρακτηριστικών
 1. Στρώματα Συνέλιξης(Convolutional Layers)
 2. Στρώματα Ενεργοποίησης ReLu(ReLU Activation Function as Layers)
 3. Στρώματα Υποδειγματοληψίας(Pooling Layers)
- Πλήρως Συνδεδεμένα Στρώματα(Fully Connected Layers)
- Επίπεδο Εξόδου

3.1.3.1 Επίπεδο Εισόδου

Το επίπεδο εισόδου δέχεται τα ακατέργαστα δεδομένα εικόνας. Αυτά τα δεδομένα καθορίζουν την τρισδιάστατη είσοδο του δικτύου, δηλαδή το πλάτος, το ύψος και τον αριθμό των καναλιών. Τις περισσότερες φορές ο αριθμός των καναλιών είναι τρία για τα RGB χρωματικά κανάλια για κάθε εικονοστοιχείο.

3.1.3.2 Επίπεδο Εξαγωγής Χαρακτηριστικών

Στρώματα Συνέλιξης(Convolutional Layers)

Τα συνελκτικά στρώματα αποτελούν τους δομικούς λίθους της αρχιτεκτονικής των συνελκτικών νευρωνικών δικτύων. Τα συνελκτικά νευρωνικά στρώματα μετασχηματίζουν τα δεδομένα εισόδου χρησιμοποιώντας νευρώνες που είναι συνδεδεμένοι με ένα τοπικό πεδίο της εικόνας εισόδου. Στη συνέχεια το ίδιο φίλτρο(με τα ίδια βάρη), εφαρμόζεται σε όλη την εικόνα γεγονός που συμβάλει στην παραμετρική αποδοτικότητα του δικτύου. Η παραγόμενη έξοδος έχει γενικά τις ίδιες ή μικρότερες χωρικές διαστάσεις, μερικές φορές όμως αυξάνει τον αριθμό των στοιχείων της τρίτης διάστασης, της διάστασης του βάθους. Σε ένα στρώμα συνέλιξης η βασική διαδικασία που υλοποιείται είναι η διαδικασία της συνέλιξης, που είναι καθοριστική για την αποδοτικότητα και την ακρίβεια του δικτύου και αναλύεται εκτενέστερα παρακάτω. Τα συνελκτικά στρώματα επιτυγχάνουν την μείωση του αριθμού των παραμέτρων που πρέπει να εκπαιδευτούν μέσω της τοπικής συνδεσιμότητας, που σημαίνει ότι κάθε νευρώνας στα στρώματα αυτά αλληλεπιδρά μόνο με ένα μικρό τμήμα της εικόνας εισόδου, και της κοινής χρήσης παραμέτρων. Αυτό οδηγεί σε πιο αποδοτικά μοντέλα τόσο όσον αφορά τη μνήμη όσο και την υπολογιστική ισχύ του δικτύου.

Συνέλιξη

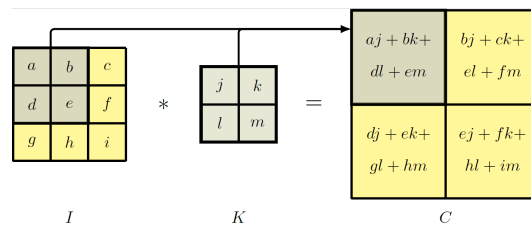
Η συνέλιξη, στο πλαίσιο των συνελκτικών νευρωνικών δικτύων, αναφέρεται στη μαθηματική διαδικασία που συνδυάζει δύο συναρτήσεις για να παράγει μία τρίτη συνάρτηση και μαθηματικά μπορεί να εκφραστεί ως εξής:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n), \quad (3.5)$$

όπου $S(i, j)$ είναι η τιμή εξόδου της συνέλιξης στη θέση (i, j) , I είναι η εισαγόμενη εικόνα ή feature map, K είναι το φίλτρο ή ο πυρήνας της συνέλιξης και m, n είναι οι δείκτες που διατρέχουν το φίλτρο/πυρήνα.

Αυτή η διαδικασία επιτυγχάνεται μέσω της εφαρμογής ενός φίλτρου ή πυρήνα(kernel) στα δεδομένα εισόδου, όπως μία εικόνα, παράγοντας έτσι έναν χάρτη χαρακτηριστικών(feature map). Η είσοδος σε μία συνέλιξη, μπορεί να είναι ακατέργαστα δεδομένα εισόδου ή ένας χάρτης χαρακτηριστικών από μία προηγούμενη συνέλιξη. Η λειτουργία της συνέλιξης εκτελείται με την ολίσθηση του πυρήνα πάνω από την εικόνα εισόδου. Σε κάθε βήμα ο πυρήνας πολλαπλασιάζεται στοιχείο προς στοιχείο με το αντίστοιχο τμήμα της εικόνας, δημιουργώντας έτσι μία μοναδική εγγραφή στον χάρτη χαρακτηριστικών. Η διαδικασία αυτή επαναλαμβάνεται, καθώς ο πυρήνας κινείται, μέχρι να καλυφθεί ολόκληρη η επιφάνεια της εικόνας και να δημιουργηθεί ο πλήρης χάρτης χαρακτηριστικών. Ο πυρήνας έχει συνήθως μικρότερες διαστάσεις από την εικόνα εισόδου και περιέχει βάρη, τα οποία εκπαιδεύονται μέσω διάφορων αλγορίθμων, όπως η μέθοδος gradient descent, επιτρέποντας έτσι την ανίχνευση συγκεκριμένων χαρακτηριστικών όπως άκρες, γωνίες ή άλλα οπτικά μοτίβα. Κάθε φίλτρο ή πυρήνας, στο επίπεδο συνέλιξης παράγει έναν χάρτη ενεργοποίησης, ο οποίος

αποτυπώνει πως τα διάφορα χαρακτηριστικά των δεδομένων εισόδου ανταποκρίνονται σε εκείνο το συγκεκριμένο φίλτρο. Όταν χρησιμοποιούνται πολλαπλά φίλτρα οι χάρτες ενεργοποίησης, που προκύπτουν από κάθε ένα φίλτρο, στοιβάζονται μαζί κατά μήκος της διάστασης βάθους. Αυτή η διαδικασία παράγει την τρισδιάστατη έξοδο του επιπέδου συνέλιξης. Το πλάτος και το ύψος του χάρτη χαρακτηριστικών εξαρτώνται από τις διαστάσεις των δεδομένων εισόδου, το μέγεθος του πυρήνα συνέλιξης(kernel size), το βήμα μετατόπισης(stride) και την ποσότητα της περιθωριοποίησης(padding) που εφαρμόζεται στα δεδομένα εισόδου. Η διάσταση του βάθους εξαρτάται από τον αριθμό των φίλτρων που χρησιμοποιούνται στο επίπεδο συνέλιξης.



Σχήμα 3.1: Παράδειγμα συνέλιξης μίας εικόνας διαστάσεων 3x3 με ένα φίλτρο διαστάσεων 2x2 που δίνει σαν αποτέλεσμα έναν feature map διαστάσεων 2x2.

Πηγή: Deep Learning: Foundations and Concepts

Υπερπαράμετροι Συνελικτικού Στρώματος

Οι υπερπαράμετροι ενός συνελικτικού στρώματος, που καθορίζουν την χωρική διάταξη και το μέγεθος της εξόδου από αυτό είναι οι εξής:

Μέγεθος φίλτρου(ή πυρήνα)

Το μέγεθος του φίλτρου αναφέρεται στο πλάτος και το ύψος του σε pixel. Τα πιο συνηθισμένα μεγέθη φίλτρων είναι 3x3, 5x5 ή και μεγαλύτερα. Το βάθος του φίλτρου, είναι αντίστοιχο του βάθους της εικόνας εισόδου. Για παράδειγμα σε μία εικόνα RGB, που έχει τρία χρωματικά κανάλια, το βάθος του φίλτρου θα είναι επίσης τρία. Η επιλογή του μεγέθους του φίλτρου εξαρτάται από το είδος των χαρακτηριστικών που πρέπει να αναγνωρίσει το δίκτυο. Μικρότερα φίλτρα είναι πιο αποδοτικά στην ανίχνευση λεπτομερειών, ενώ μεγαλύτερα στην ανίχνευση γενικών χαρακτηριστικών. Μεγαλύτερα φίλτρα καλύπτουν, μεγαλύτερη περιοχή της εικόνας εισόδου σε κάθε βήμα της συνέλιξης συγκριτικά με μικρότερα φίλτρα. Έτσι μεγαλύτερα φίλτρα μπορεί να μειώσουν το μέγεθος της εξόδου. Σημειώνεται ότι η επίδραση του μεγέθους φίλτρου μπορεί να αντισταθμιστεί από τις υπόλοιπες υπερπαραμέτρους που θα αναλυθούν στη συνέχεια.

Βάθος Εξόδου(Output Depth)

Μπορούμε να επιλέξουμε χειροκίνητα το βάθος της εξόδου. Η υπερπαράμετρος του βάθους ελέγχει τον αριθμό των νευρώνων στο συνελικτικό στρώμα που είναι συνδεδεμένοι με την ίδια περιοχή της εικόνας εισόδου. Θεωρούμε ως μία στήλη βάθους, ένα σύνολο νευρώνων που όλοι εστιάζουν στην ίδια περιοχή της εικόνας εισόδου.

Βήμα(Stride)

Το βήμα ρυθμίζει πόσο μακριά μετακινείται το παράθυρο του φίλτρου, σε κάθε εφαρμογή της συνάρτησης φίλτρου. Κάθε φορά που το φίλτρο εφαρμόζεται στην εικόνα εισόδου, δημιουργείται μία νέα στήλη βάθους στην έξοδο του συνελκτικού στρώματος. Μικρότερες τιμές βήματος, όπως για παράδειγμα το 1, δηλώνουν ότι το φίλτρο μετακινείται με πιο λεπτομερή βήματα πάνω στην εικόνα εισόδου, γεγονός που οδηγεί στη δημιουργία περισσότερων στηλών βάθους στην έξοδο. Με άλλα λόγια, το φίλτρο κινείται πιο αργά πάνω στην εικόνα εισόδου, εξετάζοντας περισσότερες περιοχές, με μεγαλύτερη επικάλυψη μεταξύ τους.

Μηδενική Συμπλήρωση(Zero-padding)

Η μηδενική συμπλήρωση αναφέρεται στην προσθήκη μηδενικών γύρω από τις άκρες της εικόνας εισόδου. Αυτό γίνεται για να διατηρηθούν οι αρχικές διαστάσεις της εικόνας και στην έξοδο του συνελκτικού στρώματος. Εκτός από την διατήρηση των αρχικών διαστάσεων η μηδενική συμπλήρωση χρησιμοποιείται σε ορισμένες περιπτώσεις και για τον έλεγχο της μείωσης των διαστάσεων. Η μηδενική συμπλήρωση μπορεί να επηρεάσει την εκπαίδευση του δικτύου, καθώς οι άκρες της εικόνας εισόδου λαμβάνονται περισσότερο υπόψιν στην επεξεργασία. Αυτό μπορεί να βοηθήσει στην καλύτερη αναγνώριση χαρακτηριστικών και μοτίβων που βρίσκονται στις άκρες των εικόνων.

Στρώματα Ενεργοποίησης ReLu

Τα στρώματα ενεργοποίησης *ReLU* εφαρμόζονται συνήθως μετά από κάθε επίπεδο συνέλιξης. Ένα στρώμα *ReLU*, εφαρμόζει τη συνάρτηση $\max(0, x)$ σε κάθε εισερχόμενη τιμή ξεχωριστά, έτσι αν η τιμή είναι θετική, παραμένει ως έχει ενώ αν είναι αρνητική αντικαθίσταται από το μηδέν. Η εφαρμογή της συνάρτησης ενεργοποίησης *ReLU* στα δεδομένα εισόδου, αλλάζει τις τιμές των pixel, αλλά δεν αλλάζει τις χωρικές διαστάσεις των δεδομένων. Αυτό σημαίνει ότι η έξοδος που θα προκύψει από το στρώμα *ReLU* θα έχει τις ίδιες διαστάσεις με την είσοδο. Τα στρώματα *ReLU* δεν έχουν δικές τους παραμέτρους ή υπερπαραμέτρους προς εκπαίδευση, γεγονός που τα κάνει απλά και αποτελεσματικά.

Στρώματα Υποδειγματοληψίας(Pooling Layers)

Ένας περιορισμός των χαρτών χαρακτηριστικών που παράγονται από τα συνελκτικά στρώματα αποτελεί το γεγονός πως καταγράφουν την ακριβή θέση των χαρακτηριστικών κατά την είσοδο. Αυτό σημαίνει ότι μικρές μετακινήσεις στη θέση του χαρακτηριστικού στην εικόνα εισόδου θα οδηγήσουν σε διαφορετικό χάρτη χαρακτηριστικών. Μικρές μετακινήσεις των χαρακτηριστικών της εικόνας εισόδου μπορεί να προκληθούν από επαναπροσδιορισμό των ορίων, περιστροφή, μετατόπιση και άλλες μικρές αλλαγές. Μία συχνή προσέγγιση για την αντιμετώπιση αυτού του προβλήματος ονομάζεται υποδειγματοληψία(downsampling), κατά την οποία δημιουργείται μία έκδοση χαμηλότερης ανάλυσης της εισερχόμενης εικόνας που διατηρεί όμως ακόμα τα σημαντικά δομικά στοιχεία. Η υποδειγματοληψία μπορεί να επιτευχθεί με συνελκτικά στρώματα, αλλάζοντας το βήμα συνέλιξης πάνω στην

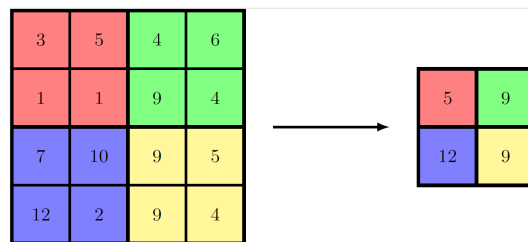
εικόνα, μία πιο ανθεκτική και κοινή προσέγγιση όμως είναι η χρήση ενός στρώματος pooling. Τα στρώματα pooling εισάγονται συνήθως μεταξύ διαδοχικών συνελκτικών στρωμάτων. Στα pooling στρώματα υλοποιείται η διαδικασία pooling, που αποτελεί τη βασική διαδικασία που υλοποιείται στα στρώματα αυτά και επιτυγχάνει τόσο τη μείωση της ποσότητας των δεδομένων που το δίκτυο θα επεξεργαστεί, όσο και τη μείωση του κινδύνου υπερπροσαρμογής.

Pooling

Το pooling είναι μία συγκεκριμένη μέθοδος υποδειγματοληψίας (downsampling) που εφαρμόζεται στα συνελκτικά νευρωνικά δίκτυα. Στο pooling, ένα παράθυρο (φίλτρο) σαρώνει την εικόνα εισόδου και εφαρμόζει μία συγκεκριμένη λειτουργία, όπως το max pooling ή το average pooling, σε κάθε υποπεριοχή. Αυτό επιτρέπει τη μείωση των χωρικών διαστάσεων (ύψος και πλάτος) της εικόνας εισόδου. Η μέθοδος pooling διαφέρει από άλλες μεθόδους downsampling, αφού εφαρμόζει μία σταθερή λειτουργία σε κάθε υποπεριοχή. Υπάρχουν δύο βασικοί τύποι pooling:

1. Max Pooling

Επιλέγεται ένα παράθυρο σάρωσης με συγκεκριμένο μέγεθος, συνήθως 2x2 ή 3x3. Το παράθυρο αυτό σαρώνει την εικόνα εισόδου, κινούμενο κάθετα και οριζόντια με βήμα που καθορίζεται από την παράμετρο stride. Σε κάθε θέση, η λειτουργία max pooling επιλέγει τη μέγιστη τιμή, από τις τιμές εντός του παραθύρου.



Σχήμα 3.2: Απεικόνιση της διαδικασίας Max Pooling, στην οποία σε κάθε μπλοκ μεγέθους 2x2 ενός feature map επιλέγεται η μέγιστη τιμή για την παραγωγή ενός νέου feature map μικρότερων διαστάσεων.

Πηγή: Deep Learning: Foundations and Concepts

2. Average Pooling

Όπως ακριβώς συμβαίνει στη λειτουργία max pooling, έτσι και κατά το average pooling επιλέγεται ένα παράθυρο σάρωσης συγκεκριμένου μεγέθους, το οποίο κινείται καλύπτοντας τις διαδοχικές υποπεριοχές της εικόνας εισόδου, με βήμα που έχει καθοριστεί από την παράμετρο stride. Σε κάθε θέση το average pooling υπολογίζει το μέσο όρο των τιμών εντός του παραθύρου. Συγκριτικά με το max pooling, το average pooling λαμβάνει υπόψιν το σύνολο των τιμών εντός του παραθύρου και όχι μόνο τη μέγιστη από αυτές. Έτσι, καθίσταται καταλληλότερο για εφαρμογές στις οποίες είναι σημαντικό να διατηρηθεί η πληροφορία από όλη την εικόνα εισόδου και όχι μόνο από τα πιο έντονα χαρακτηριστικά. Σε γενικές γραμμές το average pooling παρέχει μία πιο ομαλή αναπαράσταση των δεδομένων και είναι λιγότερο ευαίσθητο στην ύπαρξη ακραίων τιμών σε σχέση με το max pooling.

3.1.3.3 Πλήρως Συνδεδεμένα Στρώματα(Fully Connected Layers)

Τα πλήρως συνδεδεμένα στρώματα, γνωστά και ως πυκνά στρώματα(dense layers) αποτελούν έναν τύπο στρώματος στον οποίο κάθε νευρώνας από το προηγούμενο στρώμα συνδέεται με κάθε νευρώνα του τρέχοντος στρώματος. Αυτό σημαίνει ότι οι πληροφορίες από όλους τους νευρώνες του προηγούμενου στρώματος(π.χ ένα συνελκτικό ή ένα pooling στρώμα) μεταφέρονται σε κάθε νευρώνα του πλήρως συνδεδεμένου στρώματος. Ονομάζεται πλήρως συνδεδεμένο στρώμα, λοιπόν, λόγω αυτής της πλήρους σύνδεσης. Τα στρώματα αυτά βρίσκονται συνήθως προς το τέλος μιας αρχιτεκτονικής ενός συνελκτικού νευρωνικού δικτύου, ακολουθούν δηλαδή τα συνελκτικά και τα pooling στρώματα. Πριν την εισαγωγή στα πλήρως συνδεδεμένα στρώματα, οι χάρτες χαρακτηριστικών μετασχηματίζονται σε ένα μονοδιάστατο διάνυσμα. Εξαιτίας του μεγάλου αριθμού συνδέσεων, τα πλήρως συνδεδεμένα στρώματα περιέχουν συνήθως τον μεγαλύτερο αριθμό παραμέτρων προς εκπαίδευση σε ένα συνελκτικό νευρωνικό δίκτυο, γεγονός που μπορεί να αυξήσει την πολυπλοκότητα και τον χρόνο εκπαίδευσης του δικτύου. Ο αριθμός των νευρώνων στο τελευταίο πλήρως συνδεδεμένο στρώμα συνήθως αντικατοπτρίζει τον αριθμό των κλάσεων σε ένα πρόβλημα ταξινόμησης. Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης εικόνων σε 10 κλάσεις, θα υπήρχαν 10 νευρώνες στο τελευταίο πλήρως συνδεδεμένο στρώμα.

3.2 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης αποτελούν βαθμωτές συναρτήσεις που παράγουν την έξοδο των νευρώνων ενός επιπέδου προς το επόμενο επίπεδο. Μία συνάρτηση ενεργοποίησης αποφασίζει ουσιαστικά αν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι, μετατρέποντας την είσοδο(το άθροισμα των επιμέρους εισόδων επί τα αντίστοιχα βάρη συν το bias) σε μία έξοδο, συχνά σε ένα εύρος τιμών μεταξύ 0 και 1 ή -1 και 1. Οι περισσότερες συναρτήσεις ενεργοποίησης είναι μη γραμμικές, γεγονός που επιτρέπει στα νευρωνικά δίκτυα να μαθαίνουν και να μοντελοποιούν περίπλοκες σχέσεις μεταξύ των δεδομένων, που θα ήταν αδύνατο με απλές γραμμικές συναρτήσεις. Ακολουθούν μερικές από τις πιο συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης.

3.2.1 Σιγμοειδής Συνάρτηση Ενεργοποίησης(Sigmoid Activation Function)

Η σιγμοειδής συνάρτηση ενεργοποίησης, γνωστή και ως λογιστική συνάρτηση είναι μία από τις πιο γνωστές μη γραμμικές συναρτήσεις ενεργοποίησης. Η βασική της μορφή είναι:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (3.6)$$

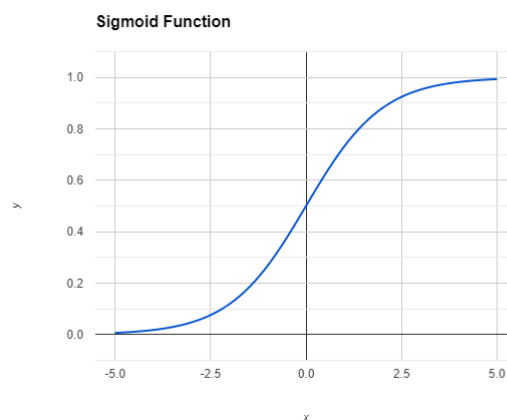
όπου, $\sigma(x)$ είναι η τιμή εξόδου της συνάρτησης και x η είσοδος στη συνάρτηση.

Όπως όλοι οι λογιστικοί μετασχηματισμοί, έτσι και η σιγμοειδής συνάρτηση έχει

την ικανότητα να μειώνει τις ακραίες τιμές χωρίς να τις απομακρύνει. Μία σιγμοειδής συνάρτηση μετατρέπει τις τιμές εισόδου, που έχουν σχεδόν άπειρο εύρος, σε απλές πιθανότητες με τιμές μεταξύ του 0 και του 1, για αυτό και χρησιμοποιείται κατά κύριο λόγο σε προβλήματα δυαδικής ταξινόμησης και προβλήματα που απαιτούν την πρόβλεψη πιθανότητας ως έξοδο. Σε ένα πρόβλημα ταξινόμησης παράγει μία ανεξάρτητη πιθανότητα για κάθε κλάση. Όταν απεικονίζεται γραφικά η συνάρτηση έχει σχήμα “S” από το οποίο προέρχεται και το όνομά της. Η συνάρτηση είναι διαφορίσιμη και κατά συνέπεια μπορούμε να βρούμε την κλίση της καμπύλης σε δύο οποιαδήποτε σημεία. Λόγω της διαφορισιμότητάς της, είναι συμβατή με τη μέθοδο backpropagation που απαιτεί την ύπαρξη παραγώγων των συναρτήσεων ενεργοποίησης για την εκπαίδευση του δικτύου. Επιπλέον το γεγονός πως η παράγωγος της συνάρτησης μπορεί να εκφραστεί μέσω της ίδιας ως:

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x)), \quad (3.7)$$

καθιστά τον υπολογισμό της κλίσης σχετικά απλό και βολικό. Στις περισσότερες περιπτώσεις, έχουμε ήδη υπολογίσει το $\sigma(x)$, κατά το forward propagation και μπορούμε να εφαρμόσουμε τη μέθοδο gradient descent χωρίς υψηλό υπολογιστικό κόστος. Η σιγμοειδής συνάρτηση έχει έντονη κλίση στο κέντρο της, δηλαδή στην περιοχή γύρω από το 0 και μικρή κλίση στα άκρα της. Το γεγονός αυτό συχνά οδηγεί στο φαινόμενο των «κορεσμένων κλίσεων», όπου οι αλλαγές στα βάρη του δικτύου έχουν μικρή ή καμία επίδραση στην έξοδο, καθιστώντας την εκπαίδευση του αναποτελεσματική ή ακόμη και σταματώντας την εντελώς. Επιπλέον η έλλειψη συμμετρίας της σιγμοειδούς συνάρτησης γύρω από το 0, μπορεί να δημιουργήσει προβλήματα κατά την εκπαίδευση, αφού η απουσία δεδομένων που κεντράρονται γύρω από το 0 μπορεί να οδηγήσει στο φαινόμενο της «μεροληψίας προς θετικές τιμές» στα βάρη του δικτύου. Για τον λόγο αυτό, είναι απαραίτητη η τυποποίηση των δεδομένων πριν τροφοδοτηθούν στο νευρωνικό δίκτυο. Οι παραπάνω περιορισμοί μπορεί να προκαλέσουν αργή σύγκλιση του δικτύου και συνεπώς η εκπαίδευση του με σιγμοειδή συνάρτηση να είναι αρκετά πιο αργή από ότι αν επιλεγεί κάποια άλλη συνάρτηση ενεργοποίησης, όπως για παράδειγμα η *ReLU* (Rectified Linear Unit).



Σχήμα 3.3: Η Σιγμοειδής Συνάρτηση Ενεργοποίησης

3.2.2 Συνάρτηση Ενεργοποίησης Υπερβολικής Εφαπτομένης (Tanh Activation Function)

Η συνάρτηση ενεργοποίησης \tanh είναι μία υπερβολική τριγωνομετρική συνάρτηση. Όπως η εφαπτομένη αντιπροσωπεύει την αναλογία της απέναντι και της προσκείμενης πλευράς, μιας οξείας γωνίας ενός ορθογωνίου τριγώνου έτσι και η συνάρτηση ενεργοποίησης υπερβολικής εφαπτομένης εκφράζει την αναλογία μεταξύ του υπερβολικού ημιτόνου και του υπερβολικού συνημιτόνου. Μαθηματικά εκφράζεται ως εξής:

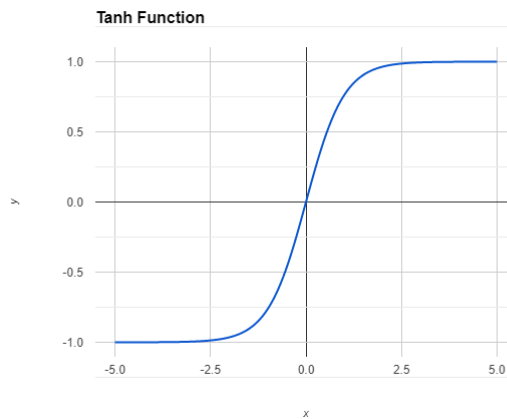
$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}, \quad (3.8)$$

όπου:

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

Έχει εύρος τιμών από το -1 έως το 1, που σημαίνει ότι για μεγάλες θετικές ή αρνητικές τιμές του x , η συνάρτηση τείνει προς τις τιμές 1 και -1 αντίστοιχα. Ένα από τα βασικά πλεονεκτήματα της συνάρτησης \tanh , σε σχέση με άλλες συναρτήσεις ενεργοποίησης είναι η ικανότητα της να χειρίζεται αποτελεσματικά τις αρνητικές τιμές. Επειδή το εύρος της \tanh περιλαμβάνει αρνητικές τιμές, αποτελεί καταλληλότερη επιλογή για περιπτώσεις όπου οι τιμές ενεργοποίησης πρέπει να είναι συμμετρικές γύρω από το 0. Σημειώνεται πως, όπως η σιγμοειδής συνάρτηση, έτσι και η συνάρτηση υπερβολικής εφαπτομένης μπορεί να οδηγήσει στο φαινόμενο των «κορεσμένων κλίσεων».

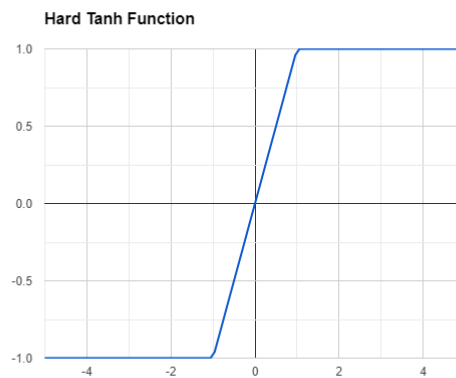


Σχήμα 3.4: Η Συνάρτηση Ενεργοποίησης $Tanh$

3.2.3 Συνάρτηση Ενεργοποίησης Hard Tanh

Η συνάρτηση ενεργοποίησης $hard\ tanh$ αποτελεί μία παραλλαγή της συνάρτησης ενεργοποίησης υπερβολικής εφαπτομένης \tanh . Όπως η \tanh έτσι και η $hard\ tanh$ παράγει εξόδους σε ένα κανονικοποιημένο εύρος, αλλά με έναν πιο αυστηρό

τρόπο, εφαρμόζοντας δηλαδή σταθερά όρια στις τιμές εξόδου. Η συνάρτηση ενεργοποίησης *hard tanh* θέτει οποιαδήποτε τιμή πάνω από 1 ίση με 1, οποιαδήποτε τιμή κάτω από -1 ίση με -1 και για όλες τις άλλες τιμές εισόδου επιστρέφει την ίδια την τιμή εισόδου. Λόγω των σταθερών ορίων παρέχει ένα πιο συγκεκριμένο και περιορισμένο όριο απόφασης συγκριτικά με την υπερβολική εφαπτομένη, έτσι χρησιμοποιείται συχνά σε νευρωνικά δίκτυα που απαιτούν αυστηρότερο έλεγχο των τιμών εξόδων των νευρώνων. Λόγω της απλής φύσης της, μπορεί να υπολογιστεί πιο αποδοτικά σε σχέση με πιο περίπλοκες συναρτήσεις ενεργοποίησης παρόλα αυτά μπορεί να οδηγήσει σε περιορισμένη εκμάθηση.



Σχήμα 3.5: Η Συνάρτηση Ενεργοποίησης *Hard Tanh*

3.2.4 Συνάρτηση Ενεργοποίησης *Softmax*

Η συνάρτηση ενεργοποίησης *Softmax* αποτελεί μία γενίκευση της λογιστικής παλινδρόμησης, καθώς μπορεί να εφαρμοστεί σε συνεχή δεδομένα και να περιλαμβάνει πολλαπλά όρια απόφασης. Είναι ιδιαίτερα χρήσιμη σε προβλήματα ταξινόμησης πολλαπλών κλάσεων, όπου ένα δείγμα ανήκει σε μία από τις πολλές διαφορετικές κατηγορίες. Μαθηματικά, η συνάρτηση ενεργοποίησης *Softmax* εκφράζεται ως εξής:

$$f(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}, \quad (3.9)$$

όπου, e είναι η βάση του φυσικού λογαρίθμου και z_i είναι η λογική τιμή, δηλαδή η αρχική πρόβλεψη του δικτύου πριν εφαρμοστεί η συνάρτηση ενεργοποίησης, για την i -οστή κλάση.

Η *Softmax* λειτουργεί αρχικά κανονικοποιώντας το διάνυσμα εισόδου, ώστε όλοι οι αριθμοί στο διάνυσμα να αθροίζονται στο 1. Στη συνέχεια, εκθετοποιεί κάθε αριθμό και τον διαιρεί με το άθροισμα όλων των εκθετοποιημένων αριθμών. Αυτό οδηγεί σε ένα διάνυσμα πιθανοτήτων. Κάθε τιμή εξόδου αποτελεί την εκτιμώμενη πιθανότητα του αντίστοιχου δείγματος να ανήκει σε μία συγκεκριμένη κατηγορία. Η συνάρτηση ενεργοποίησης *Softmax* συναντάται κατά κύριο λόγο στο επίπεδο εξόδου

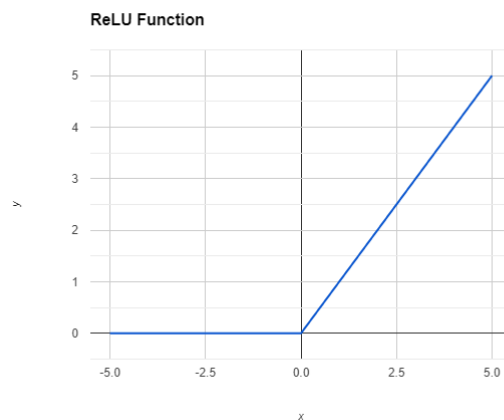
ενός ταξινομητή. Σε περιπτώσεις με πολύ μεγάλο αριθμό κατηγοριών χρησιμοποιείται μία παραλλαγή της *Softmax*, γνωστή ως ιεραρχική συνάρτηση ενεργοποίησης *Softmax*. Η παραλλαγή αυτή, χρησιμοποιεί μία δομή δέντρου για να απλοποιήσει τον υπολογισμό των πιθανοτήτων.

3.2.5 Συνάρτηση Ενεργοποίησης ReLu(Rectified Linear Unit)

Η συνάρτηση ενεργοποίησης *ReLU* αποτελεί μία από τις πιο διαδεδομένες και ευρέως χρησιμοποιούμενες συναρτήσεις στον τομέα της μηχανικής μάθησης και των νευρωνικών δικτύων. Μαθηματικά ορίζεται ως εξής:

$$g(x) = \max(0, x), \quad (3.10)$$

Αυτό σημαίνει ότι για οποιαδήποτε θετική τιμή εισόδου x , η συνάρτηση επιστρέφει την ίδια την τιμή, ενώ για μηδενικές ή αρνητικές τιμές εισόδου επιστρέφει το 0. Η συνάρτηση ενεργοποίησης *ReLU* αποτελεί μία κατά τμήματα γραμμική συνάρτηση, γεγονός που επιτρέπει στο νευρωνικό δίκτυο να διατηρεί πολλές από τις ιδιότητες που κάνουν τα γραμμικά μοντέλα εύκολα στην βελτιστοποίηση, με μεθόδους βασισμένες στην κλίση, και ικανά να γενικεύουν καλά. Επιπλέον, είναι απλή στον υπολογισμό της, γεγονός που καθιστά τα νευρωνικά δίκτυα πιο αποδοτικά σε όρους χρόνου και πόρων. Η ικανότητα της να παράγει μηδενικές τιμές, σε σύγκριση με τη σιγμοειδή συνάρτηση και τη συνάρτηση ενεργοποίησης υπερβολικής εφάπτομένης, που παράγουν τιμές πολύ κοντά στο 0, μπορεί να οδηγήσει σε απλούστερα μοντέλα που συγκλίνουν γρηγορότερα. Ένα μειονέκτημα της συνάρτησης ενεργοποίησης *ReLU* είναι πως ορισμένοι νευρώνες μπορεί να σταματήσουν να ανταποκρίνονται σε ερεθίσματα, εξαιτίας αρνητικών τιμών εισόδων, και να μην ενεργοποιηθούν ξανά κατά τη διάρκεια της εκπαίδευσης.



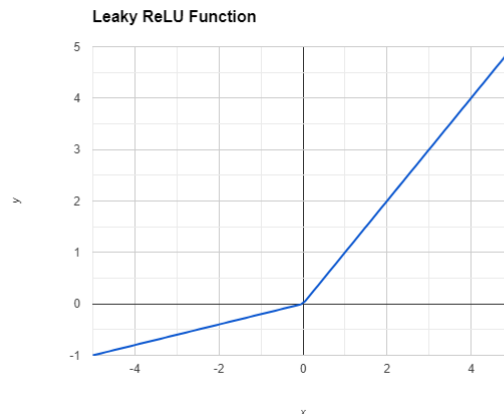
Σχήμα 3.6: Η Συνάρτηση Ενεργοποίησης *ReLU*

3.2.6 Συνάρτηση Ενεργοποίησης *Leaky ReLu*(*Leaky Rectified Linear Unit*)

Η συνάρτηση ενεργοποίησης *Leaky ReLu* αποτελεί μία παραλλαγή της συνάρτησης *ReLu*. Η συνάρτηση ενεργοποίησης *ReLu*, όταν στο δίκτυο εισέρχονται αρνητικές ή μηδενικές τιμές εισόδου επιστρέφει 0, κάτι που συχνά οδηγεί σε νευρώνες που δεν ενεργοποιούνται ποτέ κατά τη διάρκεια της εκπαίδευσης. Η *Leaky ReLu*, για να αντιμετωπίσει το πρόβλημα των «νεκρών νευρώνων», πολλαπλασιάζει αυτές τις αρνητικές τιμές εισόδου με έναν μικρό σταθερό συντελεστή, συνήθως το 0.01, επιτρέποντας έτσι τη διέλευση ενός μικρού ποσοστού της αρχικής τιμής αυτών. Με τον τρόπο αυτό διατηρεί μία μικρή, αλλά μη μηδενική κλίση, που επιτρέπει στα βάρη να συνεχίσουν να ενημερώνονται κατά την εκπαίδευση, ακόμη και αν οι είσοδοι είναι αρνητικές. Μαθηματικά η συνάρτηση ενεργοποίησης *Leaky ReLu* μπορεί να εκφραστεί ως εξής:

$$f(x) = \begin{cases} x & ,\text{αν } x > 0 \\ \alpha x & ,\text{αν } x \leq 0 \end{cases} \quad (3.11)$$

Όπου α ένας μικρός σταθερός αριθμός, όπως για παράδειγμα 0.01.



Σχήμα 3.7: Η Συνάρτηση Ενεργοποίησης *Leaky ReLu*

3.2.7 Συνάρτηση Ενεργοποίησης *Softplus*

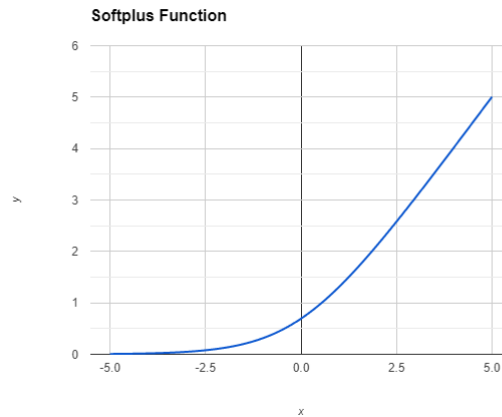
Η συνάρτηση ενεργοποίησης *Softplus* αποτελεί και αυτή μία παραλλαγή της συνάρτησης ενεργοποίησης *ReLu* και πιο συγκεκριμένα μπορεί να θεωρηθεί ως μία ομαλοποιημένη εκδοχή της. Μαθηματικά, η συνάρτηση ενεργοποίησης *Softplus* ορίζεται ως εξής:

$$f(x) = \ln(1 + e^x), \quad (3.12)$$

όπου \ln ο φυσικός αλγόριθμος και e η βάση του φυσικού αλγορίθμου.

Η συνάρτηση *Softplus* παράγει μία ομαλή μη γραμμική έξοδο που είναι πάντα ένας

θετικός αριθμός. Όταν οι τιμές εισόδου είναι μεγάλοι θετικοί αριθμοί, η *Softplus* προσεγγίζει τη γραμμική συμπεριφορά, ενώ για αρνητικές τιμές η έξοδος προσεγγίζει το μηδέν. Με τον τρόπο αυτό, δεν επιστρέφει άμεσα το μηδέν για αρνητικές τιμές εισόδου όπως η *ReLU*, αλλά προσφέρει μία πιο ομαλή μετάβαση προς αυτό. Το βασικότερο πλεονέκτημα της *Softplus* είναι πως η παράγωγός της είναι πάντα θετική, γεγονός που βοηθά στην αποφυγή του προβλήματος της εξαφάνισης της κλίσης κατά την εκπαίδευση του νευρωνικού δικτύου.



Σχήμα 3.8: Η Συνάρτηση Ενεργοποίησης *Softplus*

3.3 Συναρτήσεις Απώλειας

Η συνάρτηση απώλειας είναι η συνάρτηση που καθορίζει πόσο μακριά βρίσκεται η προβλεπόμενη από την πραγματική έξοδο. Στην περίπτωση της ταξινόμησης μετρά πόσο μακριά βρίσκεται η προβλεπόμενη από την πραγματική κλάση. Η συνάρτηση απώλειας χρησιμοποιείται κατά την διαδικασία της εκπαίδευσης ενός νευρωνικού δικτύου για να κατανοήσει το μοντέλο πόσο καλά αποδίδει και να προσαρμόσει τις διάφορες παραμέτρους του (βάρη, biases), με σκοπό να μειώσει το σφάλμα στις προβλέψεις του. Στόχος είναι η ελαχιστοποίηση της τιμής της συνάρτησης απώλειας, καθώς μικρές τιμές αυτής δηλώνουν καλή απόδοση του δικτύου και υψηλή προβλεπτική ικανότητα. Σημειώνεται πως η συνάρτηση απώλειας αναφέρεται στο σφάλμα για ένα μόνο παράδειγμα, ενώ η συνάρτηση κόστους (Cost function) αποτελεί το μέσο όρο των συναρτήσεων απώλειας για όλα τα παραδείγματα εκπαίδευσης. Ακολουθούν οι πιο δημοφιλείς συναρτήσεις απώλειας για προβλήματα ταξινόμησης.

3.3.1 Binary Cross-Entropy Loss Function-Log Loss

Η *Binary Cross – Entropy Loss* γνωστή και ως *Log Loss*, αποτελεί ένα δείκτη απόδοσης για μοντέλα ταξινόμησης που παράγουν προβλέψεις με μία τιμή

πιθανότητας, η οποία συνήθως κυμαίνεται μεταξύ 0 και 1. Αυτή η τιμή πρόβλεψης αντιστοιχεί στην πιθανότητα ένα δείγμα δεδομένων να ανήκει σε μία κατηγορία(κλάση). Στην περίπτωση της *Binary Cross – Entropy Loss* υπάρχουν δύο διακριτές κατηγορίες, συνήθως αντιστοιχισμένες στις ετικέτες 0 και 1, και συνεπώς η χρήση της είναι ευρέως διαδεδομένη σε νευρωνικά δίκτυα που απαιτούν δυαδική ταξινόμηση. Για την καλύτερη κατανόηση της *Binary Cross – Entropy Loss* ακολουθεί η επεξήγηση ορισμένων εννοιών:

- Απώλεια(Loss): είναι η ποσοτικοποίηση της διαφοράς μεταξύ της πρόβλεψης ενός αλγορίθμου μηχανικής μάθησης και της πραγματικής τιμής.
- Διασταυρούμενη Εντροπία(*Cross – Entropy*): μετρά πόσο διαφορετικές είναι δύο κατανομές πιθανοτήτων. Στην περίπτωση της μηχανικής μάθησης και της ταξινόμησης, συγκρίνουμε την κατανομή πιθανοτήτων που παράγεται από το νευρωνικό δίκτυο με την πραγματική τιμή των δεδομένων(ή τις πραγματικές ετικέτες των δεδομένων). Η Διασταυρούμενη Εντροπία μεταξύ δύο κατανομών πιθανοτήτων P και Q για ένα σύνολο δεδομένων X μπορεί να υπολογιστεί ως εξής:

$$H(P, Q) = - \sum_{x \in X} P(x) \log(Q(x)), \quad (3.13)$$

όπου P είναι η πραγματική κατανομή(συνήθως οι πραγματικές ετικέτες των Δεδομένων) και Q είναι η κατανομή πιθανοτήτων που έχει προβλέψει το νευρωνικό δίκτυο.

- Δυαδικός(Binary): αναφέρεται στην δυαδική ταξινόμηση, όπου διακρίνουμε δύο κατηγορίες Α και Β και χρησιμοποιώντας τη δυαδική αναπαράσταση η κατηγορία Α αντιστοιχίζεται με τον αριθμό 0 και η κατηγορία Β με τον αριθμό 1.

Η συνάρτηση απώλειας *Binary Cross – Entropy*, μαθηματικά εκφράζεται ως εξής:

$$L(y, f(x)) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i))], \quad (3.14)$$

όπου L είναι η συνάρτηση απώλειας *Binary Cross – Entropy Loss*, N είναι ο συνολικός αριθμός παραδειγμάτων, y_i η πραγματική δυαδική ετικέτα του δείγματος i (0 ή 1) και $f(x_i)$ η προβλεπόμενη από το νευρωνικό δίκτυο πιθανότητα για το δείγμα i .

Η συνάρτηση απώλειας *Binary Cross – Entropy Loss* «τιμωρεί» περισσότερο τις πολύ ανακριβείς προβλέψεις, δηλαδή αυτές που έχει προβλεφθεί πολύ υψηλή πιθανότητα για μία κατηγορία που δεν είναι σωστή, σε σχέση με τις λιγότερο ανακριβείς προβλέψεις. Με τον τρόπο αυτό βελτιώνεται η ακρίβεια του μοντέλου.

3.3.2 Categorical Cross-Entropy Loss Function

Η συνάρτηση απώλειας *Categorical Cross – Entropy* χρησιμοποιείται στην περίπτωση όπου ένα σύνολο δεδομένων πρέπει να ταξινομηθεί σε περισσότερες

από δύο κατηγορίες(κλάσεις). Τα μοντέλα που χρησιμοποιούν την συγκεκριμένη συνάρτηση απώλειας εκτιμούν την πιθανότητα για κάθε κατηγορία και επιλέγουν την κατηγορία με την υψηλότερη πιθανότητα ως την τελική πρόβλεψη. Μαθηματικά η συνάρτηση απώλειας *Categorical Cross – Entropy* μπορεί να εκφραστεί ως εξής:

$$L = - \sum_i y_i \log(p_i), \quad (3.15)$$

Όπου y_i είναι η πραγματική ετικέτα(απαιτούνται one-hot encoded διανύσματα) και p_i η προβλεπόμενη πιθανότητα για την i -οστή κατηγορία.

Όπως η συνάρτηση απώλειας *Binary Cross–Entropy Loss*, έτσι και η *Categorical Cross – Entropy* «τιμωρεί» περισσότερο τις προβλέψεις που απέχουν πολύ από την πραγματική ετικέτα, σε σχέση με αυτές που βρίσκονται πιο κοντά, έτσι επιτυγχάνει τη βελτίωση της απόδοσης του μοντέλου.

3.3.3 Sparse Categorical Cross-Entropy

Η συνάρτηση απώλειας *Sparse Categorical Cross – Entropy* αποτελεί μία προέκταση της συνάρτησης απώλειας *Categorical Cross – Entropy* και χρησιμοποιείται όταν οι ετικέτες εξόδου βρίσκονται σε μορφή αραιού πίνακα(sparse matrix format). Στη μορφή αυτή, οι ετικέτες αντιστοιχούν σε ένα μοναδικό δεικτοδοτημένο αριθμό, αντί ενός διανύσματος one-hot encoded. Αυτό σημαίνει ότι οι ετικέτες είναι ακέραιοι αριθμοί και όχι διανύσματα. Οι συναρτήσεις απώλειας *Categorical Cross – Entropy* και *Sparse Categorical Cross – Entropy*, έχουν την ίδια συνάρτηση απώλειας, όπως ορίστηκε παραπάνω, με μοναδική διαφορά μεταξύ των δύο τον τρόπο με τον οποίο ορίζονται οι ετικέτες. Η *Sparse Categorical Cross – Entropy* δέχεται ετικέτες σε μορφή ακεραίων αριθμών, για παράδειγμα [1], [2], [3], για ένα πρόβλημα ταξινόμησης σε τρεις κλάσεις. Χρησιμοποιείται συχνά σε προβλήματα με μεγάλο αριθμό κατηγοριών, όπου η μετατροπή των ετικετών σε one-hot encoded μορφή θα ήταν αναποτελεσματική ή ανεφάρμοστη όπως για παράδειγμα σε περιπτώσεις περιορισμένης μνήμης καθώς δεν απαιτεί τη δημιουργία μεγάλων one-hot encoded διανυσμάτων. Όπως ακριβώς οι δύο προηγούμενες συναρτήσεις απώλειας που αναφέρθηκαν, έτσι και η *Sparse Categorical Cross – Entropy* λαμβάνει υπόψιν την πραγματική κατανομή πιθανοτήτων των κλάσεων, γεγονός που διασφαλίζει πως το μοντέλο «τιμωρείται» για λανθασμένη ταξινόμηση των κλάσεων με μειωμένη εκπροσώπηση. Αυτό την καθιστά εφαρμόσιμη σε μη ισορροπημένα σύνολα δεδομένων. Σημειώνεται πως είναι ευαίσθητη σε ακραίες τιμές γεγονός που μπορεί να επηρεάσει την απόδοσή της.

Κεφάλαιο 4

Transfer Learning

Η εκπαίδευση ενός συνελκτικού νευρωνικού δικτύου από την αρχή απαιτεί, τις περισσότερες φορές, μεγάλη ποσότητα δεδομένων με ετικέτα, γεγονός που μπορεί να αποδειχθεί ιδιαίτερα δύσκολο και χρονοβόρο κυρίως σε τομείς όπου τα δεδομένα είναι σπάνια ή ακριβά να συλλεχθούν. Γενικά το Transfer Learning μπορεί να χρησιμοποιηθεί για να βελτιώσει την απόδοση σε κάποια εργασία A, για την οποία τα δεδομένα εκπαίδευσης είναι περιορισμένα, χρησιμοποιώντας δεδομένα από μια συναφή εργασία B, για την οποία τα δεδομένα είναι άφθονα. Οι δύο εργασίες πρέπει να έχουν τον ίδιο τύπο εισόδων και πρέπει να υπάρχει συνάφεια μεταξύ των εργασιών ώστε χαμηλού επιπέδου χαρακτηριστικά, ή εσωτερικές αναπαραστάσεις, που έχουν μάθει από την εργασία B να είναι χρήσιμα για την εργασία A. Τα συνελκτικά νευρωνικά δίκτυα σε πολλές εργασίες επεξεργασίας εικόνων απαιτούν παρόμοια χαμηλού επιπέδου χαρακτηριστικά που αντιστοιχούν στα πρώτα στρώματα ενός βαθιού νευρωνικού δικτύου, ενώ τα επόμενα στρώματα είναι πιο εξειδικευμένα για μια συγκεκριμένη εργασία, κάνοντας τέτοια δίκτυα κατάλληλα για εφαρμογές Transfer Learning. Έτσι, εφαρμόζοντας την τεχνική Transfer Learning αντί ένα συνελκτικό νευρωνικό δίκτυο να εκπαιδεύεται από την αρχή, χρησιμοποιείται ένα προεκπαιδευμένο μοντέλο που έχει εκπαιδευτεί σε ένα μεγάλο και γενικό σύνολο δεδομένων (όπως για παράδειγμα το ImageNet[12]) και στη συνέχεια προσαρμόζεται ή εκπαιδεύεται περαιτέρω ανάλογα με το εκάστοτε σύνολο δεδομένων κάθε φορά. Το Transfer Learning, ουσιαστικά, επιτρέπει στα μοντέλα να μεταφέρουν και να εφαρμόζουν τη γνώση που έχουν ήδη αποκτήσει από ένα πρόβλημα, σε ένα νέο. Η μέθοδος αυτή έχει επιδείξει εξαιρετική απόδοση στην ανάλυση εικόνων, ειδικά στον τομέα της ιατρικής απεικόνισης, όπως στη διάγνωση του καρκίνου του εγκεφάλου[17, 23]. Χρησιμοποιώντας το Transfer Learning τα μοντέλα μηχανικής μάθησης μπορούν να επεξεργαστούν και να αναλύσουν μεγάλα σύνολα ιατρικών εικόνων με αυξημένη ακρίβεια και αποτελεσματικότητα καθώς τα μοντέλα έχουν εκπαιδευτεί ήδη σε σχετικά δεδομένα. Αυτό μειώνει τον χρόνο και τους πόρους που απαιτούνται για την εκπαίδευση και βοηθά στην ταξινόμηση των διάφορων μορφών του καρκίνου του εγκεφάλου με μεγαλύτερη ακρίβεια. Ακολουθούν δύο περιπτώσεις χρήσης της συγκεκριμένης τεχνικής που αξίζει να σημειωθούν:

- Fine Tunning ενός ήδη υπάρχοντος μοντέλου: η διαδικασία επικεντρώνεται στην τροποποίηση και επανεκπαίδευση ορισμένων ή όλων των στρωμάτων ενός νευρωνικού δικτύου για την προσαρμογή του σε ένα νέο πρόβλημα ή

σύνολο δεδομένων.

- Χρήση ενός ήδη υπάρχοντος μοντέλου ως Feature Extractor: κατά τη διαδικασία αυτή τα πρώτα στρώματα του προ εκπαιδευμένου μοντέλου διατηρούνται ως έχουν, αφού ανιχνεύουν τα γενικά χρήσιμα χαρακτηριστικά. Η διαδικασία εστιάζει κυρίως στην τροποποίηση ή την αντικατάσταση των τελευταίων στρωμάτων, έτσι ώστε να ταιριάζουν με τις απαιτήσεις του εκάστοτε προβλήματος ή συνόλου δεδομένων.

Στη συνέχεια, αναλύονται οι αρχιτεκτονικές των προ εκπαιδευμένων μοντέλων που χρησιμοποιήθηκαν ως μοντέλα βάσης στην παρούσα διπλωματική εργασία.

4.1 ResNets/ResNet50

Μία σημαντική καινοτομία στην ανάπτυξη και βελτίωση των νευρωνικών δικτύων είναι η εισαγωγή των Residual Networks(ResNets). Η αρχιτεκτονική αυτή αποτελεί μία σημαντική αλλαγή στον τρόπο κατασκευής βαθιών δικτύων, επιλύοντας το πρόβλημα της μείωσης της απόδοσης, που προκύπτει συνήθως από την αύξηση του βάθους. Κεντρική αρχή πίσω από την αρχιτεκτονική των ResNets αποτελεί η υπολειπόμενη μάθηση(Residual Learning), μία τεχνική που αντιμετωπίζει το προαναφερθέν πρόβλημα. Μέσω της υπολειπόμενης μάθησης, εξετάζεται πώς η αρχιτεκτονική των νευρωνικών δικτύων μπορεί να βελτιωθεί, ώστε να επιτρέψει την αποτελεσματική μάθηση βαθύτερων μοντέλων και πώς η προσέγγιση αυτή συμβάλει στη γενίκευσή τους και στην αποφυγή της υπερποσαρμογής. Στο πλαίσιο, λοιπόν, της υπολειπόμενης μάθησης γίνεται η υπόθεση πως αν τα πολλαπλά μη γραμμικά stacked επίπεδα του δικτύου μπορούν να προσεγγίσουν περίπλοκες συναρτήσεις τότε μπορούν να προσεγγίσουν επίσης και τις συναρτήσεις υπολοίπων. Με βάση την παραπάνω υπόθεση τα stacked layers αντί να εκπαιδεύονται με σκοπό να προσεγγίσουν μία επιθυμητή συνάρτηση, έστω $H(x)$, εκπαιδεύονται με σκοπό να προσεγγίσουν μία άλλη συνάρτηση ,που καλείται συνάρτηση υπολοίπων, την $F(x) := H(x) - x$, τη διαφορά δηλαδή μεταξύ της επιθυμητής εξόδου και της εισόδου. Η αρχική συνάρτηση παίρνει τη μορφή $F(x) + x$. Παρά το γεγονός πως και οι δύο μορφές θα έπρεπε να μπορούν να προσεγγίσουν ασυμπτωτικά τις επιθυμητές συναρτήσεις, η εκπαίδευση του δικτύου καθίσταται ευκολότερη όταν τα stacked layers καλούνται να προσεγγίσουν τη συνάρτηση υπολοίπων ,καθώς πολλές φορές η $H(x) - x$ είναι πιο κοντά στην ταυτοτική συνάρτηση(δηλαδή η $F(x)$ βρίσκεται κοντά στο 0). Η διατύπωση του $F(x) + x$, μπορεί να υλοποιηθεί με feedforward νευρωνικά δίκτυα, με shortcut συνδέσεις .Οι shortcut συνδέσεις είναι εκείνες που παρακάμπτουν ένα ή περισσότερα επίπεδα. Στην περίπτωση των ResNets οι shortcut συνδέσεις απλώς εκτελούν την ταυτοτική συνάρτηση και οι έξοδοί τους προστίθενται στις εξόδους των stacked επιπέδων. Οι ταυτοτικές shortcut συνδέσεις δεν προσθέτουν επιπλέον παραμέτρους ή υπολογιστική πολυπλοκότητα. Ολόκληρο το δίκτυο μπορεί να εκπαιδευτεί end-to-end με τη μέθοδο Stochastic Gradient Descent μέσω backpropagation και μπορεί να υλοποιηθεί εύκολα χρησιμοποιώντας απλές βιβλιοθήκες. Η τεχνική της υπολειπόμενης μάθησης, όπως περιγράφηκε παραπάνω, εφαρμόζεται ανά ομάδα ορισμένων stacked επιπέδων και βασικό δομικό στοιχείο

της διαδικασίας αυτής αποτελεί το residual block. Τα residual blocks αποτελούνται από stacked layers και shortcut συνδέσεις, οι οποίες επιτρέπουν την πρόσθεση της εισόδου απευθείας στην έξοδο των μη γραμμικών επιπέδων. Μαθηματικά ένα residual block μπορεί να εκφραστεί ως εξής:

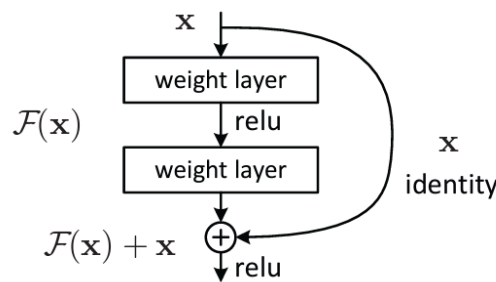
$$y = F(x, \{W_i\}) + x, \quad (4.1)$$

όπου x και y οι διανυσματικές εισοδοι και έξοδοι των επιπέδων αντίστοιχα και $F(x, W_i)$ η συνάρτηση των υπολοίπων.

Σε ένα παράδειγμα δύο επιπέδων, όπως φαίνεται στο Σχήμα 4.1, η F ορίζεται ως $F = W_2\sigma(W_1x)$, όπου σ δηλώνει τη συνάρτηση ενεργοποίησης *ReLU* και τα biases παραλείπονται για απλούστευση. Οι διαστάσεις των x και F , στην εξίσωση (4.1) πρέπει να είναι ίσες. Σε διαφορετική περίπτωση υπάρχουν τρεις επιλογές. Η πρώτη είναι να εκτελεστεί η ταυτοτική αντιστοίχιση μέσω των shortcut συνδέσεων, προσθέτοντας μηδενικές εγγραφές (zero-padding). Η δεύτερη είναι να εκτελεστεί μία γραμμική προβολή W_s , μέσω των shortcut συνδέσεων, για να ταιριάξουν οι διαστάσεις (πραγματοποιείται με συνελιξείς 1x1), επιλογή που δεν εισάγει επιπλέον παραμέτρους. Η εξίσωση (4.1), σε αυτή την περίπτωση παίρνει τη μορφή:

$$y = F(x, \{W_i\}) + W_s x, \quad (4.2)$$

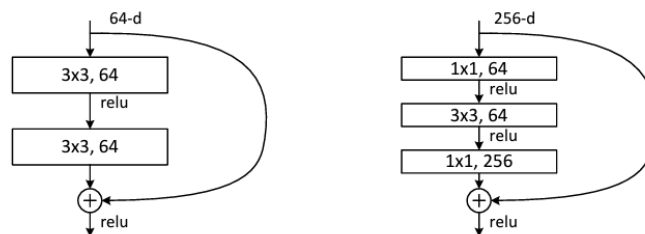
Η τρίτη επιλογή είναι να χρησιμοποιείται ο W_s μόνο όταν απαιτείται αντιστοίχιση διαστάσεων και η ταυτοτική αντιστοίχιση παντού αλλού. Σημειώνεται ότι παρά το γεγονός πως οι προηγούμενες αναλύσεις αφορούσαν πλήρως συνδεδεμένα επίπεδα για ευκολία, είναι εφαρμόσιμες και σε συνελκτικά επίπεδα. Η συνάρτηση $F(x, W_i)$, μπορεί να αντιπροσωπεύει πολλαπλά συνελκτικά επίπεδα. Η πρόσθεση στοιχείο προς στοιχείο εκτελείται μεταξύ δύο διαφορετικών feature maps, κανάλι προς κανάλι.



Σχήμα 4.1: Residual Learning: ένα residual μπλοκ 2 στρωμάτων
Πηγή: Deep Residual Learning for Image Recognition

Η βασική φιλοσοφία των Residual Networks ξεκινάει με ένα απλό νευρωνικό δίκτυο, εμπνευσμένο από την αρχιτεκτονική των VGG, χρησιμοποιώντας συνελκτικά επίπεδα που έχουν κατά κύριο λόγο 3x3 φίλτρα. Αυτή η αρχιτεκτονική ακολουθεί δύο βασικούς κανόνες σχεδιασμού: i) για ίδιο μέγεθος χάρτη χαρακτηριστικών (feature map) εξόδου, τα επίπεδα έχουν τον ίδιο αριθμό φίλτρων και ii) αν το μέγεθος του χάρτη χαρακτηριστικών (feature map) μειωθεί στο μισό, τότε

ο αριθμός των φίλτρων διπλασιάζεται ώστε να διατηρηθεί η χρονική πολυπλοκότητα ανά επίπεδο. Η υποδειγματοληψία (downsampling) πραγματοποιείται άμεσα μέσω των συνελκτικών επιπέδων, που έχουν stride ίσο με 2. Το δίκτυο τελειώνει με ένα global average pooling στρώμα και ένα πλήρως συνδεδεμένο στρώμα, όπου ο αριθμός των νευρώνων σε αυτό καθορίζεται από το εκάστοτε προς επίλυση πρόβλημα. Στη συνέχεια, για να μετατραπεί αυτό το απλό νευρωνικό δίκτυο σε residual δίκτυο, προστίθεται μία shortcut σύνδεση σε κάθε ζεύγος φίλτρων 3×3 . Η προσέγγιση αυτή, διατηρεί την απλότητα και αποδοτικότητα της αρχικής αρχιτεκτονικής. Η αρχιτεκτονική που περιγράφηκε παραπάνω αναφέρεται σε Residual Networks λίγων επιπέδων (18 ή 34). Για την εκτέλεση βαθύτερων Residual Networks (50 επιπέδων και μεγαλύτερα), το βασικό δομικό μπλοκ τροποποιείται σε ένα bottleneck μπλοκ. Για κάθε συνάρτηση υπολοίπων F , χρησιμοποιούμε stacked layers 3 επιπέδων αντί για 2 (Σχήμα 4.2). Τα τρία επίπεδα είναι 1×1 , 3×3 και 1×1 συνελίξεις, όπου τα επίπεδα 1×1 είναι υπεύθυνα για τη μείωση και στη συνέχεια την αύξηση των διαστάσεων, αφήνοντας το επίπεδο 3×3 ως ένα bottleneck με μικρότερες διαστάσεις εισόδου/εξόδου. Το Σχήμα 4.2 δείχνει ένα παράδειγμα, όπου και οι δύο σχεδιασμοί έχουν παρόμοια χρονική πολυπλοκότητα. Στις αρχιτεκτονικές, που γίνεται χρήση των bottleneck blocks οι shortcut συνδέσεις που εκτελούν την ταυτοτική αντιστοίχιση είναι ιδιαίτερα σημαντικές. Αν οι ταυτοτικές shortcut συνδέσεις, αντικατασταθούν με προβολές μπορεί ναδειχθεί ότι η χρονική πολυπλοκότητα και το μέγεθος του μοντέλου διπλασιάζονται. Έτσι, οι shortcut συνδέσεις που εκτελούν την ταυτοτική αντιστοίχιση οδηγούν σε αποδοτικότερα μοντέλα για τις αρχιτεκτονικές με bottleneck, για τον λόγο αυτό, όπως έχει προαναφερθεί, όταν οι διαστάσεις είναι διαφορετικές προτιμώνται οι εξής 2 επιλογές α) ταυτοτικές shortcut συνδέσεις με συμπλήρωση μηδενικών (zero-padding), β) προβολικές shortcut συνδέσεις για την αύξηση των διαστάσεων και όλες οι υπόλοιπες ταυτοτικές. Για την κατασκευή ενός Residual Network 50 στρωμάτων (ResNet50), αντικαθιστούμε κάθε μπλοκ 2 επιπέδων, στο δίκτυο των 34 επιπέδων, με ένα bottleneck μπλοκ 3 επιπέδων (Σχήμα 4.3). Για την αύξηση των διαστάσεων χρησιμοποιείται η επιλογή β.



Σχήμα 4.2: Μία συνάρτηση υπολοίπων F για το ImageNet. Αριστερά: ένα residual μπλοκ (σε χάρτες χαρακτηριστικών 56×56), για το ResNet34. Δεξιά: ένα bottleneck μπλοκ για τα ResNet50/101/152.

Πηγή: Deep Residual Learning for Image Recognition

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Σχήμα 4.3: Αρχιτεκτονικές για το ImageNet. Τα δομικά μπλοκ παρουσιάζονται στις αγκύλες, μαζί με τον αριθμό των stacked μπλοκ.

Πηγή: Deep Residual Learning for Image Recognition

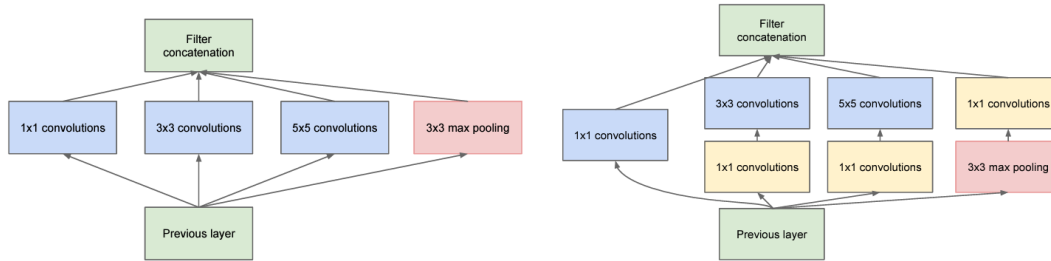
4.2 InceptionV3

Inception

Η αρχιτεκτονική Inception επικεντρώνεται στην ανάπτυξη μίας βέλτιστης τοπικά αραιής δομής σε ένα συνελκτικό νευρωνικό δίκτυο. Η βασική ιδέα πίσω από την αρχιτεκτονική αυτή βασίζεται στον εντοπισμό των τρόπων με τους οποίους η αραιή αυτή δομή μπορεί να προσεγγιστεί και να καλυφθεί, από διαθέσιμους πυκνούς παράγοντες (dense components). Δομικό στοιχείο του δικτύου αποτελούν τα συνελκτικά μπλοκ. Η προσέγγιση αυτή επιτρέπει την επανάληψη της ίδιας βέλτιστης τοπικής δομής σε διαφορετικά σημεία και επίπεδα του δικτύου, δημιουργώντας ένα βαθύ και ευρύ δίκτυο που μπορεί να αποτυπώσει σύνθετες δομές των εικόνων, με μεγάλη ακρίβεια. Ο Agha και οι συνεργάτες του [6] πρότειναν μία κατασκευή στρώμα προς στρώμα. Η διαδικασία αυτή απαιτεί την ανάλυση των συσχετίσεων στο τελευταίο στρώμα και την ομαδοποίηση των μονάδων του, σε ομάδες υψηλής συσχέτισης. Οι ομάδες αυτές στη συνέχεια αποτελούν τις μονάδες (νευρώνες) του επόμενου στρώματος και συνδέονται με τις μονάδες του προηγούμενου στρώματος. Κάθε μονάδα από το προηγούμενο στρώμα αντιστοιχεί σε κάποια περιοχή της εικόνας εισόδου και αυτές οι μονάδες συνδυάζονται σε ομάδες φίλτρων. Στα κατώτερα στρώματα του δικτύου, εκείνα δηλαδή που βρίσκονται πιο κοντά στην είσοδο, οι συσχετισμένες μονάδες θα συγκεντρωθούν σε τοπικές περιοχές. Αυτό θα οδηγήσει στη συγκέντρωση πολλών ομάδων σε μία μόνο περιοχή, που μπορούν να καλυφθούν από ένα στρώμα συνελίξεων 1x1 στο επόμενο στρώμα. Είναι αναμενόμενο να υπάρξουν λιγότερες, αλλά περισσότερο διασκορπισμένες ομάδες μονάδων που απαιτούν ανάλυση μέσω συνελίξεων με μεγαλύτερο μέγεθος φίλτρων, που συνεπάγεται την προοδευτική μείωση του αριθμού των διακριτών περιοχών καθώς αυτές εκτείνονται για να καλύψουν ευρύτερα κομμάτια της εικόνας. Η αρχιτεκτονική Inception, εστιάζοντας στην αποφυγή προβλήματος της ευθυγράμμισης των φίλτρων, περιορίζεται στη χρήση φίλτρων με διαστάσεις 1x1, 3x3, 5x5. Ο περιορισμός που αναφέρθηκε, οδηγεί σε μία αρχιτεκτονική που αποτελεί συνδυασμό διαφορετικών επιπέδων και ομάδων φίλτρων, με τις εξόδους να συγκεντρώνονται σε ένα ενιαίο διάλυμα

που αποτελεί την είσοδο για το επόμενο στάδιο. Επιπλέον, η ενσωμάτωση εναλλακτικών παράλληλων pooling μονοπατιών, σε κάθε στάδιο φαίνεται να προσφέρει πρόσθετα οφέλη στο δίκτυο, αναδεικνύοντας έτσι, την συνεισφορά των pooling λειτουργιών στην επιτυχία των σύγχρονων συνελκτικών νευρωνικών δικτύων. Καθώς τα Inception modules τοποθετούνται το ένα πάνω στο άλλο, παρατηρείται μία αλλαγή στις στατιστικές συσχετίσεις των εξόδων τους. Στα ανώτερα στρώματα, όπου ανιχνεύονται χαρακτηριστικά υψηλότερης αφαίρεσης, η χωρική συγκέντρωση αυτών φαίνεται να μειώνεται, γεγονός που υποδεικνύει αυξημένη ανάγκη για τη χρήση συνελίξεων 3×3 και 5×5 στα επίπεδα αυτά. Η χρήση των συνελίξεων αυτών οδηγεί σε ένα σημαντικό πρόβλημα, τουλάχιστον όσον αφορά τα Inception modules στη naïve μορφή τους, ακόμα και ένας όχι ιδιαίτερα μεγάλος αριθμός 5×5 συνελίξεων μπορεί να αποδειχθεί αρκετά ακριβός, σε υπολογιστικούς όρους, ειδικά όταν εφαρμόζονται σε στρώματα με πολλά φίλτρα. Το πρόβλημα εντείνεται με την προσθήκη pooling μονάδων: ο αριθμός των φίλτρων εξόδου ισούται με τον αριθμό των φίλτρων στο προηγούμενο στάδιο. Η συγχώνευση της εξόδου του pooling στρώματος με τις εξόδους των συνελκτικών στρωμάτων θα οδηγήσει σε αναπόφευκτη αύξηση του αριθμού των εξόδων από στάδιο σε στάδιο, γεγονός που θα κάνει το δίκτυο αναποτελεσματικό. Το πρόβλημα που περιγράφηκε οδηγεί στην δεύτερη βασική ιδέα της αρχιτεκτονικής Inception, που είναι η εφαρμογή μείωσης διαστάσεων και προβολών, στα σημεία όπου οι υπολογιστικές απαιτήσεις θα αυξάνονταν δυσανάλογα. Η προσέγγιση αυτή βασίζεται στην επιτυχία των embeddings. Ακόμα και τα χαμηλής διάστασης embeddings περιέχουν αρκετές πληροφορίες για σχετικά μεγάλα τμήματα της εικόνας. Τα embedding, αντιπροσωπεύουν πληροφορίες σε πυκνή, συμπιεσμένη μορφή και οι πληροφορίες αυτές είναι δυσκολότερο να μοντελοποιηθούν. Το επιθυμητό είναι να διατηρείται η αναπαράσταση των δεδομένων αραιή στις περισσότερες θέσεις, συμπιέζοντας τα σήματα μόνο όταν κρίνεται απαραίτητο. Κεντρική, στην εφαρμογή της στρατηγικής αυτής, είναι η χρήση συνελίξεων 1×1 για τη μείωση των διαστάσεων, πριν την εφαρμογή των πιο, υπολογιστικά, απαιτητικών 3×3 και 5×5 συνελίξεων. Οι 1×1 συνελίξεις δεν λειτουργούν μόνο ως τεχνική μείωσης διαστάσεων, αλλά περιλαμβάνουν και τη χρήση της γραμμικής ορθογώνιας συνάρτησης ενεργοποίησης, καθιστώντας τις 1×1 συνελίξεις ένα εργαλείο διπλού σκοπού που συμβάλει στην απόδοση του δικτύου. Τα τελικά αποτελέσματα απεικονίζονται στο Σχήμα 4.4(δεξιά). Σε γενικές γραμμές ένα δίκτυο Inception αποτελείται από μονάδες, όπως αυτές περιεγράφηκαν παραπάνω, τοποθετημένες η μία πάνω στην άλλη, με ενδιάμεσα max pooling επίπεδα, με βήμα 2, για την μείωση στο μισό της ανάλυσης του πλέγματος. Για αποδοτικότερη χρήση της μνήμης κατά την εκπαίδευση, τα Inception modules χρησιμοποιήθηκαν στα ανώτερα επίπεδα και στα κατώτερα χρησιμοποιήθηκαν παραδοσιακά συνελκτικά στρώματα. Ένα από τα βασικότερα χαρακτηριστικά της αρχιτεκτονικής Inception, που την καθιστά αποτελεσματική, αποτελεί το γεγονός πως επιτρέπει σημαντική αύξηση του αριθμού των μονάδων(νευρώνων) σε κάθε στάδιο χωρίς αυτό να συνοδεύεται από ανεξέλεγκτη αύξηση της υπολογιστικής πολυπλοκότητας. Ένα ακόμη πρακτικό όφελος του σχεδιασμού αυτού είναι ότι ανταποκρίνεται στην ιδέα, πως η επεξεργασία των οπτικών δεδομένων πρέπει να γίνεται σε πολλαπλές κλίμακες και να ακολουθεί η συγκέντρωσή τους σε ένα ενιαίο σύνολο, γεγονός που επιτρέπει στο επόμενο επίπεδο του δικτύου να συνδυάζει και να αποσπά χαρακτηριστικά από τις διάφορες αυτές κλί-

μακες ταυτόχρονα.



Σχήμα 4.4: Αριστερά: inception module naive μορφή. Δεξιά Inception module με μείωση διαστάσεων.

Πηγή: Going Deeper with Convolutions

GoogLeNet

Χρησιμοποιούμε το GoogLeNet, για να αναφερθούμε σε μία συγκεκριμένη εκδοχή της αρχιτεκτονικής Inception, η οποία και περιγράφεται στον Σχήμα 4.5. Όλες οι συνελίξεις, συμπεριλαμβανομένων και εκείνων μέσα στα Inception modules χρησιμοποιούν τη συνάρτηση ενεργοποίησης ReLu. Στο δίκτυο αυτό, κάθε νευρώνας επεξεργάζεται πληροφορίες από μία περιοχή της εικόνας με διαστάσεις 224x224 pixels, λαμβάνοντας υπόψιν τα χρωματικά κανάλια RGB με αφαίρεση της μέσης τιμής. Οι όροι «3x3 reduce» και «5x5 reduce» αναφέρονται στον αριθμό των φίλτρων 1x1 στο στρώμα μείωσης διαστάσεων που χρησιμοποιείται πριν από τις 3x3 και 5x5 συνελίξεις. Στην στήλη 'pool proj' δίνεται ο αριθμός των φίλτρων 1x1 στο στρώμα προβολής μετά την ενσωματωμένη διαδικασία max pooling. Όλα τα στρώματα μείωσης/προβολής χρησιμοποιούν επίσης τη συνάρτηση ενεργοποίησης ReLu. Το δίκτυο σχεδιάστηκε, ώστε να μπορεί να υλοποιείται σε μεμονωμένες συσκευές ακόμη και σε εκείνες που διαθέτουν περιορισμένους υπολογιστικούς πόρους. Έχει βάθος 22 στρωμάτων, όταν μετράμε μόνο τα στρώματα με παραμέτρους και επεκτείνεται στα 27 στρώματα αν μετρήσουμε και τα pooling στρώματα. Ο συνολικός αριθμός στρωμάτων, μετρώντας κάθε στρώμα εντός των Inception modules, που χρησιμοποιήθηκαν για την κατασκευή του δικτύου είναι περίπου 100. Το δίκτυο αντικαθιστά τα πλήρως συνδεδεμένα στρώματα με average pooling στρώματα, ακριβώς πριν το τελικό στάδιο της ταξινόμησης και ακολουθεί ένα επιπλέον γραμμικό στρώμα που δίνει τη δυνατότητα στο δίκτυο να προσαρμόζεται καλύτερα στα διάφορα σύνολα δεδομένων. Αυτή η αντικατάσταση βοηθάει στην αύξηση της ακρίβειας του δικτύου. Παρά το γεγονός, πως τα τελικά πλήρως συνδεδεμένα στρώματα αφαιρούνται διατηρείται η ανάγκη χρήσης dropout, για την αποφυγή της υπερπροσαρμογής. Μία καινοτομία που εφαρμόστηκε στο GoogLeNet, ήταν η ενσωμάτωση βοηθητικών ταξινομητών (auxiliary classifiers), με σκοπό την αποφυγή αναποτελεσματικής αντίστροφης διάδοσης των διαφορών (gradients) που μπορούσε να προκύψει λόγω του μεγάλου βάθους του δικτύου. Η αποτελεσματική διάδοση των διαφορών πίσω, προς τα εισαγωγικά στρώματα, για τη σωστή διαμόρφωση των βαρών του δικτύου αποτελεί θεμελιώδη διαδικασία κατά την εκπαίδευση. Οι βοηθητικοί ταξινομητές, έχουν τη μορφή μικρότερων συνελκτικών δικτύων και τοποθετούνται πάνω από

τις εξόδους των Inception modules(Σχήμα 4.5 : 4(a), 4(b)). Η ενσωμάτωση αυτή, ενθαρρύνει τη διακριτική ικανότητα των χαμηλότερων επιπέδων του δικτύου και ενισχύει το σήμα των διαφορών(gradients), που μεταδίδεται προς τα πίσω. Κατά τη διάρκεια της εκπαίδευσης, η απώλεια(loss) από αυτούς τους βοηθητικούς ταξινομητές προστίθεται στη συνολική απώλεια του δικτύου, με μειωμένη όμως βαρύτητα(συντελεστής βάρους 0.3). Στο στάδιο της συμπερασματολογίας, όταν το δίκτυο χρησιμοποιείται για την πρόβλεψη πάνω σε νέα δεδομένα, αυτοί οι βοηθητικοί ταξινομητές δεν λαμβάνονται υπόψιν και αφαιρούνται.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Σχήμα 4.5: Το δίκτυο GoogLeNet. Πηγή: Going Deeper with Convolutions

Η ακριβής δομή του επιπλέον αυτού δικτύου, περιλαμβάνοντας τον βοηθητικό ταξινομητή, είναι η εξής:

- Ένα average pooling στρώμα, με φίλτρο μεγέθους 5x5 και βήμα 3, που καταλήγει σε μία έξοδο 4x4x512 για το στάδιο (4a), και 4x4x528 για το στάδιο (4d).
- Ένα στρώμα συνέλιξης 1x1 με 128 φίλτρα, για τη μείωση των διαστάσεων, και τη γραμμική συνάρτηση ενεργοποίησης *ReLU*.
- Ένα dropout στρώμα, με 70% ποσοστό απόρριψης των εξόδων.
- Ένα γραμμικό στρώμα, με τη *Softmax* συνάρτηση απώλειας, ως ταξινομητή(προβλέποντας τις ίδιες 1000 κατηγορίες, όπως ο κύριος ταξινομητής του δικτύου, αλλά αφαιρείται κατά τη διαδικασία της συμπερασματολογίας).

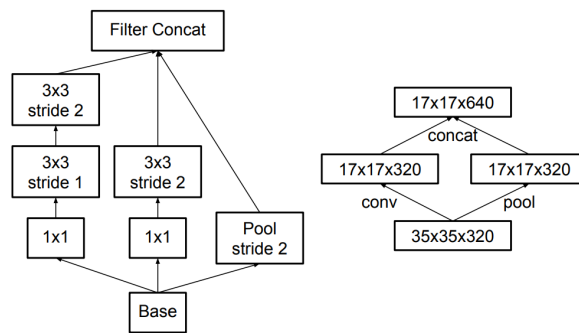
InceptionV2

Το InceptionV2 αποτελεί μία εξέλιξη του αρχικού GoogLeNet, που έφερε βελτιώσεις και καινοτομίες στην αρχική αρχιτεκτονική. Η διάταξη του δικτύου δίνεται στο Σχήμα 4.6. Βασική καινοτομία της αρχιτεκτονικής του InceptionV2, αποτελεί

η αντικατάσταση των παραδοσιακών 7x7 συνελίξεων, από ένα πολυστρωματικό δίκτυο με λιγότερες παραμέτρους, το οποίο χρησιμοποιεί το ίδιο μέγεθος εισόδου και βάθος εξόδου. Αυτό επιτυγχάνεται χρησιμοποιώντας τρία στρώματα συνελίξης 3x3, που συνολικά έχουν λιγότερες παραμέτρους και χαμηλότερο υπολογιστικό κόστος σε σύγκριση με μία 7x7 συνέλιξη. Για το Inception κομμάτι του δικτύου, χρησιμοποιούνται 3 παραδοσιακά Inception modules σε ένα πλέγμα 35x35, όπου κάθε ένα από αυτά έχει 288 φίλτρα. Χρησιμοποιώντας την τεχνική grid reduction τα δεδομένα μετατρέπονται σε 17x17, ο αριθμός των φίλτρων όμως αυξάνεται σε 768. Αυτό σημαίνει ότι το δίκτυο δημιουργεί περισσότερα χαρακτηριστικά από τα δεδομένα εισόδου. Το grid reduction αποτελεί μία τεχνική για τη μείωση των διαστάσεων των χαρτών χαρακτηριστικών σε συνελκτικά δίκτυα. Πιο συγκεκριμένα, προτείνεται η χρήση δύο παράλληλων μπλοκ, βήματος 2: ένα στρώμα pooling P(είτε average είτε max pooling) και ένα στρώμα συνελίξης C. Τα αποτελέσματα των φίλτρων των δύο μπλοκ συνδυάζονται όπως φαίνεται στο Σχήμα 4.7. Στη συνέχεια το δίκτυο περιλαμβάνει 5 factorized inception modules, όπως φαίνεται στο Σχήμα 4.8. Ακολουθεί μία επιπλέον μείωση του πλέγματος σε διαστάσεις 8x8, όπου ο αριθμός των φίλτρων αυξάνεται σε 1280. Η μείωση αυτή γίνεται ξανά με την τεχνική grid reduction. Στο τελικό στάδιο, στο πλέγμα 8x8, το δίκτυο διαθέτει δύο Inception modules, όπως φαίνεται στο Σχήμα 4.9. Κάθε ένα από τα Inception modules παράγει μία έξοδο με συνολικό μέγεθος φίλτρων 2048, για κάθε τετράγωνο του πλέγματος. Στόχος της αρχιτεκτονικής είναι η αύξηση της ακρίβειας σε συνδυασμό με το μειωμένο υπολογιστικό κόστος, έτσι παρά το γεγονός πως το InceptionV2 έχει βάθος 42 στρωμάτων, το υπολογιστικό κόστος είναι μόλις 2,5 περίπου φορές μεγαλύτερο από αυτό του GoogLeNet.

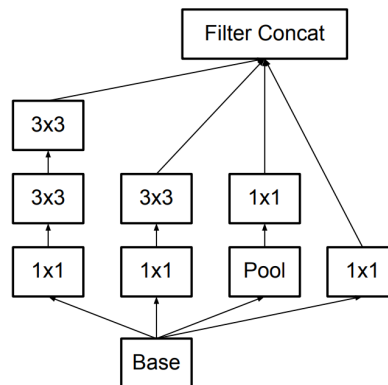
type	patch size/stride or remarks	input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	As in figure 5	35×35×288
5×Inception	As in figure 6	17×17×768
2×Inception	As in figure 7	8×8×1280
pool	8 × 8	8 × 8 × 2048
linear	logits	1 × 1 × 2048
softmax	classifier	1 × 1 × 1000

Σχήμα 4.6: Η διάταξη της αρχιτεκτονικής InceptionV2. Το μέγεθος εξόδου κάθε module είναι το μέγεθος εισόδου του επόμενου. Έχει χρησιμοποιηθεί η τεχνική grid reduction, για τη μείωση του μεγέθους των πλεγμάτων μεταξύ των Inception μπλοκ, όπου είναι εφικτό. Έχει χρησιμοποιηθεί συνέλιξη με προσθήκη μηδενικών(zero-padding), για τη διατήρηση του μεγέθους του πλέγματος. Zero-padding χρησιμοποιήθηκε, επιπλέον, μέσα στα Inception modules που δεν μειώνουν το μέγεθος του πλέγματος. Όλα τα υπόλοιπα στρώματα δεν χρησιμοποιούν zero-padding. Πηγή: Rethinking the Inception Architecture for Computer Vision

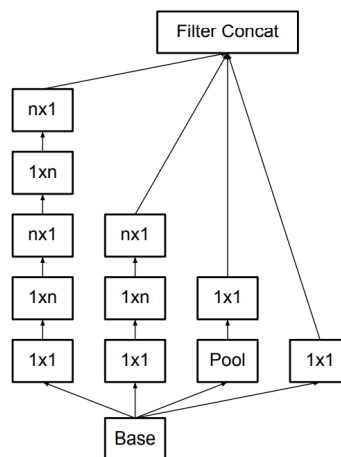


Σχήμα 4.7: Inception module που μειώνει το μέγεθος του πλέγματος, ενώ αυξάνει το μέγεθος των φίλτρων. Στα δεξιά, το διάγραμμα υπό την προοπτική των μεγεθών των πλεγμάτων και όχι των λειτουργιών που εφαρμόζονται.

Πηγή: Rethinking the Inception Architecture for Computer Vision



Σχήμα 4.8: Inception modules όπου κάθε συνέλιξη 5x5, αντικαθίσταται από δύο συνέλιξεις 3x3. Πηγή: Rethinking the Inception Architecture for Computer Vision



Σχήμα 4.9: Inception modules μετά την παραγοντοποίηση (factorization) των $n \times n$ συνέλιξεων. Στο InceptionV2, έχει επιλεγθεί $n = 7$ για το 17x17 πλέγμα. Πηγή: Rethinking the Inception Architecture for Computer Vision

Inception V3

Το InceptionV3 αποτελεί μία εξέλιξη στην αρχιτεκτονική του InceptionV2. Στην αρχιτεκτονική αυτή, γίνεται χρήση του Batch Normalization, όχι μόνο στα συνελικτικά στρώματα, αλλά και στο πλήρως συνδεδεμένο στρώμα του βοηθητικού ταξινομητή (auxiliary classifier). Αυτή η τεχνική βοηθά στη βελτίωση της απόδοσης του δικτύου και συμβάλλει στη σταθεροποίηση της διαδικασίας της εκπαίδευσης.

4.3 MobileNetV2

Βασικό δομικό στοιχείο της αρχιτεκτονικής του MobileNetV2 αποτελεί το bottleneck depth-separable συνελικτικό μπλοκ με υπολείμματα (residuals). Η ακριβής δομή ενός bottleneck depth-separable συνελικτικού μπλοκ, δίνεται στο Σχήμα 4.10. Βασική ιδέα των depthwise separable συνελίξεων αποτελεί η διάσπαση της συνέλιξης σε δύο ξεχωριστά στρώματα. Στο πρώτο στρώμα που ονομάζεται συνέλιξη depthwise, εφαρμόζεται ένα φίλτρο σε κάθε κανάλι εισόδου ξεχωριστά και στο δεύτερο στρώμα πραγματοποιείται μία 1x1 συνέλιξη, που ονομάζεται pointwise συνέλιξη και είναι υπεύθυνη για τη δημιουργία νέων χαρακτηριστικών μέσω του υπολογισμού γραμμικών συνδυασμών των εισερχόμενων καναλιών. Οι depthwise separable συνελίξεις αποτελούν μία εναλλακτική για την αντικατάσταση των παραδοσιακών συνελίξεων. Εμπειρικά λειτουργούν σχεδόν εξίσου καλά με τις παραδοσιακές αλλά το κόστος τους είναι μόνο:

$$h_i \cdot w_i \cdot d_i(k^2 + d_j),$$

που είναι το άθροισμα των depthwise και των 1x1 pointwise συνελίξεων. Ουσιαστικά, οι depthwise separable συνελίξεις μειώνουν τους υπολογισμούς, σε σύγκριση με τα παραδοσιακά συνελικτικά στρώματα, περίπου κατά έναν παράγοντα k^2 . Το MobileNetV2 χρησιμοποιεί $k = 3$ (3x3 depthwise separable συνελίξεις), έτσι το υπολογιστικό κόστος είναι 8 έως 9 φορές μικρότερο από αυτό των παραδοσιακών συνελίξεων με μία μόνο μικρή μείωση στην ακρίβεια. Σημαντικό στοιχείο της αρχιτεκτονικής του MobileNetV2 αποτελεί η χρήση των Inverted Residual Blocks, που αποτελούνται από ένα στρώμα επέκτασης (expansion layer), ακολουθεί η depthwise συνέλιξη και τέλος η συνέλιξη pointwise (1x1), η οποία μειώνει τον αριθμό των καναλιών σχηματίζοντας το bottleneck μπλοκ. Η αρχιτεκτονική του MobileNetV2 περιλαμβάνει το αρχικό πλήρως συνελικτικό στρώμα με 32 φίλτρα, ακολουθούμενο από 19 residual bottleneck μπλοκ, όπως περιγράφεται στον Πίνακα 2. Χρησιμοποιείται η *ReLU6*, ως μη γραμμική συνάρτηση λόγω της ανθεκτικότητάς της σε χαμηλής ακρίβειας υπολογισμούς. Χρησιμοποιείται μέγεθος πυρήνα 3x3 και εφαρμόζεται dropout και batch normalization κατά τη διάρκεια της εκπαίδευσης. Με εξαίρεση το πρώτο στρώμα, χρησιμοποιείται σταθερός παράγοντας εκπαίδευσης σε όλο το δίκτυο. Πειράματα έδειξαν πως οι παράγοντες επέκτασης μεταξύ 5 και 10 οδηγούν σε σχεδόν ταυτόσημες καμπύλες απόδοσης. Το MobileNetV2 προσαρμόζει την αρχιτεκτονική του για διαφορετικά επίπεδα απόδοσης, χρησιμοποιώντας δύο κύριες ρυθμιζόμενες υπερπαραμέτρους: την ανάλυση εικόνας και τον πολλαπλασιαστή πλάτους. Αυτές οι υπερπαραμέτροι επιτρέπουν στο δίκτυο να προσαρμόζεται σύμφωνα με τις διάφορες απαιτήσεις σε ότι αφορά την ακρίβεια και την από-

δοση. Το βασικό δίκτυο MobileNetV2, με πολλαπλασιαστή πλάτους 1 και εικόνες εισόδου μεγέθους 224x224 pixels, απαιτεί 300 εκατομμύρια πολλαπλασιασμούς-προσθήσεις(MAdds) και χρησιμοποιεί 3,4 εκατομμύρια παραμέτρους. Σε σύγκριση με προηγούμενα MobileNets[18], το MobileNetV2, για πολλαπλασιαστές πλάτους μικρότερους από ένα, ο πολλαπλασιαστής εφαρμόζεται σε όλα τα στρώματα εκτός από το τελευταίο συνελκτικό στρώμα. Αυτό βελτιώνει την απόδοση για μικρότερα μοντέλα. Η αποδοτική διαχείριση μνήμης είναι κρίσιμη για την ανάπτυξη γρήγορων και αποδοτικών μοντέλων. Η αρχιτεκτονική του MobileNetV2 συνδυάζει την ελαχιστοποίηση της απαιτούμενης μνήμης, διατηρώντας υψηλή την απόδοση. Με την εισαγωγή των bottleneck μπλοκ και των αντίστροφων υπολειπόμενων δομών, το MobileNetV2 καταφέρνει να υλοποιήσει ένα βαθύ μοντέλο με σημαντικά μειωμένη απαίτηση μνήμης, γεγονός ιδιαίτερα σημαντικό σε εφαρμογές επεξεργασίας εικόνων, όπως η ανάλυση εικόνων για την ανίχνευση καρκίνου του εγκεφάλου. Σε μία τυπική υλοποίηση του δικτύου, όπως αυτή πραγματοποιείται για παράδειγμα με την χρήση του TensorFlow ή του Caffe δημιουργείται ένας κατευθυνόμενος ακυκλικός υπεργράφος G . Ο υπεργράφος αυτός αποτελείται από ακμές που αντιπροσωπεύουν τις λειτουργίες και κόμβους που αντιπροσωπεύουν τα διανύσματα των ενδιάμεσων υπολογισμών. Ο υπολογισμός πραγματοποιείται με στόχο την ελαχιστοποίηση του συνολικού αριθμού των διανυσμάτων που πρέπει να αποθηκευτούν στη μνήμη. Στην πιο γενική περίπτωση, αναζητούνται όλες οι πιθανές σειρές υπολογισμού $\Sigma(G)$ και επιλέγεται αυτή που ελαχιστοποιεί. Ακολουθεί ο αντίστοιχος τύπος:

$$M(G) = \min_{\pi \in \Sigma(G)} \max_{i \in \{1..n\}} \left[\sum_{A \in R(i, \pi, G)} |A| \right] + \text{size}(\pi_i), \quad (4.3)$$

όπου $R(i, \pi, G)$ είναι η λίστα με τα ενδιάμεσα διανύσματα που είναι συνδεδεμένα με οποιονδήποτε από τους κόμβους $i..n$, $|A|$ το μέγεθος του διανύσματος A και $\text{size}(i)$ η συνολική μνήμη που απαιτείται για αποθήκευση κατά τη διάρκεια της λειτουργίας i .

Για γράφους που έχουν απλές παράλληλες δομές, όπως οι υπολειπόμενες συνδέσεις, η μέγιστη συνολική μνήμη που απαιτείται για την εκτέλεση του γράφου G , δίνεται από τον ακόλουθο απλοποιημένο τύπο:

$$M(G) = \max_{op \in G} \left[\sum_{A \in op_{inp}} |A| + \sum_{B \in op_{out}} |B| + |op| \right], \quad (4.4)$$

Πιο απλά η απαιτούμενη μνήμη, είναι η μέγιστη συνολική ποσότητα των εισόδων και των εξόδων σε όλες τις λειτουργίες του γράφου. Η δυνατότητα αντιμετώπισης των bottleneck στρωμάτων, ως μία ενιαία λειτουργία μειώνει την απαιτούμενη μνήμη, καθώς αυτή θα κυριαρχείται από το μέγεθος των διανυσμάτων του bottleneck στρώματος και όχι από τα μεγαλύτερα εσωτερικά διανύσματα του στρώματος. Ένας τελεστής στρώματος bottleneck $F(x)$, μπορεί να εκφραστεί ως σύνθεση τριών τελεστών $F(x) = [A \circ N \circ B]x$, όπου A είναι ένας γραμμικός μετασχηματισμός $A: R^{s \times s \times k} \rightarrow R^{s \times s \times n}$, N ένας μη γραμμικός μετασχηματισμός $N: R^{s \times s \times n} \rightarrow R^{s' \times s' \times n'}$ και B ένας γραμμικός μετασχηματισμός $B: R^{s' \times s' \times n'} \rightarrow R^{s' \times s' \times k'}$. Για το MobileNetV2

$N = \text{ReLU6} \circ \text{dwise} \circ \text{ReLU6}$. Υποθέτοντας ότι το μέγεθος εισόδου είναι $|x|$ και το μέγεθος της εξόδου $|y|$, η απαιτούμενη μνήμη για τον υπολογισμό του $F(x)$ μπορεί να είναι τόσο χαμηλή όσο $|s^2k| + |s'^2k'| + O(\max(s^2, s'^2))$. Ο αλγόριθμος βασίζεται στο γεγονός πως το εσωτερικό διάνυσμα I μπορεί να αναπαρασταθεί ως συνένωση t διανυσμάτων μεγέθους t/n , το καθένα, και η λειτουργία τότε μπορεί να αναπαρασταθεί ως $F(x) = F(x) = \sum_{i=1}^t (A_i \circ N \circ B_i)(x)$. Συσσωρεύοντας το άθροισμα, μόνο ένα ενδιάμεσο μπλοκ διανυσμάτων μεγέθους n/t απαιτείται να διατηρείται στη μνήμη κάθε φορά. Χρησιμοποιώντας $n = t$, διατηρείται τελικά μόνο ένα κανάλι της ενδιάμεσης αναπαράστασης κάθε φορά, μειώνοντας περαιτέρω την απαιτούμενη μνήμη. Μία σημαντική παρατήρηση για τον τρόπο εκτέλεσης του MobileNetV2, αποτελεί ότι η προσέγγιση της t -way διαίρεσης για τον υπολογισμό του $F(x)$, δεν επηρεάζεται από την τιμή του t . Ωστόσο, η χρήση πολλών μικρών πολλαπλασιαστικών πράξεων αντί μίας μεγαλύτερης μπορεί να μειώσει την απόδοση. Βρέθηκε ότι η βέλτιστη χρήση της τεχνικής αυτής επιτυγχάνεται όταν το t είναι μια μικρή σταθερά, που κυμαίνεται μεταξύ του 2 και του 5, η οποία μειώνει σημαντικά τις απαιτήσεις μνήμης αλλά επιτρέπει την αποδοτική χρήση των βελτιστοποιημένων πολλαπλασιαστικών πράξεων και τελεστών συνέλιξης.

Input	Operator	Output
$h \times w \times k$	1x1 conv2d, ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwise s=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

Σχήμα 4.10: Δομή ενός residual bottleneck block που μετασχηματίζει από k σε k' κανάλια, με stride s και expansion factor t .

Πηγή: MobileNetV2: Inverted Residuals and Linear Bottlenecks

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	-

Σχήμα 4.11: MobileNetV2: Κάθε γραμμή περιγράφει μία ακολουθία ενός ή περισσότερων ίδιων επιπέδων που επαναλαμβάνονται n φορές. Όλα τα επίπεδα, στην ίδια ακολουθία έχουν τον ίδιο αριθμό εξόδων c . Το πρώτο επίπεδο κάθε ακολουθίας έχει stride ίσο με s και όλα τα υπόλοιπα έχουν stride ίσο με 1.

Πηγή: MobileNetV2: Inverted Residuals and Linear Bottlenecks

Κεφάλαιο 5

Μεθοδολογία και Αξιολόγηση των Μοντέλων Βαθιάς Μάθησης

Στο παρόν Κεφάλαιο παρουσιάζεται αναλυτικά το Σύνολο των Δεδομένων που χρησιμοποιήθηκε για την υλοποίηση της μελέτης και η προ επεξεργασία αυτού. Ακολουθεί η περιγραφή της μεθοδολογίας που ακολουθήθηκε για την υλοποίηση των διαφορετικών μοντέλων Βαθιάς Μάθησης καθώς και τα αποτελέσματα που έδωσαν. Η μελέτη πραγματοποιήθηκε σε Python3, σε προγραμματιστικό περιβάλλον Google Colab με T4 GPU.

5.1 Σύνολο Δεδομένων

Για την εκπόνηση της παρούσας διπλωματικής εργασίας, χρησιμοποιήθηκε ένα δημόσια διαθέσιμο σύνολο δεδομένων από το Kaggle[1], το οποίο περιέχει διάφορες εικόνες MRI με εγκεφαλικούς όγκους από το Brats2015 . Αποτελείται από 3762 εικόνες, που ανήκουν σε δύο κλάσεις, παρουσία όγκου(2079 εικόνες) και απουσία όγκου(1683 εικόνες). Το σύνολο των εικόνων για την διεξαγωγή των πειραμάτων χωρίστηκε στα εξής υποσύνολα:

- 80% του συνόλου δεδομένων, δηλαδή 3046 εικόνες, χρησιμοποιήθηκαν στο σύνολο εκπαίδευσης.
- 10% του συνόλου δεδομένων, δηλαδή 339 εικόνες, χρησιμοποιήθηκαν στο σύνολο επικύρωσης.
- 10% του συνόλου δεδομένων, δηλαδή 377 εικόνες, χρησιμοποιήθηκαν στο σύνολο δοκιμής.

5.2 Εισαγωγή και Προ-επεξεργασία των Δεδομένων

Για την διεξαγωγή των πειραμάτων στην παρούσα μελέτη, τα δεδομένα φορτώθηκαν αρχικά από ένα αρχείο CSV, στο οποίο βρίσκονται οι ετικέτες και τα ονόματα των εικόνων. Παράλληλα φορτώθηκαν οι αντίστοιχες εικόνες από τον φάκελο που ήταν αποθηκευμένες. Στη συνέχεια δημιουργήθηκε ένα Data Frame στο οποίο καταγράφηκαν τα ονόματα των εικόνων μαζί με τις αντίστοιχες ετικέτες του, για την

διευκόλυνση της αντιστοίχισης κάθε εικόνας με την κατηγορία της. Για να εξασφαλιστεί η εύκολη και άμεση πρόσβαση στις εικόνες, δημιουργήθηκε ένα λεξικό που συνδέει τα ονόματα των εικόνων με τις αντίστοιχες πλήρεις διαδρομές τους στον αποθηκευτικό χώρο του υπολογιστή. Στη συνέχεια, έγινε έλεγχος για τη διασφάλιση της ύπαρξης όλων των αναφερόμενων εικόνων στον αποθηκευτικό χώρο του υπολογιστή, εξασφαλίζοντας έτσι πως όλες οι εικόνες του Data Frame είναι διαθέσιμες για χρήση. Ακολούθως έγινε ο διαχωρισμός του συνόλου των δεδομένων σε τρία υποσύνολα: το σύνολο εκπαίδευσης, το σύνολο επικύρωσης και το σύνολο δοκιμής, με τη χρήση της συνάρτησης 'train test split'. Οι αρχικές εικόνες που φορτώθηκαν είχαν διαστάσεις 240x240 pixels και όλες μετασχηματίστηκαν στα 224x224 pixels, για να διασφαλιστεί η προσαρμογή και η συμβατότητά τους στις απαιτήσεις εκπαίδευσης των μοντέλων και η μείωση του χρόνου εκπαίδευσης των μοντέλων. Επιπλέον οι εικόνες κανονικοποιήθηκαν, με τη χρήση της παραμέτρου 'rescale=1/255', που μετατρέπει τις τιμές των εικονοστοιχείων σε μία νέα κλίμακα με τιμές μεταξύ των 0 και 1. Αυτό βοηθά στη βελτίωση της απόδοσης και την ευκολότερη εκπαίδευση των νευρωνικών δικτύων, καθώς οι μικρότερες τιμές εικονοστοιχείων είναι πιο εύκολα διαχειρίσιμες κατά τις μαθηματικές πράξεις που πραγματοποιούνται στην διαδικασία της εκπαίδευσης. Μία κρίσιμη τροποποίηση που πραγματοποιήθηκε αφορά τη μετατροπή των τιμών της στήλης 'Class' από ακεραίους σε συμβολοσειρές. Η μετατροπή αυτή είναι ιδιαίτερα σημαντική για την συμβατότητα των δεδομένων με τη μέθοδο 'flow from data' της TensorFlow/ KerasImageDataGenerator, που απαιτεί τα δεδομένα να είναι σε μορφή συμβολοσειράς για τη σωστή διαχείριση των εικόνων, ειδικά όταν χρησιμοποιούνται ρυθμίσει όπως 'class mode=binary'. Η διαδικασία της κανονικοποίησης υλοποιήθηκε μέσω της 'ImageDataGenerator' του TensorFlow. Το επόμενο βήμα της διαδικασίας της προ επεξεργασίας των δεδομένων περιλάμβανε τη δημιουργία των data generators, οι οποίοι παρέχουν μικρότερες ποσότητες δεδομένων (batches) στο μοντέλο κάθε φορά, βοηθώντας στην αποδοτικότερη και πιο ομαλή διαχείριση της διαδικασίας της εκπαίδευσης. Χωρίς τη χρήση των data generators, θα έπρεπε να φορτώνεται και να επεξεργάζεται ταυτόχρονα ολόκληρο το σύνολο των δεδομένων, γεγονός που θα οδηγούσε σε αυξημένες απαιτήσεις σε υπολογιστική ισχύ και μνήμη. Με τη χρήση των bathes των 32 εικόνων τα μοντέλα βαθιάς μάθησης εκπαιδεύονται αποδοτικά, με μικρότερες ποσότητες δεδομένων σε κάθε επανάληψη, διευκολύνοντας τη διαχείριση της μνήμης και την πρόσβαση στα δεδομένα, ενώ ταυτόχρονα διατηρείται η ποιότητα και η ακρίβεια της εκπαίδευσης.

5.3 Μέτρα Αξιολόγησης της Απόδοσης

Όταν το σύνολο δεδομένων δοκιμής έχει ίσο αριθμό δειγμάτων από κάθε κατηγορία, το accuracy αποτελεί ένα χρήσιμο μέτρο αξιολόγησης της απόδοσης. Ωστόσο επειδή στο σύνολο δεδομένων, που χρησιμοποιήθηκε στην παρούσα εργασία, παρατηρήθηκε μία ελαφριά ανισορροπία μεταξύ των εγγραφών κάθε κλάσης χρησιμοποιήθηκαν περισσότερα μέτρα αξιολόγησης της απόδοσης των μοντέλων. Πιο συγκεκριμένα, για να αξιολογηθούν αποτελεσματικά οι διάφορες μέθοδοι που υλοποιήθηκαν για την ταξινόμηση των εγκεφαλικών όγκων, χρησιμοποιήθηκαν, εκτός από το accuracy, το precision, το recall (ή sensitivity), το specificity και το f1-score. Για

τον προσδιορισμό των βασικών εννοιών που χρησιμοποιούνται για τον ορισμό των παραπάνω μέτρων, άλλα και για την παροχή ενός οπτικού τρόπου κατανόησης τους, παρουσιάζεται ένα Confusion Matrix. Ένα confusion matrix, είναι ένας πίνακας που παρέχει πληροφορίες για την απόδοση και την ποιότητα ενός μοντέλου, για δύο ή περισσότερες κλάσεις, δεδομένου ενός συνόλου δεδομένων δοκιμής για τα οποία είναι γνωστές οι πραγματικές τιμές. Στο Σχήμα 5.1 που ακολουθεί, παρουσιάζεται ένα διασδιάστατο confusion matrix, για την ταξινόμηση μεταξύ δύο κλάσεων.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Σχήμα 5.1: Confusion Matrix.

Όπου,

- True Positives/TP (Θετικές Σωστές Προβλέψεις): Ένα TP αποτέλεσμα είναι εκείνο, στο οποίο το μοντέλο προβλέπει με ακρίβεια τη θετική κλάση.
- True Negatives/TN (Αρνητικές Σωστές Προβλέψεις): Ένα TN αποτέλεσμα είναι εκείνο, στο οποίο το μοντέλο προβλέπει με ακρίβεια την αρνητική κλάση.
- False Positives/FP (Θετικές Λανθασμένες Προβλέψεις): Ένα FP αποτέλεσμα, είναι εκείνο στο οποίο το μοντέλο προβλέπει λανθασμένα τη θετική κλάση.
- False Negatives/FN (Αρνητικές Λανθασμένες Προβλέψεις): Ένα FN αποτέλεσμα, είναι εκείνο στο οποίο το μοντέλο προβλέπει λανθασμένα την αρνητική κλάση.

Με βάση τα παραπάνω, θα ορίσουμε και τα υπόλοιπα μέτρα αξιολόγησης της απόδοσης ενός μοντέλου.

Accuracy: είναι ένα στατιστικό μέτρο που χρησιμοποιείται για την περιγραφή της απόδοσης του μοντέλου γενικά, σε όλες τις κλάσεις του. Ορίζεται ως το πηλίκο του αριθμού των συνολικών σωστών προβλέψεων προς τον συνολικό αριθμό των προβλέψεων που πραγματοποίησε το μοντέλο.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (5.1)$$

Specificity: η ειδικότητα, είναι το ποσοστό των πραγματικών αρνητικών που προβλέφθηκαν ως αρνητικά, δηλαδή τα αρνητικά σωστά. Η ειδικότητα ποσοτικοποιεί το πόσο καλά το μοντέλο αποφεύγει τις λανθασμένες θετικές προβλέψεις.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5.2)$$

Precision: γνωστή και ως θετική προβλεπτική αξία, ορίζεται ως ο λόγος του αριθμού των σωστών θετικών προβλέψεων, προς τον συνολικό αριθμό των θετικών προβλέψεων.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5.3)$$

Recall ή Sensitivity: γνωστή και ως ρυθμός επιτυχίας, ορίζεται ως ο λόγος των σωστών θετικών προβλέψεων προς το άθροισμα των σωστών θετικών και των λανθασμένων αρνητικών. Ουσιαστικά είναι το ποσοστό των σωστών θετικών που το μοντέλο αναγνώρισε ορθά.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5.4)$$

F1-score: Ορίζεται ως ο σταθμισμένος ή αρμονικός μέσος του Precision και του Recall.

$$\text{F1-score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \quad (5.5)$$

5.4 Μεθοδολογία

Στην παρούσα διπλωματική εργασία, υλοποιήθηκε μία σειρά διαφορετικών αρχιτεκτονικών βασισμένων στην τεχνική του Transfer Learning, με τη χρήση τριών προεκπαιδευμένων μοντέλων ως βάση: τα InceptionV3, ResNet50, και MobileNetV2 και δύο ειδικά κατασκευασμένων συνελκτικών νευρωνικών δικτύων. Σκοπός ήταν η σύγκριση των αποτελεσμάτων των διάφορων αρχιτεκτονικών και παραμέτρων εκπαίδευσης στην απόδοση των μοντέλων για την ανίχνευση και ταξινόμηση εγκεφαλικών όγκων σε δύο κλάσεις: απουσία και παρουσία εγκεφαλικού όγκου, μέσω της ανάλυσης εικόνων MRI. Το σύνολο των δεδομένων που χρησιμοποιήθηκε, καθώς και η εισαγωγή και προ επεξεργασία των δεδομένων περιγράφηκαν παραπάνω.

Αναφορικά με τα προεκπαιδευμένα μοντέλα, στο κομμάτι της εκπαίδευσης, τα στρώματα αυτών διατηρήθηκαν παγωμένα, για να διασφαλιστεί ότι τα βάρη τους δεν θα ενημερωθούν κατά τη διάρκεια της εκπαίδευσης. Χρησιμοποιήθηκε η ρύθμιση 'include top=False', που αφαίρεσε τα τελευταία πλήρως συνδεδεμένα στρώματα, για την προσθήκη νέων προσαρμοσμένων στο συγκεκριμένο πρόβλημα στρωμάτων. Τα τρία μοντέλα χρησιμοποίησαν τα βάρη του ImageNet, που αποτελεί μία εκτενή βάση δεδομένων. Η επιλογή αυτή, επέτρεψε την άμεση αξιοποίηση της προηγούμενης γνώσης για την αποτελεσματική αντιμετώπιση της διαδικασίας ανίχνευσης και ταξινόμησης εγκεφαλικών όγκων. Πριν από την αρχικοποίηση των βαρών, των πλήρως συνδεδεμένων και συνελκτικών στρωμάτων που προστέθηκαν, ορίστηκε μία σταθερή τιμή seed ίση με το 0 και στη συνέχεια χρησιμοποιήθηκε η μέθοδος 'glorot uniform'. Η 'glorot uniform' αποτελεί μία μέθοδο αρχικοποίησης των

βαρών για τα στρώματα ενός νευρωνικού δικτύου, η οποία προσπαθεί να διατηρήσει την κλίμακα των gradients σταθερή σε όλα τα επίπεδα κατά την αρχικοποίηση. Επιλέγει τα βάρη από μία ομοιόμορφη κατανομή μέσα σε ένα διάστημα που εξαρτάται από τον αριθμό των εισόδων και των εξόδων του στρώματος. Η επιλογή της μεθόδου, έγινε για διασφαλιστεί πως κάθε φορά που ξεκινά η εκπαίδευση των μοντέλων, τα βάρη θα αρχικοποιούνται με τον ίδιο τρόπο, γεγονός που μειώνει την τυχαιότητα που μπορεί να οφείλεται σε ιδιαίτερα καλή ή κακή αρχικοποίηση των βαρών και καθιστά άμεση την σύγκριση μεταξύ διαφορετικών αρχιτεκτονικών. Επιπλέον, στη συνάρτηση 'train test split' το 'random state' ορίστηκε 0, για να εξασφαλιστεί πως κάθε φορά που εκτελείται ο κώδικας, η διαδικασία διαχωρισμού θα παράγει τα ίδια ακριβώς αποτελέσματα, δηλαδή τα σύνολα εκπαίδευσης, επικύρωσης και δοκιμής αποτελούνται κάθε φορά από τις ίδιες ακριβώς εικόνες. Με δεδομένο ότι όλα τα μοντέλα ξεκινούν με βάρη που αρχικοποιήθηκαν με τον ίδιο τρόπο, διατηρούν τις τιμές των υπόλοιπων υπερπαραμέτρων ίδιες και τα επιμέρους υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής αποτελούνται κάθε φορά από τις ίδιες εικόνες, οι διαφορές στην απόδοσή τους μπορεί να αποδοθεί στη διαφορετική δομή των επιμέρους αρχιτεκτονικών. Χρησιμοποιήθηκε ο Adam optimizer με ρυθμό μάθησης 0.0001, που αποτελεί την πιο κοινή επιλογή, καθώς είναι υπολογιστικά αποδοτικός, εύκολος στην υλοποίηση, έχει μικρές απαιτήσεις μνήμης και λειτουργεί καλά ακόμα και σε προβλήματα που περιλαμβάνουν μεγάλες ποσότητες δεδομένων ή παραμέτρων [21]. Η διαδικασία της εκπαίδευσης περιλάμβανε 10 εποχές και ελέγχθηκε μέσω της χρήσης του Early Stopping callback της TensorFlow, το οποίο χρησιμοποιήθηκε για την διακοπή της εκπαίδευσης στην περίπτωση που δεν υπήρχε βελτίωση της απόδοσης για πέντε συνεχόμενες εποχές. Σκοπός ήταν η εξοικονόμηση χρόνου και πόρων, αλλά και η πρόληψη της υπερεκπαίδευσης. Η συνάρτηση απώλειας που χρησιμοποιήθηκε, ήταν η binary crossentropy, μία τυπική επιλογή για προβλήματα δυαδικής ταξινόμησης. Αρχικά, στα μοντέλα τα τελευταία πλήρως συνδεδεμένα στρώματα αντικαταστάθηκαν από ένα Global Average Pooling στρώμα, με σκοπό τη μείωση των διαστάσεων εξόδου των μοντέλων και την συμπύκνωση της πληροφορίας. Ακολούθησε ένα πλήρως συνδεδεμένο στρώμα (Dense Layer), με ένα νευρώνα και συνάρτηση ενεργοποίησης τη σιγμοειδή συνάρτηση, για την τελική ταξινόμηση, αντικατοπτρίζοντας τη δυαδική φύση του προβλήματος (απουσία ή παρουσία εγκεφαλικού όγκου). Οι επόμενες δύο υλοποιήσεις, στηρίχθηκαν στην προηγούμενη μεθοδολογία, αλλά την ενίσχυσαν με μία σειρά επιπλέον στρωμάτων που προέκυψαν από συνδυασμό και τροποποιήσεις αρχιτεκτονικών που έχουν προταθεί στη σχετική βιβλιογραφία [17],[23],[7]. Πιο συγκεκριμένα, αμέσως μετά το Global Average Pooling στρώμα προστέθηκε ένα Batch Normalization στρώμα, για την κανονικοποίηση των εισόδων και τη βελτίωση της σταθερότητας και ταχύτητας εκπαίδευσης των μοντέλων. Ακολούθως προστέθηκε ένα πλήρως συνδεδεμένο στρώμα με συνάρτηση ενεργοποίησης τη *ReLU* και δοκιμάστηκε μία φορά με 64 νευρώνες και μία με 128 νευρώνες. Έπειτα προστέθηκε ένα στρώμα Dropout, με τη χρήση του οποίου τυχαία επιλεγμένοι νευρώνες, στην περίπτωσή μας 20% αυτών, αγνοούνται κατά τη διάρκεια της εκπαίδευσης σε κάθε εποχή. Η χρήση του στρώματος αυτού βοηθά στην πρόληψη του φαινομένου της υπερπροσαρμογής, μειώνοντας την εξάρτηση του εκάστοτε δικτύου από συγκεκριμένα βάρη. Το τελικό πλήρως συνδεδεμένο στρώμα, για την ταξινόμηση, παρέμεινε όπως ακριβώς περιγράφηκε

παραπάνω. Στην επόμενη υλοποίηση των μοντέλων, υιοθετήθηκε μία διαφορετική αρχιτεκτονική για τα προστιθέμενα στρώματα προέρχεται από τη μελέτη [10]. Παρά το γεγονός πως η συγκεκριμένη αρχιτεκτονική έχει προταθεί για ένα πρόβλημα ανίχνευσης και ταξινόμησης εγκεφαλικών όγκων σε τρεις κλάσεις: το γλοίωμα, το μηνιγγίωμα και τον όγκο της υπόφυσης, εδώ θα ελεγχθεί η απόδοσή της στο πρόβλημα της δυαδικής ταξινόμησης των εγκεφαλικών, που εξετάζεται. Πιο συγκεκριμένα, έχουν προστεθεί 3 συνελκτικά στρώματα με φίλτρα 32, 64, και 128 αντίστοιχα, διαστάσεων 3x3, με stride 1x1 και οι διαστάσεις των εικόνων έχουν διατηρηθεί με την τεχνική padding 'same'. Κάθε συνελκτικό στρώμα ακολουθείται από ένα Max pooling στρώμα για τη μείωση των διαστάσεων της εικόνας και αμέσως μετά έχει τοποθετηθεί ένα στρώμα Dropout ποσοστού 20%. Σε όλα τα συνελκτικά στρώματα, ως συνάρτηση ενεργοποίησης έχει χρησιμοποιηθεί η συνάρτηση *ReLU*. Ακολούθως, έχει προστεθεί ένα Flatten στρώμα για την εισαγωγή των δεδομένων σε ένα πλήρως συνδεδεμένο στρώμα 128 νευρώνων. Τέλος, το πλήρως συνδεδεμένο στρώμα εξόδου, με ένα νευρώνα για την δυαδική ταξινόμηση και συνάρτηση ενεργοποίησης τη σιγμοειδή. Η επόμενη υλοποίηση των μοντέλων, αφορούσε την αντικατάσταση των τελευταίων πλήρως συνδεδεμένων στρωμάτων με μία σειρά νέων, ενσωματώνοντας και συνδυάζοντας στοιχεία από τις προηγούμενες αρχιτεκτονικές που περιγράφηκαν. Αρχικά, χρησιμοποιήθηκε ένα συνελκτικό στρώμα με 32 φίλτρα διαστάσεων 3x3, με stride 1x1 και οι διαστάσεις των εικόνων διατηρήθηκαν με τη χρήση της τεχνικής padding 'same'. Έπειτα προστέθηκαν δύο ακόμη συνελκτικά στρώματα με 64 και 128 φίλτρα διαστάσεων 3x3 αντίστοιχα, αυξάνοντας έτσι τη δυνατότητα του δικτύου να αναγνωρίζει περισσότερο περίπλοκα χαρακτηριστικά και μοτίβα των δεδομένων εισόδου. Στα συνελκτικά στρώματα χρησιμοποιήθηκε ως συνάρτηση ενεργοποίησης η *ReLU*. Ακολούθως χρησιμοποιήθηκε ένα Global Average Pooling στρώμα για τη μείωση των διαστάσεων των feature maps και συνεπώς τη μείωση της πολυπλοκότητας και των υπολογιστικών απαιτήσεων του δικτύου. Η επιλογή του Global Average Pooling στρώματος έναντι του Flatten γίνεται για την αποφυγή του φαινομένου της υπερπροσαρμογής, που επιτυγχάνεται μέσω της μείωσης του αριθμού των παραμέτρων και της πολυπλοκότητας, όπως προαναφέρθηκε. Τέλος, προστέθηκαν δύο πλήρως συνδεδεμένα στρώματα, το πρώτο με 16 νευρώνες και συνάρτηση ενεργοποίησης τη *ReLU* και το δεύτερο με ένα νευρώνα και συνάρτηση ενεργοποίησης τη σιγμοειδή, για την τελική δυαδική ταξινόμηση των εικόνων. Σημειώνεται ότι στην τελευταία αυτή προσέγγιση τα μοντέλα χρειάστηκαν περισσότερες εποχές εκπαίδευσης. Πιο συγκεκριμένα το μοντέλο που χρησιμοποίησε ως βάση το InceptionV3 εκπαιδεύτηκε για 30 εποχές, ενώ αυτά που χρησιμοποίησαν το ResNet50 και το MobileNetV2 για 50 εποχές.

Στη συνέχεια ακολουθήθηκε μία διαφορετική μεθοδολογία, για την ανίχνευση και ταξινόμηση των εγκεφαλικών όγκων. Η μεθοδολογία είναι εμπνευσμένη από τη σχετική βιβλιογραφία αλλά έχει υποστεί τροποποιήσεις [19], [20], [24]. Δημιουργήθηκαν και εκπαιδεύτηκαν δύο συνελκτικά νευρωνικά δίκτυα, απλά στη δομή τους, με τη χρήση της βιβλιοθήκης TensorFlow και του πακέτου Keras. Το σύνολο των δεδομένων που χρησιμοποιήθηκε, καθώς και η εισαγωγή και προ επεξεργασία των δεδομένων είναι ακριβώς ίδια με αυτά των μοντέλων που έχουν ήδη περιγραφεί. Αρχικά, για το πρώτο συνελκτικό νευρωνικό δίκτυο, δημιουργήθηκε το μοντέλο

με τη μέθοδο ‘models.Sequential’ και ακολούθως προστέθηκαν τρία επίπεδα συνελίξης με 32, 64 και 128 φίλτρα διαστάσεων 3x3 αντίστοιχα. Μετά από κάθε ένα από τα συνελικτικά στρώματα, τοποθετήθηκε ένα Max pooling στρώμα. Σε κάθε συνελικτικό στρώμα, ως συνάρτηση ενεργοποίησης χρησιμοποιήθηκε η συνάρτηση *ReLU*. Ακολούθως προστέθηκε ένα Global Average Pooling στρώμα και στη συνέχεια δύο πλήρως συνδεδεμένα στρώματα, το πρώτο αποτελούμενο από 32 νευρώνες και συνάρτηση ενεργοποίησης τη *ReLU* και το δεύτερο από 1 νευρώνα και συνάρτηση ενεργοποίησης τη σιγμοειδή, για την τελική ταξινόμηση. Στα συνελικτικά και πλήρως συνδεδεμένα στρώματα, χρησιμοποιήθηκε η μέθοδος ‘glorot uniform’ με σταθερό seed, για την διασφάλιση της επαναληψιμότητας και της άμεσης σύγκρισης των διαφορετικών αρχιτεκτονικών που έχουν υλοποιηθεί. Στη συνέχεια, το ίδιο μοντέλο υλοποιήθηκε ξανά με την προσθήκη ενός Max pooling στρώματος αμέσως μετά το τρίτο συνελικτικό στρώμα και ακολούθως ένα επιπλέον συνελικτικό στρώμα με 256 φίλτρα διαστάσεων 3x3, έτσι προέκυψε το δεύτερο συνελικτικό νευρωνικό δίκτυο. Τα μοντέλα εκπαιδεύτηκαν για 50 εποχές, χρησιμοποιήθηκε ο Adam optimizer, με αρχικό learning rate ίσο με 0.0001, η binary crossentropy συνάρτηση απώλειας και η μετρική accuracy, για την εκπαίδευση. Για τη βελτιστοποίηση της διαδικασίας της εκπαίδευσης χρησιμοποιήθηκαν δύο callbacks, το ‘EarlyStopping’ και το ‘LearningRateSchedulerStepDecay’. Το πρώτο χρησιμοποιήθηκε για την αποφυγή του φαινομένου της υπερπροσαρμογής, καθώς λειτουργεί διακόπτοντας την εκπαίδευση εάν η τιμή της απώλειας του συνόλου επικύρωσης δεν βελτιωθεί για πέντε συνεχόμενες εποχές. Το δεύτερο επιτρέπει την προσαρμογή του ρυθμού μάθησης (learning rate), μειώνοντας τον, καθ’ όλη τη διάρκεια της εκπαίδευσης. Ο αρχικός ρυθμός μάθησης ήταν 0.0001 και μειωνόταν κατά το ήμισυ κάθε 5 εποχές σύμφωνα με τις παραμέτρους ‘drop rate’ και ‘epochs drop’ που ορίστηκαν. Οι αρχιτεκτονικές, αλλά και ο αριθμός των παραμέτρων κάθε στρώματος των δύο συνελικτικών νευρωνικών δικτύων φαίνονται αναλυτικά στα Σχήματα 5.2 και 5.3.

```

Model: "sequential"

```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_2 (Conv2D)	(None, 52, 52, 128)	73856
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 32)	4128
dense_1 (Dense)	(None, 1)	33

```

=====
Total params: 97409 (380.50 KB)
Trainable params: 97409 (380.50 KB)
Non-trainable params: 0 (0.00 Byte)

```

Σχήμα 5.2: Το Συνελικτικό Νευρωνικό Δίκτυο 8 Επιπέδων

```

Model: "sequential_1"
-----
Layer (type)                Output Shape                Param #
-----
conv2d_4 (Conv2D)           (None, 222, 222, 32)       896
max_pooling2d_3 (MaxPoolin  (None, 111, 111, 32)       0
g2D)
conv2d_5 (Conv2D)           (None, 109, 109, 64)       18496
max_pooling2d_4 (MaxPoolin  (None, 54, 54, 64)         0
g2D)
conv2d_6 (Conv2D)           (None, 52, 52, 128)        73856
max_pooling2d_5 (MaxPoolin  (None, 26, 26, 128)        0
g2D)
conv2d_7 (Conv2D)           (None, 24, 24, 256)        295168
global_average_pooling2d_1  (None, 256)                 0
(GlobalAveragePooling2D)
dense_2 (Dense)              (None, 32)                  8224
dense_3 (Dense)              (None, 1)                   33
-----
Total params: 396673 (1.51 MB)
Trainable params: 396673 (1.51 MB)
Non-trainable params: 0 (0.00 Byte)

```

Σχήμα 5.3: Το Συνελικτικό Νευρωνικό Δίκτυο 10 Επιπέδων

5.5 Αποτελέσματα

Στη συνέχεια θα παρουσιαστούν και ερμηνευτούν τα κύρια αποτελέσματα, των μοντέλων που περιγράφηκαν παραπάνω, για την ανίχνευση και ταξινόμηση των εγκεφαλικών όγκων. Δεδομένης της ελαφριάς ανισορροπίας που παρατηρήθηκε στο σύνολο των δεδομένων που χρησιμοποιήθηκε, με την κλάση της παρουσίας όγκου να υπερεκπροσωπείται (2079 εικόνες παρουσίας όγκου έναντι 1683 εικόνων απουσίας όγκου), η αξιολόγηση της απόδοσης των διάφορων μοντέλων εκτός από το accuracy περιλαμβάνει επιπλέον τις μετρικές: precision, recall, f1-score και specificity. Το precision μετρά την ικανότητα ενός μοντέλου να προβλέπει σωστά την παρουσία όγκου, καταδεικνύοντας με τον τρόπο αυτό το ποσοστό των ψευδώς θετικών αποτελεσμάτων, που θα μπορούσαν να οδηγήσουν σε περαιτέρω ιατρικές παρεμβάσεις και περιττή ανησυχία των ασθενών. Το recall μετρά την ικανότητα ενός μοντέλου να εντοπίζει τις πραγματικές περιπτώσεις όγκου, προσδιορίζοντας ουσιαστικά το ποσοστό των ασθενών που εμφανίζουν όγκο και δεν θα παραλειφθούν από το σύστημα. Το f1-score αποτελεί ένα μέτρο που συνδυάζει τα precision και recall, παρέχοντας έναν ενιαίο δείκτη για την εκτίμηση της συνολικής αποδοτικότητας των μοντέλων. Έτσι, το f1-score στην ουσία αξιολογεί την ισορροπία μεταξύ της δυνατότητας του εκάστοτε μοντέλου να ανιχνεύει και ταξινομεί σωστά τις περιπτώσεις εμφάνισης εγκεφαλικού όγκου (precision) και της ικανότητας του να ανακαλύπτει το σύνολο των πραγματικών περιστατικών που εμφάνισαν εγκεφαλικό όγκο (recall). Τέλος, το specificity καταδεικνύει την ικανότητα ενός μοντέλου να αναγνωρίζει τις περιπτώσεις χωρίς όγκο, παρέχοντας με τον τρόπο αυτό μία εκτίμηση για το πόσο καλά το εκάστοτε μοντέλο αποφεύγει την υπερδιάγνωση. Η ανάλυση όλων των μετρικών που περιγράφηκαν επιτρέπει μία πιο ολοκληρωμένη κατανόηση της απόδοσης ενός μοντέλου, αναδεικνύοντας τις δυνατότητες αλλά και τα περιθώρια βελτίωσης της διαγνωστικής ακρίβειας. Είναι ιδιαίτερα σημαντικό, ιδίως σε ιατρικές εφαρμογές

όπως αυτή που μελετάται στην παρούσα διπλωματική εργασία, η επίτευξη ισορροπημένης απόδοσης που μεγιστοποιεί την αναγνώριση των πραγματικά θετικών περιπτώσεων, που χρήζουν άμεσης ιατρικής παρέμβασης, και ελαχιστοποιεί τον αριθμό των ψευδώς θετικών διαγνώσεων. Ακολουθεί ο πίνακας με τα αναλυτικά αποτελέσματα, όλων των μοντέλων που περιγράφηκαν και υλοποιήθηκαν.

Model	Accuracy	Precision	Recall	F1-score	Specificity
InceptionV3	0.87	0.87	0.87	0.87	0.89
ResNet50	0.76	0.77	0.77	0.77	0.86
MobileNetV2	0.84	0.84	0.84	0.84	0.86
InceptionV3+ANN1	0.93	0.93	0.93	0.93	0.93
ResNet50+ANN1	0.87	0.87	0.87	0.87	0.88
MobileNetV2+ANN1	0.93	0.93	0.93	0.93	0.93
InceptionV3+ANN2	0.95	0.95	0.95	0.95	0.96
ResNet50+ANN2	0.88	0.88	0.88	0.88	0.90
MobileNetV2+ANN2	0.94	0.94	0.94	0.94	0.95
InceptionV3+CNN1	0.95	0.95	0.95	0.95	0.94
ResNet50+CNN1	0.86	0.86	0.86	0.86	0.88
MobileNetV2+CNN1	0.94	0.94	0.94	0.94	0.91
InceptionV3+CNN2	0.94	0.94	0.94	0.94	0.94
ResNet50+CNN2	0.83	0.83	0.83	0.83	0.82
MobileNetV2+CNN2	0.92	0.92	0.92	0.92	0.92
CNN-8layers	0.82	0.82	0.82	0.82	0.86
CNN-10layers	0.87	0.87	0.87	0.87	0.88

Πίνακας 5.1: Μοντέλα Ταξινόμησης και Αποτελέσματα

Από τον παραπάνω πίνακα παρατηρείται πως τα μοντέλα κατέγραψαν συνολικά εξαιρετικά αποτελέσματα στην ανίχνευση και ταξινόμηση των εγκεφαλικών όγκων. Διατήρησαν, παρά τις τροποποιήσεις και διαφορετικές προσεγγίσεις που ακολουθήθηκαν, τη συνέπεια τους αναφορικά με την απόδοση που είχαν στην εκάστοτε υλοποίηση, με το InceptionV3 να καταγράφει συνολικά τα καλύτερα αποτελέσματα και το ResNet50 τα λιγότερο ικανοποιητικά. Το MobileNetV2 διατήρησε εξαιρετική απόδοση στο σύνολο των υλοποιήσεων, αρκετά κοντά σε αυτή του πρωτοπόρου InceptionV3. Στην πρώτη υλοποίηση των μοντέλων που περιγράφηκαν παραπάνω το InceptionV3 κατέγραψε accuracy, precision, recall και f1-score 87% ενώ το specificity ήταν 89%. Το ResNet50 κατέγραψε accuracy 76%, recall, f1-score, precision 77% και specificity 86%, ενώ το MobileNetV2 έδωσε καλύτερα αποτελέσματα με accuracy, precision, recall, f1-score στο 84% και specificity 86%. Στην δεύτερη υλοποίηση των μοντέλων που ακολούθησε, παρατηρήθηκαν εξαιρετικά αποτελέσματα με το InceptionV3 και το MobileNetV2 να καταγράφουν accuracy, precision, recall, f1-score και specificity 93%, ενώ το ResNet50 να καταγράφει accuracy, precision, recall, f1-score 87% και specificity 88%. Τα τρία μοντέλα στην υλοποίηση αυτή κατάφεραν να βελτιώσουν σημαντικά την απόδοσή τους σε σύγκριση με τα αντίστοιχα της προηγούμενης υλοποίησης, με τα InceptionV3 και MobileNetV2 να δίνουν εξαιρετικά αποτελέσματα στην ανίχνευση και ταξινόμηση εγκεφαλικών όγκων. Η τρίτη υλοποίηση έδωσε συνολικά τα υψηλότερα αποτελέ-

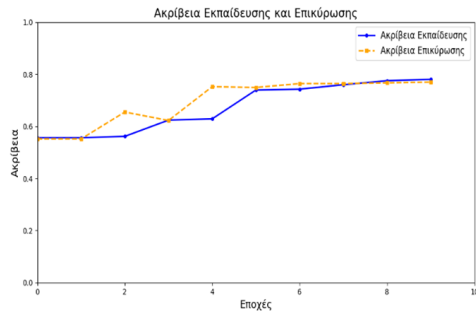
σματα και για τα τρία μοντέλα, κατά τη διερεύνηση που πραγματοποιήθηκε στην παρούσα εργασία. Το InceptionV3 κατέγραψε accuracy, precision, recall, f1-score 95% και specificity 96% ενώ το ResNet50 κατέγραψε accuracy, precision, recall, f1-score 88% και specificity 90%. Το MobileNetV2 κατέγραψε accuracy, precision, recall, f1-score 94% και specificity 95%. Η επόμενη προσέγγιση, δεν κατάφερε να βελτιώσει τις αποδόσεις των μοντέλων συγκριτικά με την προηγούμενη διατήρησε όμως υψηλά τα ποσοστά των μέτρων απόδοσης, με το InceptionV3 να καταγράφει accuracy, precision, recall, f1-score 95% και specificity 94% και το MobileNetV2 να καταγράφει accuracy, precision, recall, f1-score 94% και specificity 91%. Το ResNet50 κατέγραψε accuracy, precision, recall, f1-score 86% και specificity 88%. Η τελευταία προσέγγιση έδωσε λίγο χαμηλότερα αποτελέσματα συγκριτικά με την προηγούμενη. Το InceptionV3 κατέγραψε accuracy, precision, recall, f1-score και specificity 94% , το MobileNetV2 κατέγραψε accuracy, precision, recall, f1-score και specificity 92%. Τέλος, το ResNet50 κατέγραψε accuracy, precision, recall, f1-score 83% και specificity 82%. Τα δύο συνελκτικά νευρωνικά δίκτυα που υλοποιήθηκαν αξιολογήθηκαν με τα ίδια μέτρα απόδοσης, με αυτά που αξιολογήθηκαν και τα μοντέλα που βασίστηκαν στην τεχνική του Transfer Learning. Το πρώτο κατέγραψε accuracy, precision, recall, f1-score 82% και specificity 86%. Το δεύτερο βελτίωσε τα αποτελέσματα του πρώτου, δεν έδωσε όμως ιδιαίτερα ικανοποιητικά αποτελέσματα. Πιο συγκεκριμένα κατέγραψε accuracy, precision, recall, f1-score 87% και specificity 88%. Σημειώνεται ότι τα μοντέλα αυτά δεν κατάφεραν σε γενικές γραμμές να ξεπεράσουν τα αποτελέσματα των μοντέλων που χρησιμοποίησαν την τεχνική Transfer Learning.

5.6 Καμπύλες Ακρίβειας κα Απώλειας

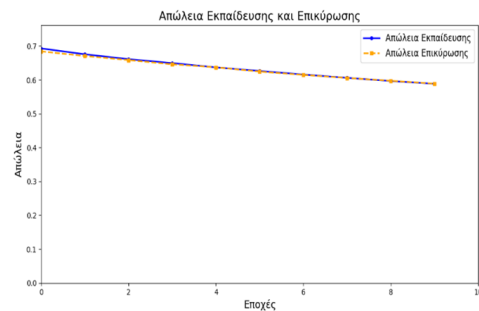
Για τη βελτίωση και αξιολόγηση της εκπαίδευσης και της ικανότητας των μοντέλων βαθιάς μάθησης να γενικεύουν, η ανάλυση των καμπύλων ακρίβειας και απώλειας για τα σύνολα εκπαίδευσης και επικύρωσης αποτελεί μία σημαντική διαδικασία. Οι καμπύλες αυτές, παρέχουν έναν οπτικό και διαισθητικό τρόπο για την παρακολούθηση των μοντέλων, αναφορικά με την ικανότητά τους να μαθαίνουν και να προσαρμόζονται στα δεδομένα με την πάροδο του χρόνου. Ιδανικά, οι καμπύλες ακρίβειας θα πρέπει να αυξάνουν και οι καμπύλες απώλειας να μειώνουν σταδιακά και για τα δύο σύνολα, με τις καμπύλες εκπαίδευσης και επικύρωσης να συγκλίνουν και στις δύο περιπτώσεις. Αυτό υποδεικνύει ότι το εκάστοτε μοντέλο, έχει την ικανότητα να γενικεύει αποτελεσματικά σε νέα δεδομένα και δεν εμφανίζει σημάδια υπερπροσαρμογής. Συνεπώς, μεγάλη απόκλιση στις καμπύλες ακρίβειας και απώλειας, μεταξύ των συνόλων εκπαίδευσης και επικύρωσης μπορεί να υποδεικνύει προβλήματα υπερπροσαρμογής, που σημαίνει ότι το μοντέλο είναι υπερβολικά εξαρτημένο από τα δεδομένα του συνόλου εκπαίδευσης και αποτυγχάνει να προσαρμοστεί καλά σε νέα δεδομένα. Στην περίπτωση που παρατηρηθεί χαμηλή ακρίβεια και υψηλή απώλεια στα σύνολα εκπαίδευσης και επικύρωσης, αποτελεί σημάδι υποπροσαρμογής(underfitting), γεγονός που υποδεικνύει ότι το μοντέλο δεν έχει εκπαιδευτεί επαρκώς από τα δεδομένα εκπαίδευσης. Για κάθε υλοποίηση και μοντέλο που περιγράφηκε παραπάνω σχεδιάστηκαν οι καμπύλες ακρίβειας και απώλειας. Παρατηρήθηκε ότι κανένα μοντέλο δεν εμφάνισε σημάδια υπερπροσαρμογής

εκτός από τα μοντέλα InceptionV3 και MobileNetV2 της 4^{ης} προσέγγισης και υλοποίησης. Ενδεικτικά παρουσιάζονται και σχολιάζονται, οι καμπύλες ακρίβειας και απώλειας ορισμένων από τα μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία.

Οι καμπύλες ακρίβειας και απώλειας για το ResNet50 της πρώτης υλοποίησης.



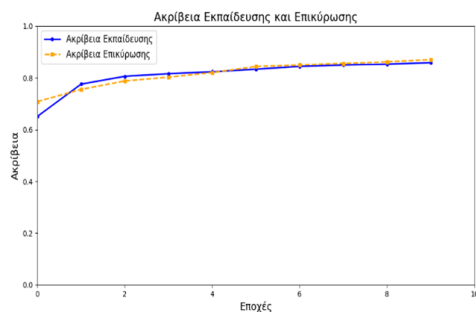
Σχήμα 5.4: ResNet50:Καμπύλες Ακρίβειας.



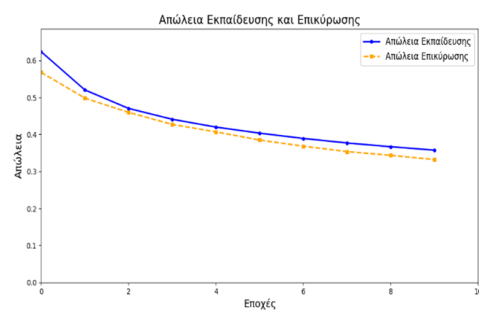
Σχήμα 5.5: ResNet50:Καμπύλες Απώλειας.

Από το γράφημα της ακρίβειας για τα σύνολα εκπαίδευσης και επικύρωσης παρατηρείται, ιδιαίτερα μετά την τέταρτη εποχή εκπαίδευσης, ότι οι καμπύλες αρχίζουν να έχουν σταθερή αυξητική πορεία και να συγκλίνουν. Οι καμπύλες απώλειας των δύο συνόλων επίσης παρουσιάζουν φθίνουσα πορεία κατά τη διάρκεια της εκπαίδευσης και όπως και οι καμπύλες ακρίβειας, συγκλίνουν. Με βάση τις καμπύλες ακρίβειας και επικύρωσης το δίκτυο δεν υπερπροσαρμόζεται στα δεδομένα, γεγονός που το καθιστά κατάλληλο για γενίκευση σε νέα δεδομένα. Παρατηρείται επιπλέον, ότι η τιμή της ακρίβειας για το σύνολο εκπαίδευσης δεν είναι ιδιαίτερα υψηλή και αντίστοιχα η τιμή της απώλειας του συνόλου εκπαίδευσης δεν είναι ιδιαίτερα χαμηλή. Αυτό μπορεί να εξαρτάται από μία σειρά παραγόντων, το πιο πιθανό όμως είναι ότι είτε το μοντέλο χρειάζεται περισσότερες εποχές εκπαίδευσης, είτε αύξηση της πολυπλοκότητάς του, με την εισαγωγή επιπλέον στρωμάτων, που θα το βοηθήσουν να μάθει πιο περίπλοκα μοτίβα και σχέσεις των δεδομένων εκπαίδευσης.

Ακολουθούν οι καμπύλες ακρίβειας και απώλειας του MobileNetV2, της ίδιας προσέγγισης



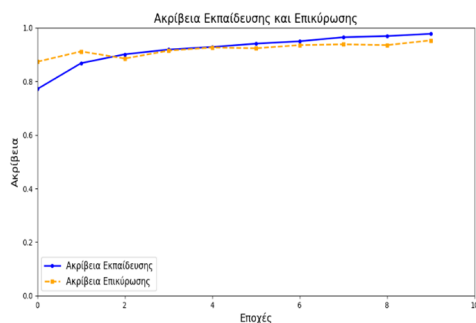
Σχήμα 5.6: MobileNetV2:Καμπύλες Ακρίβειας.



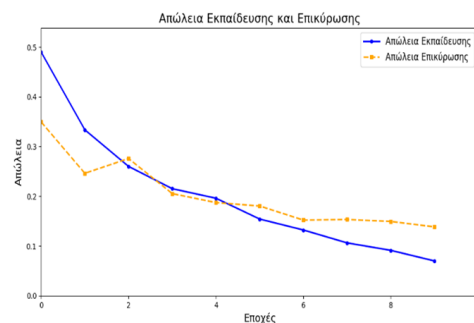
Σχήμα 5.7: MobileNetV2:Καμπύλες Απώλειας.

Οι καμπύλες ακρίβειας και απώλειας του MobileNetV2 οδηγούν σε αντίστοιχα με τα προηγούμενα συμπεράσματα. Παρότι καταγράφονται υψηλότερες και χαμηλότερες τιμές της ακρίβειας και της απώλειας αντίστοιχα του συνόλου εκπαίδευσης, δεν είναι ιδιαίτερα ικανοποιητικές. Παρατηρείται επιπλέον στην περίπτωση αυτή, ότι οι καμπύλες του συνόλου επικύρωσης βρίσκονται πάνω και κάτω από αυτές του συνόλου εκπαίδευσης στο γράφημα της ακρίβειας και της απώλειας αντίστοιχα. Το γεγονός αυτό, οφείλεται κατά πάσα πιθανότητα στην έλλειψη επαρκούς εκπαίδευσης του δικτύου. Το δίκτυο δεν έχει κατορθώσει να «μάθει» επαρκώς από τα δεδομένα εκπαίδευσης, γεγονός που πιθανώς οφείλεται στον μικρό αριθμό των εποχών εκπαίδευσης και την έλλειψη πολυπλοκότητας του δικτύου.

Ακολουθούν τα γραφήματα με τις καμπύλες Ακρίβειας και Απώλειας του InceptionV3 της 4^{ης} προσέγγισης και υλοποίησης.



Σχήμα 5.8: InceptionV3+CNN1: Καμπύλες Ακρίβειας.

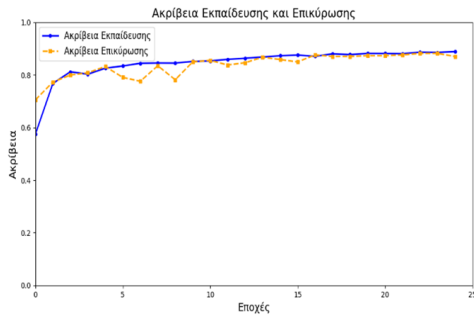


Σχήμα 5.9: InceptionV3+CNN1: Καμπύλες Απώλειας.

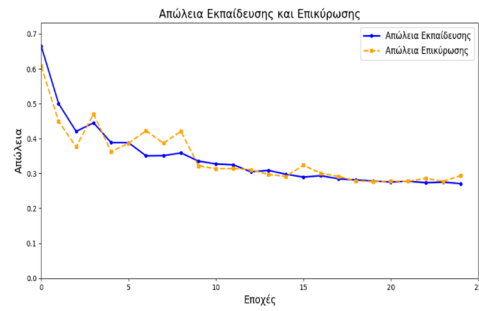
Από τα γραφήματα είναι εμφανές ότι τόσο οι τιμές της ακρίβειας όσο και οι τιμές της απώλειας του συνόλου εκπαίδευσης ήταν εξαιρετικά υψηλές και χαμηλές αντίστοιχα. Αυτό υποδεικνύει την εξαιρετική ικανότητα του μοντέλου στην ανίχνευση μοτίβων και σχέσεων του συνόλου εκπαίδευσης. Οι καμπύλες ακρίβειας των συνόλων εκπαίδευσης και επικύρωσης έχουν καθ' όλη τη διάρκεια της εκπαίδευσης σταθερή αυξητική πορεία, ενώ οι καμπύλες της απώλειας των δύο συνόλων σταθερή φθίνουσα πορεία, παρατηρείται όμως ιδίως μετά την 6η εποχή εκπαίδευσης πως οι καμπύλες των δύο συνόλων αρχίζουν να απομακρύνονται η μία από την άλλη. Το φαινόμενο αυτό είναι αρκετά πιο εμφανές στο γράφημα με τις καμπύλες απώλειας των δύο συνόλων και υποδεικνύει ότι από την εποχή εκείνη και μετά αρχίζουν να υπάρχουν σημάδια ελαφριάς υπερπροσαρμογής του δικτύου στα δεδομένα εκπαίδευσης.

Τέλος παρουσιάζονται τα γραφήματα με τις καμπύλες ακρίβειας και απώλειας των συνόλων εκπαίδευσης και επικύρωσης για το συνελκτικό νευρωνικό δίκτυο των 10 επιπέδων.

Από τα παραπάνω γραφήματα, παρατηρούμε ότι το μοντέλο δεν κατόρθωσε να καταγράψει ιδιαίτερα υψηλές και χαμηλές τιμές για την ακρίβεια και την απώλεια του συνόλου εκπαίδευσης αντίστοιχα, γεγονός που συνδέεται, στην περίπτωση αυτή, με την έλλειψη επαρκούς πολυπλοκότητας του δικτύου. Οι καμπύλες ακρίβειας των συνόλων εκπαίδευσης και επικύρωσης έχουν σταθερή αύξουσα πορεία,



Σχήμα 5.10: CNN 10 layers: Καμπύλες Ακρίβειας.



Σχήμα 5.11: CNN 10 layers: Καμπύλες Απώλειας.

ενώ οι καμπύλες απώλειας των δύο συνόλων σταθερή φθίνουσα πορεία. Παρατηρείται μία ελαφριά απόκλιση των καμπύλων, και στα δύο γραφήματα, από την 25^η εποχή εκπαίδευσης και μετά. Το γεγονός αυτό υποδεικνύει ότι το μοντέλο δεν πρέπει να εκπαιδευτεί άλλο καθώς θα χάσει τη δυνατότητά του να γενικεύει σε νέα δεδομένα. Σημειώνεται, πως το callback EarlyStopping που χρησιμοποιήθηκε διέκοψε σωστά την εκπαίδευση μετά την 25^η εποχή.

Κεφάλαιο 6

Συμπεράσματα

Για την αποτελεσματική ανίχνευση και ταξινόμηση εγκεφαλικών όγκων, από ένα σύνολο δεδομένων αποτελούμενο από 3762 εικόνες MRI, υλοποιήθηκαν και συγκρίθηκαν μεταξύ τους διαφορετικές προσεγγίσεις και τροποποιήσεις των προ εκπαιδευμένων μοντέλων InceptionV3, ResNet50 και MobileNetV2 με χρήση της τεχνικής Transfer Learning και δύο ειδικά κατασκευασμένα, συνελκτικά νευρωνικά δίκτυα 8 και 10 επιπέδων αντίστοιχα. Σκοπός κάθε προσέγγισης ήταν η βελτίωση της απόδοσης των μοντέλων, χωρίς όμως αυτή να συνοδεύεται από υπερβολική αύξηση της πολυπλοκότητάς τους. Ταυτόχρονα, για να μπορέσει αυτή η μελέτη να καταλήξει σε προσεγγίσεις που θα είναι άμεσα εφαρμόσιμες στην πράξη, δόθηκε έμφαση στην προσπάθεια αποφυγής του φαινομένου της υπερπροσαρμογής. Τα μοντέλα που δεν εμφανίζουν σημάδια υπερπροσαρμογής, δεν εξαρτώνται από το σύνολο δεδομένων με βάση το οποίο εκπαιδεύονται και κατά συνέπεια μπορούν να χρησιμοποιηθούν και να αποδώσουν το ίδιο καλά σε νέα δεδομένα. Οι προσεγγίσεις που υλοποιήθηκαν προέρχονταν είτε αυτούσιες από τη βιβλιογραφία, είτε έγινε προσαρμογή τους στα δεδομένα, είτε συνδυασμός αυτών. Ορισμένες από τις αρχιτεκτονικές αυτές σχεδιάστηκαν και υλοποιήθηκαν, αρχικά, για προβλήματα ταξινόμησης εγκεφαλικών όγκων σε πολλαπλές κλάσεις, σκοπός στην εργασία αυτή ήταν η αξιολόγηση των μεθόδων αυτών αναφορικά με την ικανότητα τους να γενικεύουν και να αποδίδουν είτε αυτούσιες είτε με προσαρμογές, σε άλλα προβλήματα ταξινόμησης εγκεφαλικών όγκων, όπως αυτό της δυαδικής ταξινόμησης που μελετάται στην παρούσα διπλωματική εργασία. Τα συνελκτικά νευρωνικά δίκτυα που προτείνονται, προέκυψαν από εκτενή μελέτη και πειραματισμό των υπερπαραμέτρων και του αριθμού των επιπέδων τους. Το σύνολο των δεδομένων που χρησιμοποιήθηκε παρουσίασε μία ελαφριά ανισορροπία, με την κλάση της παρουσίας όγκου να υπερεκπροσωπείται, παρόλα αυτά δεν χρησιμοποιήθηκε κάποια μέθοδος εξισορρόπησης των δύο κλάσεων. Η απόφαση αυτή σχετίζεται με την προσπάθεια κατασκευής και αξιολόγησης μεθόδων που θα ανταποκρίνονται στην πραγματικότητα και θα μπορούν άμεσα να χρησιμοποιηθούν στην κλινική πράξη, στην οποία σπάνια συναντώνται σύνολα δεδομένων με τον ίδιο αριθμό περιπτώσεων σε κάθε κλάση. Σε γενικές γραμμές, παρατηρήθηκε ότι τα μοντέλα που υλοποιήθηκαν με την τεχνική Transfer Learning κατέγραψαν υψηλότερες επιδόσεις, και στα τρία υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής, σε μικρότερο αριθμό εποχών εκπαίδευσης και ήταν λιγότερο επιρρεπή στην εμφάνιση υπερπροσαρμογής. Από αυτά, οι αρχιτεκτο-

νικές που χρησιμοποίησαν ως βάση το InceptionV3 κατέγραψαν τα πιο ικανοποιητικά αποτελέσματα φτάνοντας το 95% για τα accuracy, precision, recall, f1-score και το 96% για το specificity. Οι αρχιτεκτονικές που χρησιμοποίησαν ως βάση το ResNet50 κατέγραψαν, σε όλες τις προσεγγίσεις, τα χαμηλότερα αποτελέσματα με τις πιο υψηλές τιμές των μέτρων accuracy, precision, recall, f1-score να καταγράφονται στο 88% και του specificity στο 90%. Οι αρχιτεκτονικές που χρησιμοποίησαν ως βάση το MobileNetV2 κατέγραψαν εξαιρετικά αποτελέσματα στο σύνολο των υλοποιήσεων τους, τα οποία είχαν μικρές αποκλίσεις από αυτά του πρωτοπόρου InceptionV3. Παρατηρήθηκε λοιπόν μία συνέπεια στην απόδοση των μοντέλων, σε όλες τις προσεγγίσεις που ακολουθήθηκαν στην συγκεκριμένα εργασία, όπου δεδομένης της τήρησης των παραμέτρων ίδιων για όλα τα μοντέλα μας εμπορεύεται να συμπεράνουμε, ότι οφείλεται στην ειδική δομή ή αρχιτεκτονική καθενός από τα τρία προ εκπαιδευμένα μοντέλα που χρησιμοποιήθηκαν. Είναι γνωστό, πως κάθε μία από τις αρχιτεκτονικές αυτές έχει σχεδιαστεί με συγκεκριμένο τρόπο, ώστε να αντιμετωπίζει διαφορετικούς τύπους προβλημάτων αναγνώρισης μοτίβων και εκμάθησης χαρακτηριστικών από δεδομένα εικόνας, ενσωματώνοντας διάφορες τεχνικές και καινοτομίες που ενδέχεται να τα καθιστούν πιο αποτελεσματικά σε συγκεκριμένες εφαρμογές ή σύνολα δεδομένων. Επιπλέον παρατηρήθηκε ότι οι προσεγγίσεις που ενσωμάτωναν συνελκτικά στρώματα, αύξησαν την πολυπλοκότητα και απαιτήσαν μεγαλύτερο αριθμό εποχών για την εξασφάλιση υψηλής αποδοτικότητας. Τέλος, τα συνελκτικά νευρωνικά δίκτυα που κατασκευάστηκαν δεν κατόρθωσαν να αποδώσουν το ίδιο ικανοποιητικά με τα δίκτυα που χρησιμοποίησαν την τεχνική Transfer Learning. Παρατηρήθηκε όμως ότι η αύξηση της πολυπλοκότητας, με την προσθήκη επιπλέον στρωμάτων, βελτίωσε αισθητά τα αποτελέσματα τους χωρίς όμως να καταγράφουν αξιοσημείωτες επιδόσεις. Κατά πάσα πιθανότητα η περαιτέρω αύξηση των επιπέδων και της πολυπλοκότητας τους θα οδηγήσει σε καλύτερα αποτελέσματα, ωστόσο η εκπαίδευση από την αρχή ενός νευρωνικού δικτύου απαιτεί πόρους χρόνου και μνήμης. Μεταξύ των δύο συνελκτικών νευρωνικών δικτύων που κατασκευάστηκαν τα υψηλότερα ποσοστά καταγράφηκαν από αυτό των 10 επιπέδων με τα accuracy, precision, recall, f1-score στο 87% και το specificity στο 85%. Σημειώνεται πως στα δύο συνελκτικά νευρωνικά δίκτυα το callback EarlyStopping διέκοψε την εκπαίδευση τους στις 42 και 25 εποχές αντίστοιχα.

6.1 Προτάσεις μελλοντικής έρευνας και βελτίωσης

Στην παρούσα διπλωματική εργασία, με σκοπό την ταξινόμηση εικόνων MRI σύμφωνα με την απουσία ή την παρουσία εγκεφαλικού όγκου υλοποιήθηκε μία σειρά συνελκτικών νευρωνικών δικτύων. Οι προσεγγίσεις βασίστηκαν σε δύο βασικές μεθοδολογίες, τη χρήση της τεχνικής Transfer Learning και την εκπαίδευση, από την αρχή, συνελκτικών νευρωνικών δικτύων (CNNs). Οι προσεγγίσεις στις οποίες έγινε χρήση της τεχνικής Transfer Learning χρησιμοποίησαν ως βάση τρία προ εκπαιδευμένα μοντέλα: InceptionV3, ResNet50 και MobileNetV2. Τα αποτελέσματα της μελέτης που πραγματοποιήθηκε ανέδειξαν ως καταλληλότερες και πιο αποδοτικές τις προσεγγίσεις εκείνες που χρησιμοποίησαν την τεχνική Transfer Learning, καθώς αυτές κατέγραψαν υψηλότερες τιμές των μέτρων απόδοσης και κατόρθωσαν να εκπαιδευτούν επαρκώς στα δεδομένα που τους δόθηκαν σε λίγες εποχές

εκπαίδευσης. Παρόλα αυτά, έγινε φανερό πως η συγκεκριμένη δομή και αρχιτεκτονική καθενός από τα τρία προ εκπαιδευμένα μοντέλα που χρησιμοποιήθηκαν, επηρέασε σε μεγάλο βαθμό τα αποτελέσματα. Με βάση την παρατήρηση αυτή, προτείνεται η χρήση και η διενέργεια πειραμάτων με διαφορετικά προ εκπαιδευμένα μοντέλα όπως το MobileNetV3, το EfficientNetV2B2, το Xception κλπ. Τα μοντέλα που κατασκευάστηκαν, όπως προαναφέρθηκε, χρησιμοποιήθηκαν για την αντιμετώπιση ενός προβλήματος δυαδικής ταξινόμησης, στις κλάσεις απουσία και παρουσία όγκου. Η προτεινόμενη μεθοδολογία, μπορεί να επεκταθεί και να χρησιμοποιηθεί σε προβλήματα ταξινόμησης εγκεφαλικών όγκων σε πολλαπλές κλάσεις, όπως το γλοίωμα, το μηνιγγίωμα και ο όγκος τη υπόφυσης. Προτείνεται, επιπλέον, η χρήση της μεθοδολογίας που προτάθηκε στην παρούσα εργασία για την ανίχνευση και ταξινόμηση διαφορετικών όγκων, σε άλλα σημεία του σώματος, όπως για παράδειγμα ο όγκος του μαστού, των πνευμόνων κλπ. Η εργασία αυτή, παρά το γεγονός πως κατέγραψε εξαιρετικά αποτελέσματα χρησιμοποίησε ένα δημόσια διαθέσιμο σύνολο δεδομένων και τα αποτελέσματά της δεν έχουν επικυρωθεί σε πραγματική κλινική μελέτη. Προτείνεται, λοιπόν, η χρήση των μοντέλων σε πραγματικά κλινικά δεδομένα ή αν δεν υπάρχουν αρκετά η επαύξηση του συνόλου των δεδομένων που χρησιμοποιήθηκε με ένα μικρό σύνολο εικόνων MRI από πραγματικούς ασθενείς. Αναφορικά με τα συνελκτικά νευρωνικά δίκτυα που κατασκευάστηκαν και εκπαιδεύτηκαν από την αρχή προτείνεται ο επιπλέον πειραματισμός με τις τιμές των παραμέτρων και η προσθήκη επιπλέον στρωμάτων, καθώς παρατηρήθηκε καλύτερη προσαρμογή και απόδοση του CNN με τα 10 επίπεδα έναντι αυτού των 8. Η προσθήκη επιπλέον πολυπλοκότητας κατά πάσα πιθανότητα να οδηγήσει στην εμφάνιση του φαινομένου της υπερπροσαρμογής, δεδομένου του μικρού σχετικά συνόλου δεδομένων που έχει χρησιμοποιηθεί, για το λόγο αυτό προτείνεται να εξεταστεί το ενδεχόμενο συνδυασμού διαφορετικών διαθέσιμων συνόλων δεδομένων, ώστε να προκύψει ένα μεγαλύτερο ή η ενσωμάτωση τεχνικών image augmentation. Στην περίπτωση που ενσωματωθούν τεχνικές image augmentation, αυτό θα πρέπει να γίνει προσεκτικά δεδομένης της φύσης του προβλήματος που εξετάζεται. Στην ιατρική απεικόνιση η ακριβής θέση, μορφή και το μέγεθος του όγκου αποτελούν κρίσιμα χαρακτηριστικά για τη διάγνωση. Ορισμένες μέθοδοι image augmentation όπως η αναστροφή και η περιστροφή μπορεί να παράγουν εικόνες που δεν αντιπροσωπεύουν πραγματικές παθολογικές καταστάσεις, ενώ άλλες μπορεί να οδηγήσουν σε εικόνες χαμηλότερης ποιότητας προσθέτοντας θόρυβο ή διαστρεβλώνοντας σημαντικές λεπτομέρειες. Για την αποφυγή των προβλημάτων αυτών είναι σημαντικό να εφαρμοστούν τεχνικές image augmentation με προσοχή και επίγνωση των περιορισμών και των απαιτήσεων της συγκεκριμένης ιατρικής εφαρμογής, με βάση την ανάγκη διατήρησης της διαγνωστικής ακεραιότητας των εικόνων. Τέλος, προτείνεται, ο πειραματισμός με μεθόδους ανίχνευσης αντικειμένων(Object Detection) . Αλγόριθμοι όπως ο YOLO(You Only Look Once) και ο SSD(Single Shot Detector), μπορούν να εκπαιδευτούν για να εντοπίζουν την παρουσία και την τοποθεσία των όγκων μέσα σε μία εικόνα.

Βιβλιογραφία

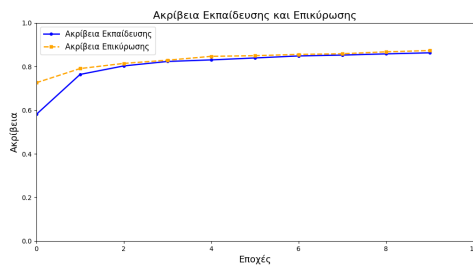
- [1] Brain Tumor. <https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor>.
- [2] Christie Adams, Joan Sullivan, and Todd W. Vitaz. Clinical Presentation of Brain Tumors. In Terry Lichtor, editor, *Molecular Considerations and Evolving Surgical Management Issues in the Treatment of Patients with a Brain Tumor*. InTech, March 2015.
- [3] Muhammad Afridi, Abhi Jain, Mariam Aboian, and Seyedmehdi Payabvash. Brain Tumor Imaging: Applications of Artificial Intelligence. *Seminars in Ultrasound, CT and MRI*, 43(2):153--169, April 2022.
- [4] Belal Amin, Romario Sameh Samir, Youssef Tarek, Mohammed Ahmed, Rana Ibrahim, Manar Ahmed, and Mohamed Hassan. Brain tumor multi classification and segmentation in MRI images using deep learning, June 2023. arXiv:2304.10039 [cs, eess].
- [5] Javeria Amin, Muhammad Sharif, Anandakumar Haldorai, Yasmin Mussarat, and Ramesh Nayak. Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex Intelligent Systems*, 8, 11 2021.
- [6] Sanjeev Arora, Aditya Bhaskara, Rong Ge, and Tengyu Ma. Provable Bounds for Learning Some Deep Representations, October 2013. arXiv:1310.6343 [cs, stat].
- [7] Gopinath Balaji, Ranit Sen, and Harsh Kirty. Detection and Classification of Brain tumors Using Deep Convolutional Neural Networks, August 2022. arXiv:2208.13264 [cs, eess].
- [8] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [9] Christopher M. Bishop and Hugh Bishop. *Deep Learning: Foundations and Concepts*. Springer International Publishing, Cham, 2024.
- [10] Sunanda Das, O. F. M. Riaz Rahman Aranya, and Nishat Nayla Labiba. Brain Tumor Classification Using Convolutional Neural Network. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1--5, Dhaka, Bangladesh, May 2019. IEEE.

- [11] Lisa M. DeAngelis. Brain Tumors. *New England Journal of Medicine*, 344(2):114--123, January 2001.
- [12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248--255, 2009.
- [13] Prमित Dutta, Khaleda Akhter Sathi, and Md Saiful Islam. Multi-Classification of Brain Tumor Images Using Transfer Learning Based Deep Neural Network, June 2022. arXiv:2206.08543 [cs, eess].
- [14] Dmytro Filatov and Ghulam Nabi Ahmad Hassan Yar. Brain Tumor Diagnosis and Classification via Pre-Trained Convolutional Neural Networks, July 2022. arXiv:2208.00768 [cs, eess].
- [15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts, 2016.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. arXiv:1512.03385 [cs].
- [17] Javed Hossain, Md Touhidul Islam, and Md Taufiqul Haque Khan Tusar. Streamlining Brain Tumor Classification with Custom Transfer Learning in MRI Images. In *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, pages 522--526, May 2023. arXiv:2310.13108 [cs, eess].
- [18] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017.
- [19] Hassan Ali Khan, Wu Jue, Muhammad Mushtaq, and Muhammad Umer Mushtaq. Brain tumor classification in MRI image using convolutional neural network. *Mathematical biosciences and engineering: MBE*, 17(5):6203--6216, September 2020.
- [20] Md. Saikat Islam Khan, Anichur Rahman, Tanoy Debnath, Md. Razaul Karim, Mostofa Kamal Nasir, Shahab S. Band, Amir Mosavi, and Iman Dehzangi. Accurate brain tumor detection using deep convolutional neural network. *Computational and Structural Biotechnology Journal*, 20:4733--4745, August 2022.
- [21] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].
- [22] Abu Kaisar Mohammad Masum, Nusrat Badhon, S. M. Saiful Islam Badhon, Nushrat Jahan Ria, Sheikh Abujar, Muntaser Mansur Syed, and Naveed Mahmud. Comparative Evaluation of Transfer Learning for Classification of Brain Tumor Using MRI, September 2023. arXiv:2310.02270 [cs, eess].

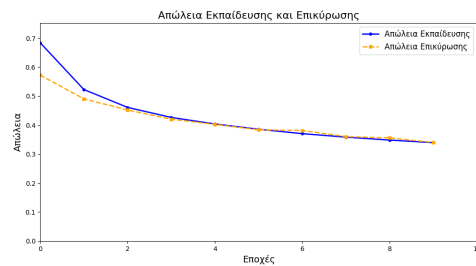
- [23] Amr Mohamed, Mahmoud Rabea, Aya Sameh, and Ehab Kamal. Brain Tumor Radiogenomic Classification, January 2024. arXiv:2401.09471 [cs, eess].
- [24] Vasileios E. Papageorgiou, Pantelis Dogoulis, and Dimitrios-Panagiotis Papageorgiou. A convolutional neural network of low complexity for tumor anomaly detection. volume 696, pages 973--983. 2024. arXiv:2301.09861 [eess, stat].
- [25] J. Patterson and A. Gibson. *Deep Learning: A Practitioner's Approach*. O'Reilly, 2017.
- [26] Nader Pouratian and David Schiff. Management of Low-Grade Glioma. *Current Neurology and Neuroscience Reports*, 10(3):224--231, May 2010.
- [27] Jing Qian, Michael G. Herman, Debra H. Brinkmann, Nadia N. Laack, Bradley J. Kemp, Christopher H. Hunt, Val Lowe, and Deanna H. Pafundi. Prediction of MGMT Status for Glioblastoma Patients Using Radiomics Feature Extraction From 18F-DOPA-PET Imaging. *International Journal of Radiation Oncology*Biography*Physics*, 108(5):1339--1346, December 2020.
- [28] Zenghui Qian, Yiming Li, Yongzhi Wang, Lianwang Li, Runtong Li, Kai Wang, Shaowu Li, Ke Tang, Chuanbao Zhang, Xing Fan, Baoshi Chen, and Wenbin Li. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Letters*, 451:128--135, June 2019.
- [29] Allan H. Ropper, Martin A. Samuels, and Joshua Klein. *Adams and Victor's principles of neurology*. McGraw-Hill Education, New York, eleventh edition edition, 2019.
- [30] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018.
- [31] Roger Stupp, Warren P. Mason, Martin J. Van Den Bent, Michael Weller, Barbara Fisher, Martin J.B. Taphoorn, Karl Belanger, Alba A. Brandes, Christine Marosi, Ulrich Bogdahn, Jürgen Curschmann, Robert C. Janzer, Samuel K. Ludwin, Thierry Gorlia, Anouk Allgeier, Denis Lacombe, J. Gregory Cairncross, Elizabeth Eisenhauer, and René O. Mirimanoff. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine*, 352(10):987--996, March 2005.
- [32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions, September 2014. arXiv:1409.4842 [cs].

- [33] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818--2826, Las Vegas, NV, USA, June 2016. IEEE.
- [34] Tahia Tazin, Sraboni Sarker, Punit Gupta, Fozayel Ibn Ayaz, Sumaia Islam, Mohammad Monirujjaman Khan, Sami Bourouis, Sahar Ahmed Idris, and Hammam Alshazly. A Robust and Novel Approach for Brain Tumor Classification Using Convolutional Neural Network. *Computational Intelligence and Neuroscience*, 2021:1--11, December 2021.
- [35] Σωτήριος Μπερσίμης, Γεώργιος Μπάρτζης, Γεώργιος Παπαδάκης, Αθανάσιος Σαχλάς. *Εφαρμοσμένη Στατιστική και Στατιστική Μηχανική Μάθηση με χρήση των IBM SPSS Statistics, R, Python*. Εκδόσεις Τζιόλα, 2021.

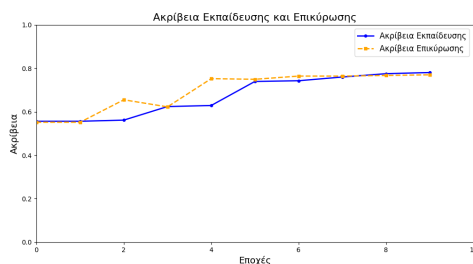
Παράρτημα: Γραφήματα Καμπυλών Ακρίβειας και Απώλειας



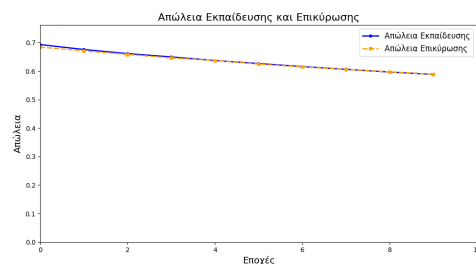
Σχήμα 1: InceptionV3: Καμπύλες Ακρίβειας



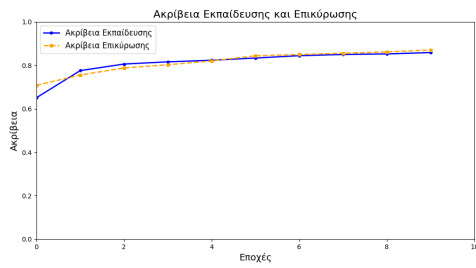
Σχήμα 2: InceptionV3: Καμπύλες Απώλειας



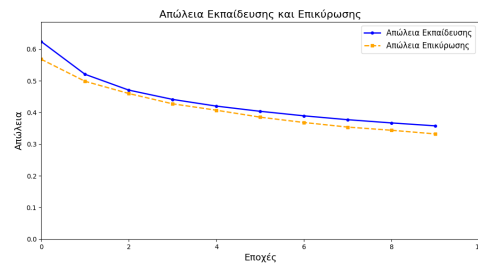
Σχήμα 3: ResNet50: Καμπύλες Ακρίβειας



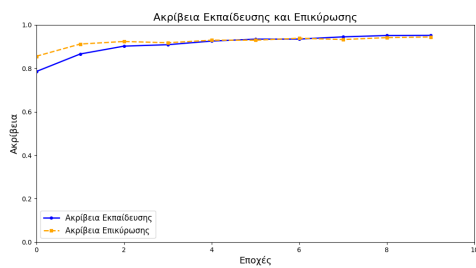
Σχήμα 4: ResNet50: Καμπύλες Απώλειας



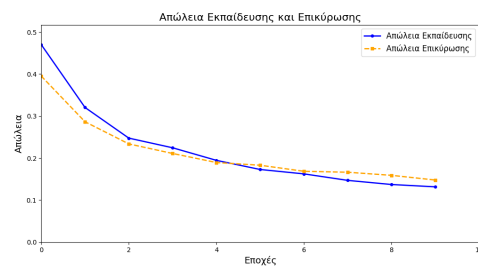
Σχήμα 5: MobileNetV2: Καμπύλες Ακρίβειας



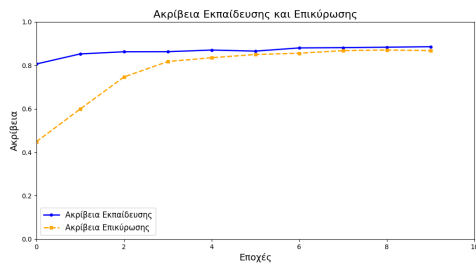
Σχήμα 6: MobileNetV2: Καμπύλες Απώλειας



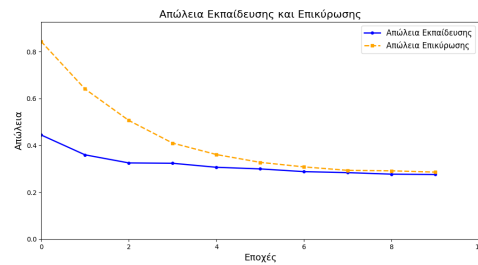
Σχήμα 7: InceptionV3+ANN1: Καμπύλες Ακρίβειας



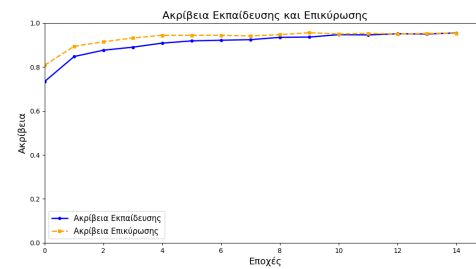
Σχήμα 8: InceptionV3+ANN1: Καμπύλες Απώλειας



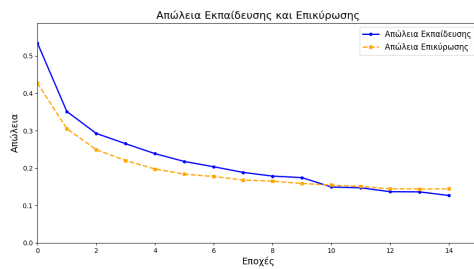
Σχήμα 9: ResNet50+ANN1: Καμπύλες Ακρίβειας



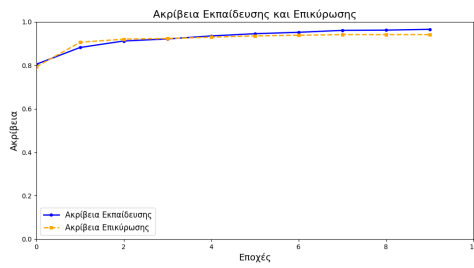
Σχήμα 10: ResNet50+ANN1: Καμπύλες Απώλειας



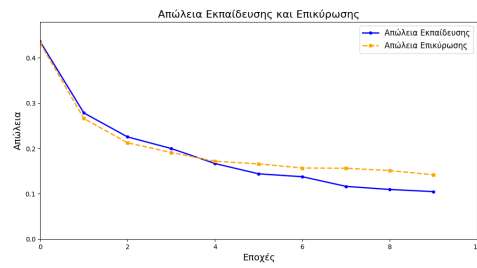
Σχήμα 11: MobileNetV2+ANN1: Καμπύλες Ακρίβειας



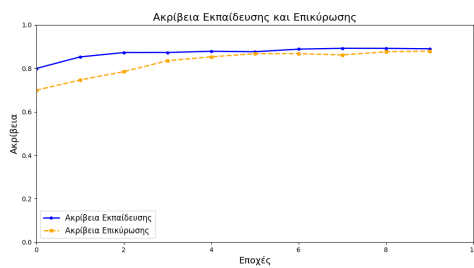
Σχήμα 12: MobileNetV2+ANN1: Καμπύλες Απώλειας



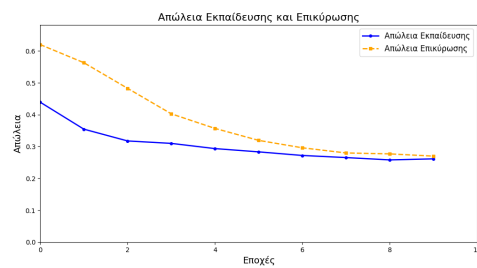
Σχήμα 13: InceptionV3+ANN2: Καμπύλες Ακρίβειας



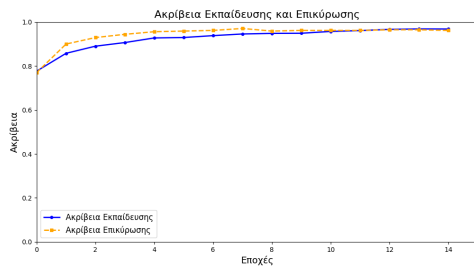
Σχήμα 14: InceptionV3+ANN2: Καμπύλες Απώλειας



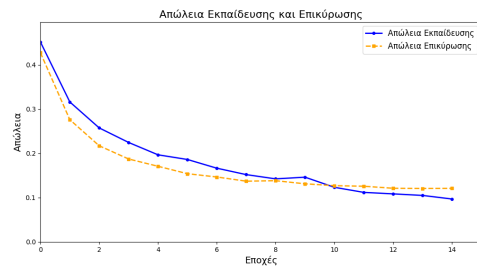
Σχήμα 15: ResNet50+ANN2: Καμπύλες Ακρίβειας



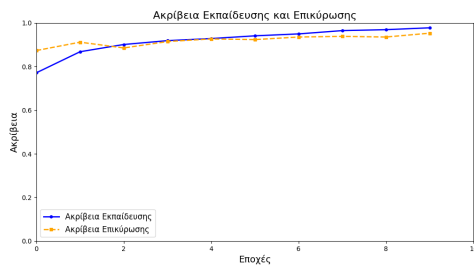
Σχήμα 16: ResNet50+ANN2: Καμπύλες Απώλειας



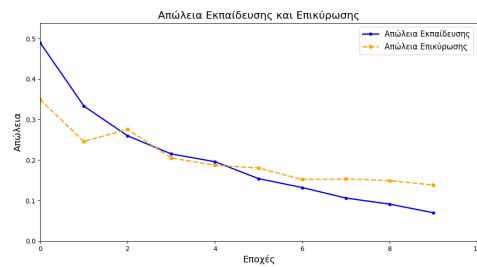
Σχήμα 17: MobileNetV2+ANN2: Καμπύλες Ακρίβειας



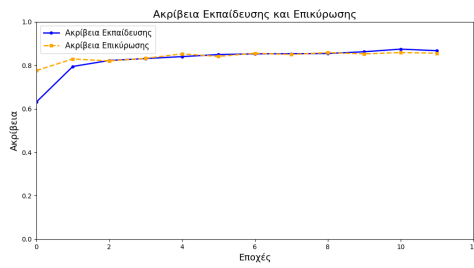
Σχήμα 18: MobileNetV2+ANN2: Καμπύλες Απώλειας



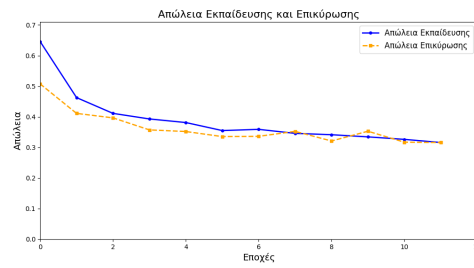
Σχήμα 19: InceptionV3+CNN1: Καμπύλες Ακρίβειας



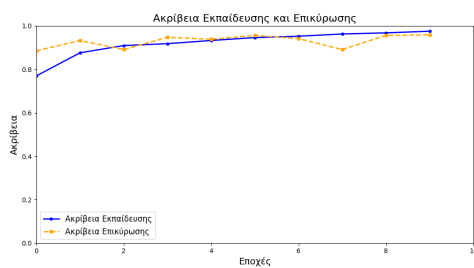
Σχήμα 20: InceptionV3+CNN1: Καμπύλες Απώλειας



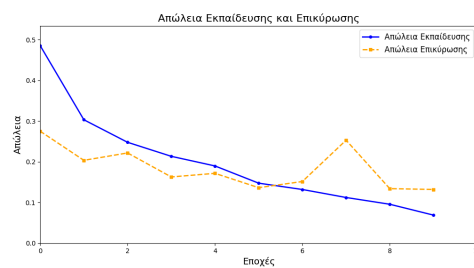
Σχήμα 21: ResNet50+CNN1: Καμπύλες Ακρίβειας



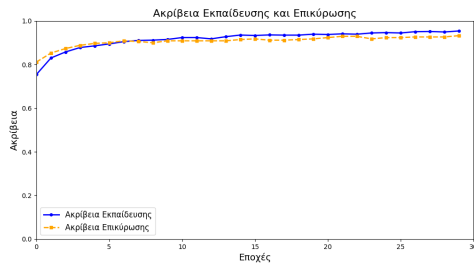
Σχήμα 22: ResNet50+CNN1: Καμπύλες Απώλειας



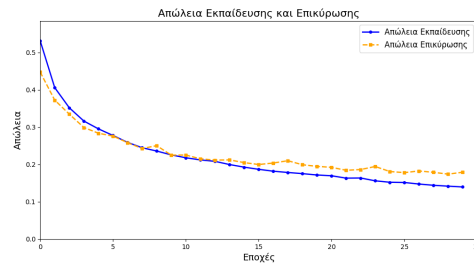
Σχήμα 23: MobileNetV2+CNN1: Καμπύλες Ακρίβειας



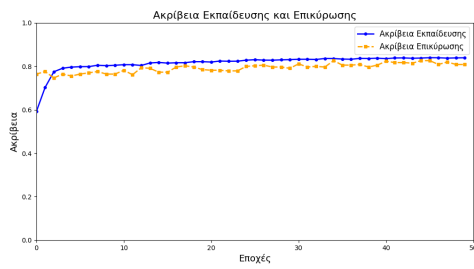
Σχήμα 24: MobileNetV2+CNN1: Καμπύλες Απώλειας



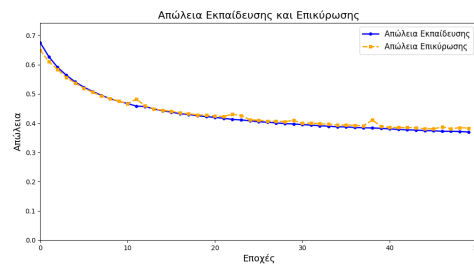
Σχήμα 25: InceptionV3+CNN2: Καμπύλες Ακρίβειας



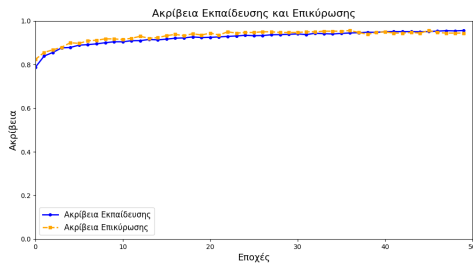
Σχήμα 26: InceptionV3+CNN2: Καμπύλες Απώλειας



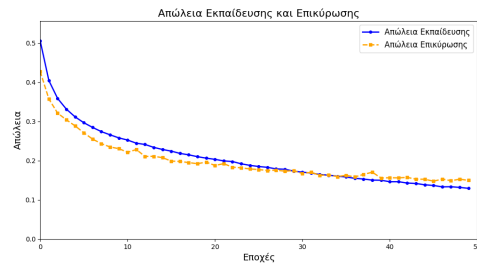
Σχήμα 27: ResNet50+CNN2: Καμπύλες Ακρίβειας



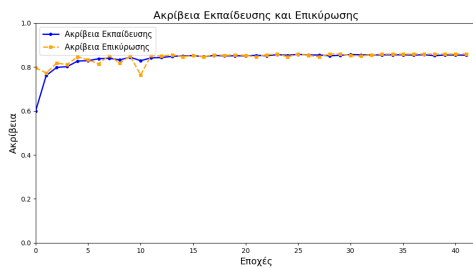
Σχήμα 28: ResNet50+CNN2: Καμπύλες Απώλειας



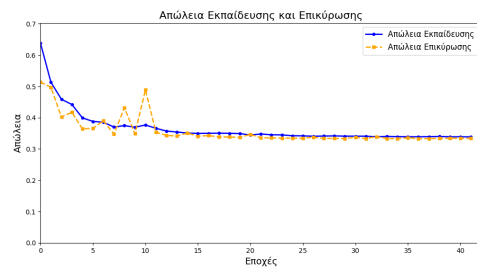
Σχήμα 29: MobileNetV2+CNN2: Καμπύλες Ακρίβειας



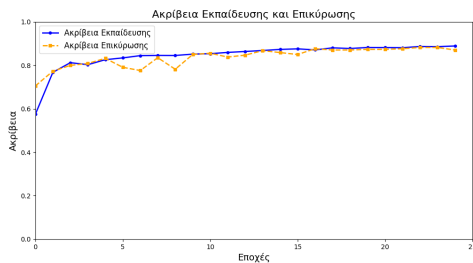
Σχήμα 30: MobileNetV2+CNN2: Καμπύλες Απώλειας



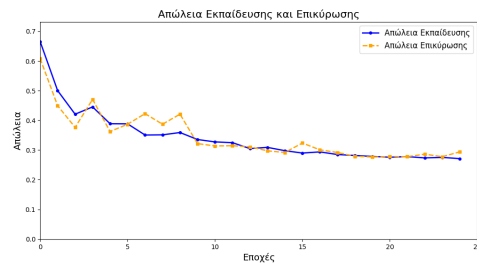
Σχήμα 31: CNN 8 layers: Καμπύλες Ακρίβειας



Σχήμα 32: CNN 8 layers: Καμπύλες Απώλειας



Σχήμα 33: CNN 10 layers: Καμπύλες Ακρίβειας



Σχήμα 34: CNN 10 layers: Καμπύλες Απώλειας