



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Κατάτμηση και Κατηγοριοποίηση Εγκεφαλικών Όγκων
με χρήση Βαθιάς Μάθησης**

Θεόδωρος Κ. Μπαρτσώκας

Επιβλέπων: Παναγιώτης Σταματόπουλος, Επίκουρος Καθηγητής

ΑΘΗΝΑ

ΟΚΤΩΒΡΙΟΣ 2021

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Κατάτμηση και Κατηγοριοποίηση Εγκεφαλικών Όγκων με χρήση Βαθιάς Μάθησης

Θεόδωρος Κ. Μπαρτσώκας

A.M.: 1115201700096

ΕΠΙΒΛΕΠΟΝΤΕΣ: Παναγιώτης Σταματόπουλος, Επίκουρος Καθηγητής

ΠΕΡΙΛΗΨΗ

Σκοπός αυτής της εργασίας είναι η ανάπτυξη μεθόδων για τη διάγνωση εγκεφαλικών όγκων χρησιμοποιώντας τεχνικές βαθιάς μάθησης. Αρχικά, παρατίθεται το απαραίτητο θεωρητικό υπόβαθρο που αφορά θεμελιώδεις έννοιες και αρχιτεκτονικές των νευρωνικών δικτύων. Κατόπιν, ερευνώνται ποικίλες μέθοδοι διάγνωσης εγκεφαλικών όγκων από συναφείς εργασίες του παρελθόντος. Στην προτεινόμενη προσέγγιση που παρουσιάζεται μετέπειτα, χρησιμοποιούνται τα πλεονεκτήματα της μεταφοράς μάθησης για τη δημιουργία πέντε ανεξάρτητων μοντέλων νευρωνικών δικτύων. Ειδικότερα, η δημιουργία ενός εξ αυτών βασίζεται στην αρχιτεκτονική Mask R-CNN και χρησιμοποιείται για την κατάτμηση των εγκεφαλικών όγκων. Τα υπόλοιπα τέσσερα μοντέλα χρησιμοποιούνται για την κατηγοριοποίηση των εγκεφαλικών όγκων και προκύπτουν από τον συνδυασμό των δύο παραμετροποιήσεων VGG-16 και VGG-19 της αρχιτεκτονικής VGGNet, με δύο τεχνικές μεταφοράς μάθησης. Αυτά τα πέντε μοντέλα αποτελούν τη βάση για τις δύο διαφορετικών προσεγγίσεις που ακολουθούνται. Στην πρώτη προσέγγιση, οι εγκεφαλικοί όγκοι κατηγοριοποιούνται χρησιμοποιώντας τις αρχικές εικόνες του σετ δεδομένων, τέσσερις ανεξάρτητες φορές, μία φορά για κάθε μοντέλο κατηγοριοποίησης. Στην δεύτερη προσέγγιση, πρώτα εκτελείται κατάτμηση των όγκων και στη συνέχεια εκτελείται κατηγοριοποίηση χρησιμοποιώντας μόνο το κατατμημένο μέρος του εγκεφαλικού όγκου. Η διαδικασία της δεύτερης προσέγγισης εφαρμόζεται τέσσερις ανεξάρτητες φορές, μία φορά για κάθε πιθανό συνδυασμό του μοντέλου κατάτμησης με τα μοντέλα κατηγοριοποίησης. Οι δύο προσεγγίσεις παράγουν συνολικά οκτώ διαφορετικά αποτελέσματα, τα οποία καθιστούν ξεκάθαρα τα πλεονεκτήματα και μειονεκτήματα κάθε επιλογής, καθώς και την προοπτική που έχει η βαθιά μάθηση να αποτελέσει καθοριστικό παράγοντα στη διάγνωση εγκεφαλικών όγκων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Όραση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Κατάτμηση, Κατηγοριοποίηση, Διάγνωση Εγκεφαλικού Όγκου

ABSTRACT

The purpose of this paper is the development of methods for diagnosing brain tumors using deep learning techniques. First of all, the necessary theoretical background concerning fundamental concepts and architectures of neural networks are presented. Then, various methods for diagnosing brain tumors from related work of the past are investigated. In the proposed approach that is presented afterwards, the advantages of learning transfer are used to create five independent neural network models. In particular, the creation of one of them is based on the Mask R-CNN architecture and is used for brain tumor segmentation. The other four models are used for brain tumor classification and result from the combination of two VGGNet architecture's configurations, VGG-16 and VGG-19, with two learning transfer techniques. These five models form the basis of the two different approaches taken. In the first approach, brain tumors are classified using the original data in the data set, four independent times, once for each classification model. In the second approach, tumors are first segmented and then classified using only the segmented part of the brain tumor. The procedure of the second approach is applied four independent times, once for each possible combination of the segmentation model with the classification models. The two approaches produce a total of eight different results, which elucidate the advantages and disadvantages of each option, as well as the prospect of deep learning becoming a key factor in the diagnosis of brain tumors.

SUBJECT AREA: Computer Vision

KEYWORDS: Deep Learning, Convolutional Neural Networks, Segmentation, Classification, Brain Tumor Diagnosis

Στους γονείς μου, Βίκυ και Κώστα.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον κύριο Παναγιώτη Σταματόπουλο για την υποστήριξη που μου παρείχε καθ' όλη τη διάρκεια εκπόνησης της παρούσας εργασίας.

ΠΕΡΙΕΧΟΜΕΝΑ

| | |
|--|-----------|
| 1. ΕΙΣΑΓΩΓΗ | 13 |
| 2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ | 15 |
| 2.1 Τεχνητός νευρώνας (Artificial neuron) | 15 |
| 2.1.1 McCulloch-Pitts νευρώνας (McCulloch-Pitts neuron) | 15 |
| 2.1.2 Αισθητήρας (Perceptron) | 16 |
| 2.1.3 Πόλωση (Bias) | 17 |
| 2.2 Συνάρτηση ενεργοποίησης (Activation function) | 19 |
| 2.2.1 Βηματική συνάρτηση (Heaviside Step Function) | 21 |
| 2.2.2 Γραμμική συνάρτηση (Linear function) | 22 |
| 2.2.3 Λογιστική συνάρτηση (Logistic function) | 23 |
| 2.2.4 Κανονικοποιημένη εκθετική συνάρτηση (Softmax) | 23 |
| 2.2.5 Υπερβολική εφαπτομένη (Tanh function) | 24 |
| 2.2.6 Ανορθωμένη Γραμμική Μονάδα (Rectified Linear Unit – ReLU) | 24 |
| 2.2.7 Leaky ReLU | 25 |
| 2.3 Στρώματα (Layers) | 26 |
| 2.4 Συνάρτηση κόστους (Cost function) | 27 |
| 2.4.1 Μέσο απόλυτο σφάλμα (Mean absolute error) | 28 |
| 2.4.2 Μέσο τετραγωνικό σφάλμα (Mean squared error) | 28 |
| 2.4.3 Ριζικό μέσο τετραγωνικό σφάλμα (Root mean squared error) | 29 |
| 2.4.4 Δυαδική διασταυρωμένη εντροπία (Binary cross entropy) | 29 |
| 2.4.5 Κατηγορική διασταυρωμένη εντροπία (Categorical cross entropy) | 29 |
| 2.5 Κάθοδος με βάση την κλίση (Gradient descent) | 30 |
| 2.6 Οπισθοδιάδοση (Backpropagation) | 32 |
| 2.7 Αναδρομικό νευρωνικό δίκτυο (Recurrent neural network – RNN) | 34 |
| 2.8 Νευρωνικό δίκτυο μακροχρόνιας και βραχυχρόνιας μνήμης (Long short-term memory - LSTM) | 37 |
| 2.9 Συνελικτικό νευρωνικό δίκτυο (Convolutional neural network – CNN) | 40 |
| 3. ΣΥΝΑΦΕΙΣ ΕΡΓΑΣΙΕΣ | 46 |
| 3.1 Μέθοδοι βασισμένες στην κατωφλίωση (Thresholding) | 46 |
| 3.2 Μέθοδοι βασισμένες σε περιοχές της εικόνας (Region-based) | 46 |
| 3.3 Μέθοδοι βασισμένες στη στατιστική (Statistics-based) | 47 |
| 3.3.1 Μαρκοβιανά Τυχαία Πεδία (Markov Random Fields) | 47 |
| 3.3.2 Γκαουσιανό μοντέλο ανάμειξης (Gaussian mixture model) | 48 |

| | | |
|-----------|---|-----------|
| 3.4 | Μέθοδοι βασισμένες στη βαθιά μάθηση (Deep learning-based) | 48 |
| 4. | ΠΡΟΤΕΙΝΟΜΕΝΗ ΠΡΟΣΕΓΓΙΣΗ | 50 |
| 4.1 | Περιγραφή προβλήματος (Problem description) | 50 |
| 4.2 | Προκλήσεις Προβλήματος | 50 |
| 4.3 | Μεταφορά μάθησης (Transfer learning) | 52 |
| 4.4 | Σετ δεδομένων (Dataset) | 53 |
| 4.5 | Προεπεξεργασία δεδομένων (Data preprocessing) | 54 |
| 4.6 | Επαύξηση δεδομένων | 55 |
| 4.7 | VGGNet | 57 |
| 4.8 | Mask R-CNN | 59 |
| 4.9 | Μετρικές (Metrics) | 63 |
| 4.9.1 | Μετρικές μοντέλων κατηγοριοποίησης | 63 |
| 4.9.2 | Μετρικές μοντέλου κατάτμησης | 66 |
| 4.10 | Εκπαίδευση | 67 |
| 4.10.1 | Εκπαίδευση μοντέλων κατηγοριοποίησης | 67 |
| 4.10.2 | Εκπαίδευση μοντέλων κατάτμησης | 77 |
| 4.11 | Αποτελέσματα | 79 |
| 5. | ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ | 82 |
| 5.1 | Συμπεράσματα | 82 |
| 5.2 | Μελλοντικές εργασίες | 83 |
| | ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ | 84 |
| | ΒΙΒΛΙΟΓΡΑΦΙΑ | 86 |

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

| | | |
|------|---|----|
| 2.1 | Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα κατά το μοντέλο McCulloch και Pitts. | 16 |
| 2.2 | Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα κατά το μοντέλο Perceptron. | 17 |
| 2.3 | Παράδειγμα συνάρτησης εξόδου ενός νευρώνα χωρίς τη χρήση πόλωσης. | 18 |
| 2.4 | Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα χρησιμοποιώντας τον επικρατέστερο συμβολισμό. | 19 |
| 2.5 | Απεικόνιση της πλήρους λειτουργίας ενός τεχνητού νευρώνα. | 21 |
| 2.6 | Απεικόνιση της βηματικής συνάρτησης ενεργοποίησης και της παραγώγου της. | 22 |
| 2.7 | Απεικόνιση της γραμμικής συνάρτησης ενεργοποίησης και της παραγώγου της. | 22 |
| 2.8 | Απεικόνιση της λογιστικής συνάρτησης ενεργοποίησης και της παραγώγου της. | 23 |
| 2.9 | Απεικόνιση της συνάρτησης ενεργοποίησης ReLU και της παραγώγου της. | 24 |
| 2.10 | Απεικόνιση της συνάρτησης ενεργοποίησης ReLU και της παραγώγου της. | 25 |
| 2.11 | Απεικόνιση της συνάρτησης Leaky ReLU και της παραγώγου της. | 26 |
| 2.12 | Απεικόνιση της σύγκλισης του αλγορίθμου της καθόδου με βάση την κλίση. | 30 |
| 2.13 | Παράδειγμα δομής αναδρομικού νευρωνικού δικτύου. | 35 |
| 2.14 | Απεικόνιση δύο αναπαραστάσεων ενός νευρώνα της αρχιτεκτονικής RNN. | 36 |
| 2.15 | Απεικόνιση ενός κελιού RNN. | 36 |
| 2.16 | Απεικόνιση ενός κελιού LSTM. | 38 |
| 2.17 | Παράδειγμα εφαρμογής της συνέλιξης. | 41 |
| 2.18 | Παράδειγμα εφαρμογής της μέγιστης συγκέντρωσης. | 42 |
| 2.19 | Παράδειγμα εφαρμογής της μέσης συγκέντρωσης. | 43 |
| 2.20 | Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο μετατοπισμένο σε διαφορετικές κατευθύνσεις. | 43 |
| 2.21 | Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο με διαφορετική περιστροφή. | 43 |
| 2.22 | Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο με διαφορετικό μέγεθος. | 44 |
| 2.23 | Παράδειγμα του τρόπου εξαγωγής χαρακτηριστικών ανά συνελκτικό στρώμα. | 44 |
| 2.24 | Παράδειγμα αρχιτεκτονικής συνελκτικού νευρωνικού δικτύου. | 45 |
| 4.1 | Διάγραμμα της πρώτης προσέγγισης. | 50 |
| 4.2 | Διάγραμμα της δεύτερης προσέγγισης. | 50 |
| 4.3 | Παράδειγμα δύο ίδιου είδους όγκων με διαφορετικά χαρακτηριστικά. | 51 |
| 4.4 | Παράδειγμα εφαρμογής CLAHE. | 55 |
| 4.5 | Παράδειγμα επαύξησης των δεδομένων στην πρώτη προσέγγιση. | 56 |
| 4.6 | Παράδειγμα επαύξησης των δεδομένων για το μοντέλο κατάτμησης της δεύτερης προσέγγισης. | 56 |
| 4.7 | Παράδειγμα επαύξησης δεδομένων για το μοντέλο κατηγοριοποίησης της δεύτερης προσέγγισης. | 57 |
| 4.8 | Σχηματική απεικόνιση της παραμετροποίησης VGG-16. | 58 |
| 4.9 | Σχηματική απεικόνιση της παραμετροποίησης VGG-19. | 59 |
| 4.10 | Διάγραμμα κύριων βημάτων αρχιτεκτονικής Mask R-CNN. | 60 |

| | |
|---|----|
| 4.11 Αρχιτεκτονική Δικτύου Πρότασης Περιοχών. | 61 |
| 4.12 Απεικόνιση εφαρμογής RoI Align, σε υψηλό επίπεδο. | 62 |
| 4.13 Λεπτομερέστερη αναπαράσταση των τελικών βημάτων παραγωγής των τριών εξόδων της αρχιτεκτονικής Mask R-CNN. | 62 |
| 4.14 Αναπαράσταση ενός πίνακα σύγχυσης. | 64 |
| 4.15 Απεικόνιση λειτουργίας της μετρικής IoU. | 67 |
| 4.16 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 ως εξαγωγέα χαρακτηριστικών, για την πρώτη προσέγγιση. Συνολικά 94 εποχές εκπαίδευσης. | 70 |
| 4.17 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 με fine-tuning, για την πρώτη προσέγγιση. Συνολικά 73 εποχές εκπαίδευσης. | 71 |
| 4.18 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-19 ως εξαγωγέα χαρακτηριστικών, για την πρώτη προσέγγιση. Συνολικά 108 εποχές εκπαίδευσης. | 72 |
| 4.19 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-19 με fine-tuning, για την πρώτη προσέγγιση. Συνολικά 99 εποχές εκπαίδευσης. | 73 |
| 4.20 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 ως εξαγωγέα χαρακτηριστικών, για τη δεύτερη προσέγγιση. Συνολικά 136 εποχές εκπαίδευσης. | 74 |
| 4.21 Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 με fine-tuning, για τη δεύτερη προσέγγιση. Συνολικά 78 εποχές εκπαίδευσης. | 75 |
| 4.22 Γραφήματα της εκπαίδευσης του μοντέλου που είναι βασισμένο στην παραμετροποίηση VGG-19 ως εξαγωγέα χαρακτηριστικών, για τη δεύτερη προσέγγιση. Συνολικά 32 εποχές εκπαίδευσης. | 76 |
| 4.23 Γραφήματα της εκπαίδευσης του μοντέλου που είναι βασισμένο στην παραμετροποίηση VGG-19 με fine-tuning, για τη δεύτερη προσέγγιση. Συνολικά 87 εποχές εκπαίδευσης. | 77 |
| 4.24 Γράφημα των συναρτήσεων απώλειας του Δικτύου Πρότασης Περιοχών για το σετ εκπαίδευσης. | 78 |
| 4.25 Γράφημα των συναρτήσεων απώλειας του συνολικού μοντέλου Mask R-CNN για το σετ εκπαίδευσης. | 78 |
| 4.26 Γράφημα των συναρτήσεων απώλειας του Δικτύου Πρότασης Περιοχών για το σετ επικύρωσης. | 79 |
| 4.27 Γράφημα της εκπαίδευσης συνολικά του μοντέλου Mask R-CNN για το σετ επικύρωσης. | 79 |
| 4.28 Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.95. | 80 |
| 4.29 Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.89. | 80 |
| 4.30 Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.93. | 81 |

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

| | | |
|-----|--|----|
| 4.1 | Λεπτομέρειες της παραμετροποίησης VGG-16. | 58 |
| 4.2 | Λεπτομέρειες της παραμετροποίησης VGG-19. | 59 |
| 4.3 | Τελική μορφή αρχιτεκτονικής των μοντέλων που βασίζονται στην παραμετροποίηση VGG-16. | 68 |
| 4.4 | Τελική μορφή αρχιτεκτονικής των μοντέλων που βασίζονται στην παραμετροποίηση VGG-19. | 69 |
| 4.5 | Αποτελέσματα πρώτης προσέγγισης. | 81 |
| 4.6 | Αποτελέσματα δεύτερης προσέγγισης. | 81 |

ΠΡΟΛΟΓΟΣ

Η παρούσα εργασία έχει ιδιαίτερη σημασία για μένα καθώς το πέρας της σηματοδοτεί την ολοκλήρωση των προπτυχιακών σπουδών μου στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Ήταν μεγάλη χαρά που είχα τη δυνατότητα να έρθω σε επαφή με την έρευνα και να μελετήσω ένα τόσο ενδιαφέρον θέμα, μέσω αυτής της εργασίας.

1. ΕΙΣΑΓΩΓΗ

Ο καρκίνος είναι μια ασθένεια η οποία έχει επηρεάσει, άμεσα ή έμμεσα, ένα μεγάλο μέρος του παγκόσμιου πληθυσμού. Το 2020 καταγράφηκαν, κατά προσέγγιση, 19.3 εκατομμύρια νέα κρούσματα και 10 εκατομμύρια θάνατοι σχετιζόμενοι με τον καρκίνο, καθιστώντας την ασθένεια αυτή μία από τις συχνότερες αιτίες θανάτου ανά τον κόσμο. Αξιοσημείωτη ιδιότητα του καρκίνου αποτελεί η ποικιλομορφία που τον διακατέχει με μία από τις μορφές του να είναι, δυνητικά, ο όγκος εγκεφάλου.

Ένας εγκεφαλικός όγκος είναι μία μάζα ανωμάλων κυττάρων τα οποία πολλαπλασιάζονται ανεξέλεγκτα, ενώ συχνά ο λόγος εμφάνισής του δεν είναι ξεκάθαρος. Συνολικά υπάρχουν πάνω από 130 είδη εγκεφαλικών όγκων με τον κάθε ένα να χαρακτηρίζεται από ένα διαφορετικό συνδυασμό μορφολογικών χαρακτηριστικών, όπως είναι το μέγεθος και το σχήμα. Μάλιστα, συχνό φαινόμενο είναι ακόμα και εγκεφαλικοί όγκοι ίδιου είδους να έχουν διαφορετικά μορφολογικά χαρακτηριστικά. Κάθε εγκεφαλικός όγκος, εκτός από το είδος που ανήκει, ανήκει και σε δύο κατηγορίες από τις οποίες μπορεί να χαρακτηριστεί ως πρωτογενής ή δευτερογενής και καλοήθης ή κακοήθης. Ένας πρωτογενής εγκεφαλικός όγκος αποτελείται αποκλειστικά από εγκεφαλικά κύτταρα και έχει ξεκινήσει την ανάπτυξή του στον εγκέφαλο. Από την άλλη πλευρά, ένας δευτερογενής εγκεφαλικός όγκος είναι το αποτέλεσμα μεταφοράς καρκινικών κυττάρων από άλλο μέρος του σώματος στον εγκέφαλο, τα οποία δημιουργούν τον όγκο στον εγκέφαλο. Ο όγκος που δημιουργείται σε αυτή την περίπτωση δεν αποτελείται από εγκεφαλικά κύτταρα αλλά αποκλειστικά από κύτταρα που ανήκουν στο άλλο μέρος του σώματος. Ένας καλοήθης εγκεφαλικός όγκος δεν θεωρείται καρκίνος καθώς αναπτύσσεται αργά, έχει σαφή όρια και τα κύτταρά του δε μπορούν να μεταφερθούν σε άλλο μέρος του σώματος δημιουργώντας έναν δευτεροπαθή όγκο σε αυτό. Βέβαια, η τοποθεσία στην οποία έχει αναπτυχθεί ενδέχεται να παρεμβαίνει με άλλα κύτταρα του εγκεφάλου που αντιστοιχούν σε ζωτικής σημασίας λειτουργίες του ανθρώπου, δημιουργώντας σημαντικά προβλήματα στην καθημερινότητά του. Αντιθέτως, ένας κακοήθης όγκος εγκεφάλου αποτελείται από καρκινικά κύτταρα τα οποία αναπτύσσονται ραγδαία, δεν έχουν σαφή όρια και αποτελούν μεγάλη απειλή για τη ζωή του ανθρώπου. Μάλιστα, ένας δευτεροπαθής εγκεφαλικός όγκος είναι πάντα κακοήθης.

Τα παραπάνω γεγονότα καθιστούν ξεκάθαρη τη σημαντικότητα της έγκαιρης και σωστής διάγνωσης ενός εγκεφαλικού όγκου. Μία διάγνωση ενός όγκου εγκεφάλου μπορεί να είναι είτε επεμβατική είτε μη επεμβατική. Ο επεμβατικός τρόπος διάγνωσης ονομάζεται βιοψία και απαιτεί τη λήψη ενός τμήματος του ήδη ανεπτυγμένου όγκου, το οποίο στη συνέχεια εξετάζεται με τη χρήση μικροσκοπίου. Η βιοψία είναι μεν ο πιο αξιόπιστος τρόπος διάγνωσης ενός εγκεφαλικού όγκου, αλλά είναι ακριβός και η διαδικασία περάτωσης του ενδέχεται να καταπονεί σωματικά και ψυχικά τον άνθρωπο στον οποίο εφαρμόζεται. Ένας από τους μη επεμβατικούς τρόπους διάγνωσης ενός εγκεφαλικού όγκου είναι η τομογραφία εγκεφάλου η οποία μπορεί ληφθεί με τη χρήση δύο διαφορετικών τεχνικών. Η μία τεχνική είναι η μαγνητική τομογραφία (magnetic resonance imaging – MRI) και η άλλη είναι η αξονική τομογραφία (computed tomography). Και οι δύο τεχνικές είναι φθηνότερες και ασφαλέστερες από τη βιοψία και ταυτόχρονα δεν καταπονούν τον άνθρωπο στον οποίο εφαρμόζονται. Μεταξύ των δύο αυτών τεχνικών, το οικονομικό κόστος της αξονικής τομογραφίας είναι μικρότερο, ωστόσο η μαγνητική τομογραφία προσφέρει εικόνα υψηλότερης ποιότητας χωρίς να χρησιμοποιεί την δυνητικά επιβλαβή για την ανθρώπινη υγεία ιοντίζουσα ακτινοβολία που χρησιμοποιεί η αξονική τομογραφία. Για αυτούς τους λόγους η χρήση της μαγνητικής τομογραφίας για την διάγνωση ενός εγκεφαλικού όγκου

προτιμάται πολύ συχνά. Μετά τη λήψη της μαγνητικής τομογραφίας, απαιτείται η εξέτασή της από εξειδικευμένο προσωπικό. Αυτό σημαίνει ότι δεσμεύει χρόνο από τη λειτουργία μιας κλινικής, ο οποίος θα μπορούσε να διοχετευτεί αλλού. Επίσης, η ορθότητα της τελικής διάγνωσης είναι μια παράμετρος που επηρεάζεται και από την ικανότητα και εμπειρία του ιατρικού προσωπικού, αφήνοντας πάντα το ενδεχόμενο ανθρώπινου λάθους ανοιχτό.

Κρίνεται, λοιπόν, σκόπιμη η εφαρμογή μεθόδων οι οποίες θα αυξήσουν τη συνολική ταχύτητα διάγνωσης και θα προσθέσουν ακόμα μεγαλύτερη αξιοπιστία σε αυτή. Η ραγδαία ανάπτυξη της τεχνητής νοημοσύνης (artificial intelligence – AI) τα τελευταία χρόνια, την έχει εισάγει σε πολλούς ζωτικής σημασίας τομείς της ανθρώπινης ζωής, όπως είναι η ιατροφαρμακευτική περίθαλψη. Κατά συνέπεια, ποικίλες έρευνες που βασίζονται στην τεχνητή νοημοσύνη έχουν διεξαχθεί με σκοπό τη δημιουργία συστημάτων διάγνωσης εγκεφαλικών όγκων με τη βοήθεια υπολογιστή (computer assisted diagnosis – CAD). Δύο από τα βασικότερα θέματα των ερευνών είναι η κατάτμηση (segmentation) και η κατηγοριοποίηση (classification) των εγκεφαλικών όγκων. Στόχος αυτής της εργασίας είναι δημιουργία ενός τέτοιου συστήματος βασισμένο στη βαθιά μάθηση (deep learning), η οποία αποτελεί ένα υποσύνολο της τεχνητής νοημοσύνης. Ειδικότερα, στη συγκεκριμένη εργασία παρουσιάζονται τέσσερα διαφορετικά μοντέλα κατηγοριοποίησης και ένα μοντέλο κατάτμησης μέσω δύο διαφορετικών προσεγγίσεων. Συνολικά το υπόλοιπο μέρος της εργασίας είναι δομημένο ως εξής: στην ενότητα 2 παρέχεται θεωρητικό υπόβαθρο που αφορά τη βαθιά μάθηση, στην ενότητα 3 περιλαμβάνονται εργασίες που έχουν γίνει στο παρελθόν πάνω στην κατηγοριοποίηση και κατάτμηση εγκεφαλικών όγκων, στην ενότητα 4 περιγράφεται η ανάπτυξη των προτεινόμενων μοντέλων και, εν τέλει, στην ενότητα 5 αποτυπώνονται ορισμένα συμπεράσματα και πιθανές μελλοντικές εργασίες.

2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Τεχνητός νευρώνας (Artificial neuron)

Ο τεχνητός νευρώνας αποτελεί το πιο θεμελιώδες συστατικό του τεχνητού νευρωνικού δικτύου. Είναι εμπνευσμένος από τους νευρώνες που βρίσκονται στον ανθρώπινο εγκέφαλο και κατά συνέπεια έχει παρόμοια συμπεριφορά με αυτούς. Ένας ανθρώπινος νευρώνας, σε υψηλό επίπεδο, απαρτίζεται από τρεις κύριους μηχανισμούς: τον μηχανισμό εισόδου, τον μηχανισμό υπολογισμού και τον μηχανισμό εξόδου. Σε αντιστοιχία, ένας τεχνητός νευρώνας είναι μία μαθηματική συνάρτηση η οποία δέχεται δυνητικά πολλαπλές εισόδους, εκτελεί μαθηματικές πράξεις πάνω σε αυτές και παράγει μία έξοδο. Η έξοδος αυτή μπορεί να αποτελέσει είσοδο για περισσότερους από έναν τεχνητούς νευρώνες. Εν τέλει, η λειτουργία που επιτελεί ένας τεχνητός νευρώνας τον οδηγεί σε δύο πιθανές καταστάσεις: ενεργός ή ανενεργός.

2.1.1 McCulloch-Pitts νευρώνας (McCulloch-Pitts neuron)

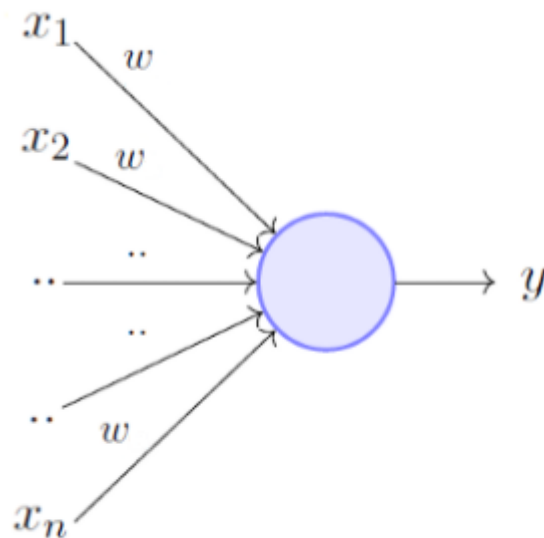
Το 1943 οι McCulloch και Pitts πρότειναν το πρώτο μοντέλο που προσομοιώνει την συμπεριφορά ενός ανθρώπινου νευρώνα [24]. Το συγκεκριμένο μοντέλο, αναφερόμενο ως McCulloch-Pitts νευρώνας, λειτουργεί κάτω από συγκεκριμένες υποθέσεις:

- Ο νευρώνας μπορεί να δεχτεί πολλαπλές εισόδους αλλά παράγει μία έξοδο.
- Κάθε είσοδος x_i και η έξοδος y είναι δυαδικές.
- Υπάρχει μία σταθερά θ που αντιστοιχεί στον νευρώνα και λειτουργεί ως κατώφλι.
- Διακρίνονται δύο κατηγορίες εισόδου: οι διεγερτικές και οι ανασταλτικές.
- Αρκεί μόνο μία ανασταλτική είσοδος για να ακυρώσει όλες τις διεγερτικές εισόδους καθιστώντας τον νευρώνα ανενεργό.
- Όλες οι διεγερτικές εισοδοί που δέχεται ο νευρώνας έχουν πανομοιότυπο βάρος w .
- Ο νευρώνας καθίσταται ενεργός αν και μόνο αν το άθροισμα των πολλαπλασιασμένων με το βάρος w διεγερτικών εισόδων είναι μεγαλύτερο ή ίσο με το κατώφλι θ και ο νευρώνας δεν έχει δεχθεί καμία ανασταλτική είσοδο.

Δεδομένων των παραπάνω υποθέσεων η έξοδος του τεχνητού νευρώνα μπορεί να αναπαρασταθεί μέσω της εξής δίκλαδης συνάρτησης :

$$y = \begin{cases} 1, & \sum_{i=1}^n w * x_i \geq \theta \text{ και δεν υπάρχει ανασταλτική είσοδος} \\ 0, & \text{αλλιώς} \end{cases}$$

όπου n είναι ο αριθμός των εισόδων ενός νευρώνα.



Εικόνα 2.1: Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα κατά το μοντέλο McCulloch και Pitts.

Η συγκεκριμένη μοντελοποίηση, παρά την επιτυχή προσέγγιση της συμπεριφοράς ενός πραγματικού νευρώνα, δεν παύει να έχει κάποιες σημαντικές αδυναμίες. Ειδικότερα, η ύπαρξη εισόδου η οποία είναι αποκλειστικά δυαδική και η αντιστοίχιση κάθε εισόδου ενός νευρώνα με τον ίδιο βαθμό σημαντικότητας w περιορίζουν την ευελιξία του συγκεκριμένου μοντέλου. Τέλος, η σταθερά θ πρέπει να οριστεί από τον άνθρωπο που υλοποιεί το μοντέλο, αυξάνοντας την πιθανότητα ανεπιθύμητης συμπεριφοράς από την επιλογή ακατάλληλου κατωφλίου.

2.1.2 Αισθητήρας (Perceptron)

Το 1957 ο Frank Rosenblatt πρότεινε ένα νέο και καταφανώς βελτιωμένο μοντέλο με όνομα Perceptron [30] το οποίο κατάφερε να εξαλείψει τις αδυναμίες του μοντέλου των McCulloch και Pitts. Το μοντέλο Perceptron είναι το πιο διαδεδομένο στις μέρες μας και αποτελεί τη βάση για ποικίλες αρχιτεκτονικές νευρωνικών δικτύων. Παρακάτω διακρίνονται οι εξής σημαντικότερες διαφορές:

- Δεν υπάρχουν πλέον ανασταλτικές εισοδοί.
- Κάθε είσοδος x_i μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός.
- Κάθε είσοδος x_i αντιστοιχίζεται με το δικό της βάρος w_i .

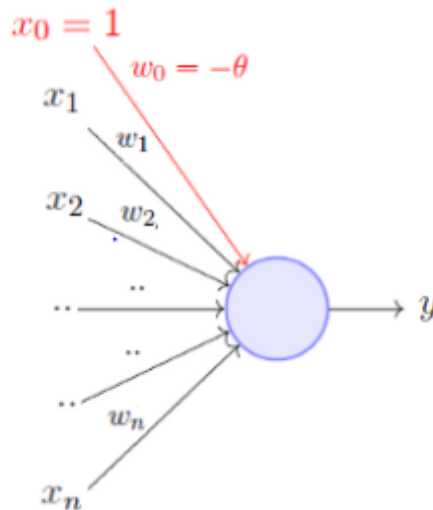
- Κάθε βάρος w_i μπορεί να είναι διαφορετικό από τα υπόλοιπα.
- Ο νευρώνας παρέχει μηχανισμό εκμάθησης. Αυτό σημαίνει πως τα βάρη που αντιστοιχούν στις εισόδους του νευρώνα και το κατώφλι μπορούν να επιλεγθούν από αυτόν. Για την αξιοποίηση αυτού του μηχανισμού, το κατώφλι θ αντιμετωπίζεται και αυτό ως είσοδος στο νευρώνα.

Λαμβάνοντας υπόψη αυτές τις τροποποιήσεις, καταλήγουμε στην παρακάτω συνάρτηση εξόδου :

$$y = \begin{cases} 1, & \sum_{i=1}^n w_i * x_i \geq \theta \\ 0, & \text{αλλιώς} \end{cases} \Leftrightarrow$$

$$y = \begin{cases} 1, & \sum_{i=1}^n w_i * x_i - \theta \geq 0 \\ 0, & \text{αλλιώς} \end{cases} \Leftrightarrow$$

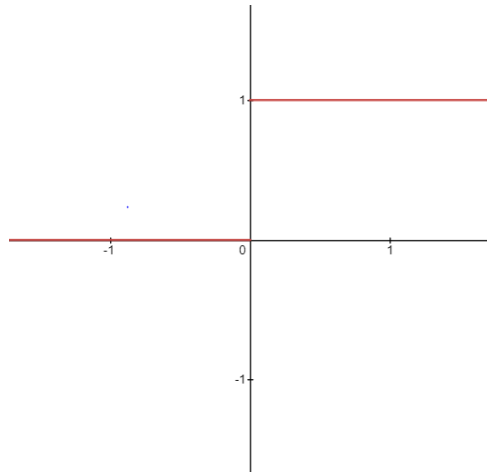
$$y = \begin{cases} 1, & \sum_{i=0}^n w_i * x_i \geq 0 \\ 0, & \text{αλλιώς} \end{cases} , x_0 = 1 \text{ και } w_0 = -\theta$$



Εικόνα 2.2: Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα κατά το μοντέλο Perceptron.

2.1.3 Πόλωση (Bias)

Όπως απεικονίζεται στην Εικόνα 2.2 ένας νευρώνας, εκτός από τις πραγματικές εισόδους του, μπορεί να έχει μια επιπλέον είσοδο, έστω με τιμή 1, στην οποία αντιστοιχίζεται ένα βάρος θ , που ονομάζεται κατώφλι. Απολύτως ισοδύναμα η ίδια επιπλέον είσοδος μπορεί να έχει τιμή -1 και βάρος ίσο με $-\theta$. Σε αυτή την περίπτωση, το βάρος ονομάζεται πόλωση. Η ύπαρξη της πόλωσης στον νευρώνα είναι εξαιρετικά σημαντική καθώς αυξάνει αισθητά την ευελιξία του μοντέλου. Συγκεκριμένα, αν θεωρήσουμε πως δεν υπάρχει πόλωση στον νευρώνα, τότε η συνάρτηση εξόδου θα απεικονίζεται από την Εικόνα 2.3.



Εικόνα 2.3: Παράδειγμα συνάρτησης εξόδου ενός νευρώνα χωρίς τη χρήση πόλωσης.

Όπως είναι εμφανές, ο νευρώνας θα είναι ενεργός αν και μόνο αν:

$$\sum_{i=1}^n w_i * x_i \geq 0$$

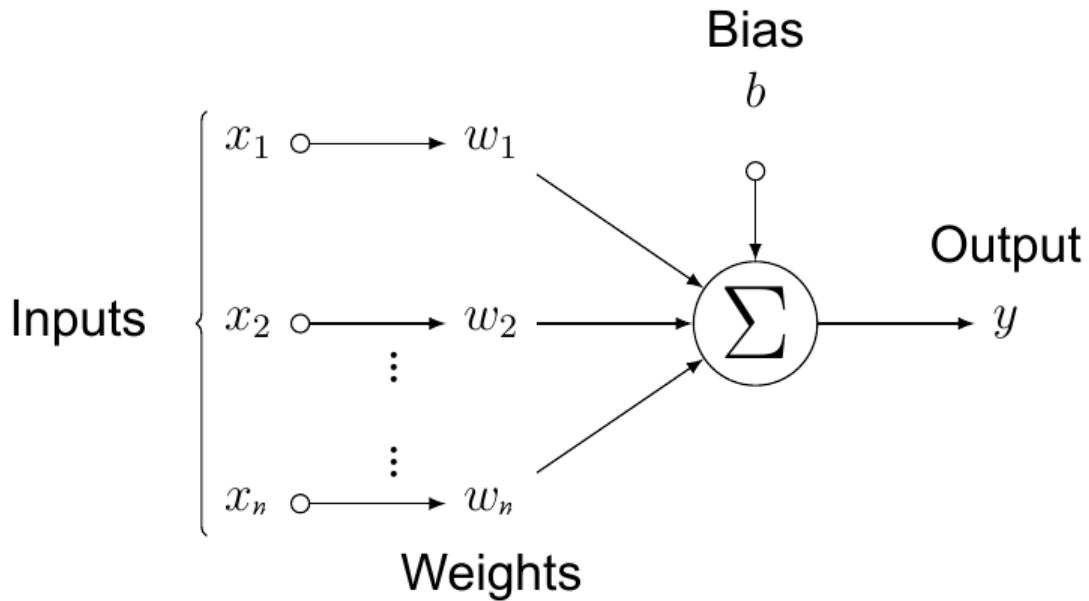
Ωστόσο, αυτός ο πολύ συγκεκριμένος κανόνας πιθανώς να μην είναι ιδανικός για όλα τα είδη των εισόδων, οδηγώντας τον νευρώνα σε μη επιθυμητή συμπεριφορά. Για παράδειγμα, ας υποθέσουμε πως για ένα σύνολο δεδομένων ο νευρώνας συμπεριφέρεται βέλτιστα αν και μόνο αν :

$$\sum_{i=1}^n w_i * x_i \geq 2$$

Χωρίς την πόλωση είναι προφανές πως θα επιτευχθεί υποβέλτιστη συμπεριφορά. Βάζοντάς την όμως σε χρήση, δίνεται η δυνατότητα στον νευρώνα να προσαρμόζει την έξοδο του κατάλληλα για την επίτευξη της βέλτιστης συμπεριφοράς. Επιπροσθέτως, παρόλο που, η πόλωση δε διαφέρει σε τίποτα από τα υπόλοιπα βάρη ενός νευρώνα, ο συμβολισμός που έχει επικρατήσει διαχωρίζει τις δύο αυτές έννοιες, περιγράφοντας ισοδύναμα την έξοδο του νευρώνα μέσω της εξίσωσης

$$y = \begin{cases} 1, & WX + b \geq 0 \\ 0, & \text{αλλιώς} \end{cases}$$

όπου W είναι ένα διάνυσμα που περιέχει όλα τα βάρη που αντιστοιχούν στις εισόδους του νευρώνα, X είναι ένα διάνυσμα που περιέχει όλες τις εισόδους του νευρώνα και b είναι η πόλωση του νευρώνα. Εν τέλει, η Εικόνα 2.4 απεικονίζει την λειτουργία ενός νευρώνα με βάση τον συμβολισμό που έχει επικρατήσει και τον οποίο θα ακολουθήσουμε.



Εικόνα 2.4: Απεικόνιση της λειτουργίας ενός τεχνητού νευρώνα χρησιμοποιώντας τον επικρατέστερο συμβολισμό.

2.2 Συνάρτηση ενεργοποίησης (Activation function)

Όπως είδαμε, η έξοδος που παράγει ένας νευρώνας εξαρτάται και από τη συνάρτηση εξόδου ή συνάρτηση ενεργοποίησης, όπως θα την αποκαλούμε από εδώ και πέρα. Για την κατανόηση της σημαντικότητας ύπαρξης συναρτήσεων ενεργοποίησης ας υποθέσουμε πως δεν υπάρχουν. Σε αυτή την περίπτωση δημιουργούνται δύο σημαντικά προβλήματα:

1. Κάθε άθροισμα που παράγεται από έναν νευρώνα θα μεταβιβάζεται σε άλλους νευρώνες αυτούσιο. Αυτοί με τη σειρά τους θα δημιουργούν νέα αθροίσματα που θα τα μεταβιβάζουν σε άλλους νευρώνες. Είναι ευκατανόητο ότι οι τιμές που μπορεί να παράγει ένας νευρώνας δεν έχουν κανένα όριο και πιθανώς να οδηγήσουν σε υπολογιστικά προβλήματα.
2. Τα αθροίσματα αυτά δεν παύουν να είναι γραμμικοί συνδυασμοί, δηλαδή ολόκληρο το νευρωνικό δίκτυο θα μπορούσε να αναπαρασταθεί ως ένας πολλαπλασιασμός πινάκων. Ωστόσο τα περισσότερα προβλήματα που συναντάμε στον πραγματικό κόσμο δεν μπορούν να αναπαρασταθούν επαρκώς με γραμμικό τρόπο και για αυτό κρίνεται απαραίτητη η εισαγωγή της μη-γραμμικότητας (non-linearity) στο νευρωνικό δίκτυο.

Η συνάρτηση ενεργοποίησης καθορίζει την έξοδο ενός νευρώνα, δεδομένου ενός συνόλου εισόδων. Με άλλα λόγια, επιλέγει σε ποια από τις δύο πιθανές καταστάσεις θα βρεθεί ένας νευρώνας έχοντας, κατ' επέκταση, μεγάλη επιρροή στην συμπεριφορά και στις δυνατότητες ενός νευρωνικού δικτύου διότι κάθε νευρώνας αντιστοιχίζεται με μία συνάρτηση ενεργοποίησης. Συγκεκριμένα, δέχεται ως είσοδο τον γραμμικό μετασχηματισμό που δημιουργεί ο νευρώνας από τις εισόδους του και παράγει μία έξοδο. Με τη χρήση συναρτήσεων ενεργοποίησης η έξοδος του νευρώνα ενδέχεται να μην είναι δυαδική. Μάλιστα, η συντριπτική πλειοψηφία των συναρτήσεων ενεργοποίησης που χρησιμοποιούνται σήμερα παρέχουν ένα ικανοποιητικά μεγάλο διάστημα τιμών από την

οποία επέρχεται η έξοδος, στοχεύοντας στη μεγιστοποίηση της ευελιξίας του μοντέλου. Αυτό δίνει παράλληλα και μία διαφορετική διάσταση στην έννοια της κατάστασης ενός νευρώνα καθώς σε ένα νευρωνικό δίκτυο ένας νευρώνας μπορεί να είναι περισσότερο ή λιγότερο ενεργός/ανενεργός σε σχέση με έναν άλλον. Όσο μεγαλύτερη τιμή έχει η έξοδος, τόσο πιο ενεργός καθίσταται ένας νευρώνας. Όλες οι διαδοσόμενες συναρτήσεις ενεργοποίησης παράγουν τιμές που ανήκουν σε ένα διάστημα. Επιλέγοντας έξοδο από ένα διάστημα, ακόμα και μικρές αλλαγές στα βάρη των εισόδων αποτυπώνονται πολύ καλύτερα στην έξοδο, κάτι που δεν είναι εφικτό με τη χρήση δυαδικής εξόδου.

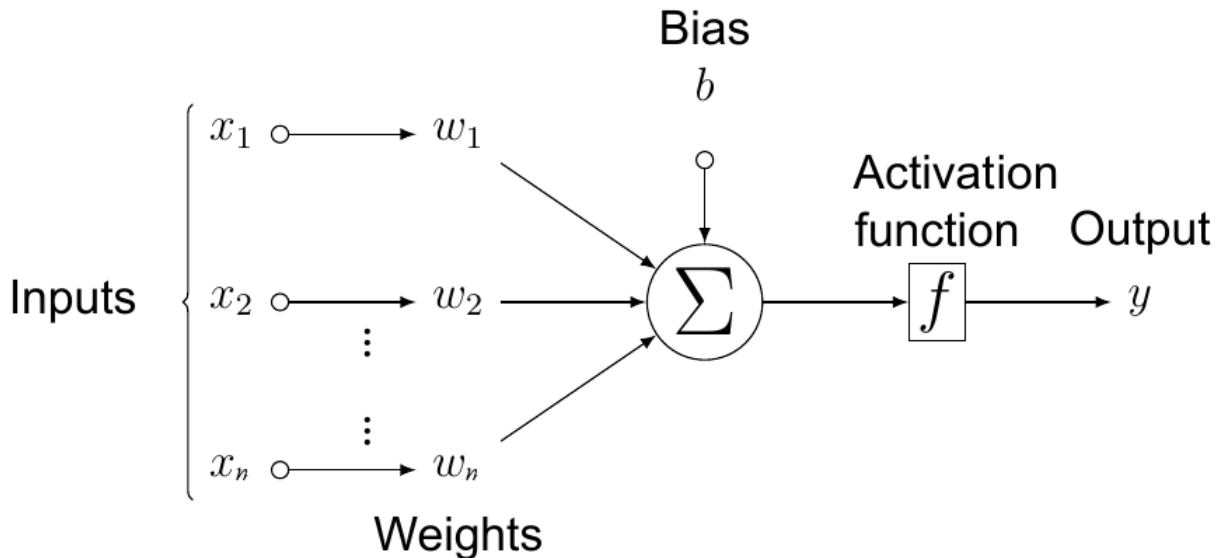
Παρά την ύπαρξη ποικίλων συναρτήσεων ενεργοποίησης, όλες μπορούν να κατηγοριοποιηθούν σε μία από τις δύο κατηγορίες: γραμμικές και μη γραμμικές. Οι γραμμικές συναρτήσεις ενεργοποίησης μπορούν να λύσουν μόνο το πρώτο από τα δύο προβλήματα που αναφέρθηκαν καθώς μπορούν να λειτουργήσουν βέλτιστα αποκλειστικά σε γραμμικά διαχωρίσιμα σύνολα δεδομένων, συνεπώς η χρήση τους δεν συνηθίζεται. Από την άλλη πλευρά, η χρήση μη-γραμμικής συνάρτησης αποτελεί γενικά την ορθότερη επιλογή γιατί λύνει και τα δύο προβλήματα απογειώνοντας κατακόρυφα τις δυνατότητες ενός νευρωνικού δικτύου.

Η επιλογή κατάλληλης συνάρτησης ενεργοποίησης δεν είναι κάτι που περιλαμβάνεται στο μηχανισμό εκμάθησης ενός νευρωνικού δικτύου, συνεπώς υπεύθυνος για αυτή είναι ο άνθρωπος. Ως κριτήρια επιλογής, εκτός της ικανότητας εξάλειψης των δύο προαναφερθέντων προβλημάτων, χρησιμοποιούνται και τα παρακάτω χαρακτηριστικά που πιθανόν να έχει μία συνάρτηση ενεργοποίησης :

- **Παραγωγίσιμη:** Ο υπολογισμός της κλίσης μιας συνάρτησης ενεργοποίησης καθίσταται χρήσιμη σε πολλές φάσεις κατά την εκπαίδευση ενός νευρωνικού δικτύου. Συνεπώς, είναι σημαντικό, αλλά όχι αναγκαία απαραίτητο, η παράγωγος της συνάρτησης να υπάρχει σε κάθε σημείο του πεδίου ορισμού της.
- **Ανοχή στο πρόβλημα των εξαφανισμένων κλίσεων (Vanishing Gradient Problem):** Το συγκεκριμένο πρόβλημα εξηγείται στην ενότητα 2.5. Με λίγα λόγια, είναι σημαντικό για μία συνάρτηση ενεργοποίησης να μην έχει πολύ μικρή κλίση σε οποιοδήποτε σημείο του πεδίου ορισμού της. Ο υπολογισμός της κλίσης μιας συνάρτησης γίνεται μέσω της παραγώγου.
- **Δυνατότητα εφαρμογής οπισθοδιάδοσης (backpropagation) στο νευρωνικό δίκτυο:** Η διαδικασία της οπισθοδιάδοσης αφορά τον τρόπο εκμάθησης ενός νευρωνικού δικτύου αλλά επηρεάζεται από την συνάρτηση ενεργοποίησης. Περισσότερα για την οπισθοδιάδοση αναφέρονται στην ενότητα 2.6. Συνοπτικά, η οπισθοδιάδοση δεν είναι εφικτή εάν η παράγωγος της συνάρτησης ενεργοποίησης είναι ανεξάρτητη από την είσοδο.
- **Κεντραρισμένη στο μηδέν (Zero-centered):** Μια συνάρτηση είναι κεντραρισμένη στο μηδέν όταν το πεδίο τιμών της περιλαμβάνει θετικές και αρνητικές τιμές. Αν μία συνάρτηση ενεργοποίησης δεν είναι κεντραρισμένη στο μηδέν τότε η έξοδος της θα είναι πάντα είτε θετική είτε αρνητική. Αυτή η συμπεριφορά δεν είναι απαραίτητα καταστροφική για ένα νευρωνικό δίκτυο, ωστόσο ενδέχεται να χρειαστούν περισσότερες εποχές (epochs) για την εκμάθηση των κατάλληλων βαρών στους νευρώνες.
- **Χαμηλό υπολογιστικό κόστος:** Η συνάρτηση ενεργοποίησης μαζί με την παράγωγό της ενδέχεται να υπολογιστούν εκατομμύρια φορές από ένα νευρωνικό δίκτυο. Συνεπώς, για εξοικονόμηση χρόνου, πολύ σημαντικό παράγοντα αποτελεί

η χαμηλή απαίτηση υπολογιστικών πόρων και για τις δύο συναρτήσεις.

Γνωρίζοντας πλέον για την ύπαρξη της συνάρτησης ενεργοποίησης, μπορούμε να έχουμε την απεικόνιση της πλήρους λειτουργίας ενός νευρώνα η οποία φαίνεται στην Εικόνα 2.5.

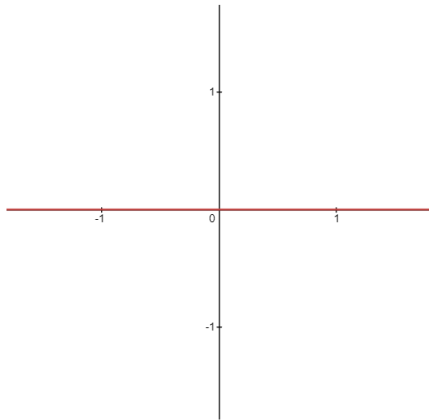


Εικόνα 2.5: Απεικόνιση της πλήρους λειτουργίας ενός τεχνητού νευρώνα.

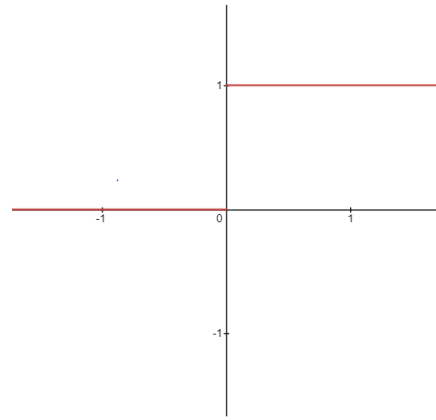
2.2.1 Βηματική συνάρτηση (Heaviside Step Function)

Η βηματική συνάρτηση αποτελεί την ίδια συνάρτηση με αυτή που είδαμε στον ορισμό του μοντέλου Perceptron και αποτελεί την πιο απλή συνάρτηση ενεργοποίησης. Η συγκεκριμένη συνάρτηση δεν είναι παραγωγίσιμη στο μηδέν. Στα υπόλοιπα σημεία του πεδίου ορισμού της, η παράγωγος της ισούται με μηδέν αναδεικνύοντας το πρόβλημα των εξαφανισμένων κλίσεων, ενώ παράλληλα η οπισθοδιάδοση δεν είναι εφικτή διότι το μηδέν είναι μια σταθερά ανεξάρτητη από την είσοδο. Επίσης, δεν είναι κεντραρισμένη στο μηδέν διότι δεν παρέχει καμία αρνητική έξοδο. Παρά το χαμηλό υπολογιστικό κόστος τόσο της ίδιας όσο και της παραγωγού της, η χρησιμότητά της περιορίζεται στη δυαδική κατηγοριοποίηση (binary classification) ενός σετ δεδομένων.

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



$$H'(x) = 0, \quad x \neq 0$$

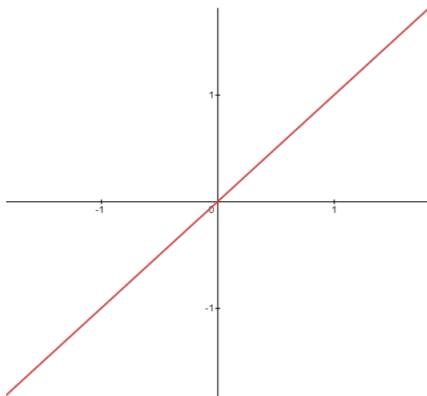


Εικόνα 2.6: Απεικόνιση της βηματικής συνάρτησης ενεργοποίησης και της παραγώγου της.

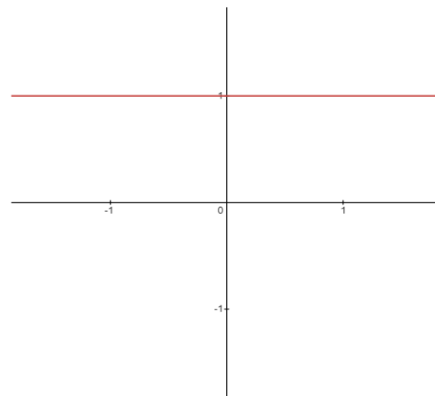
2.2.2 Γραμμική συνάρτηση (Linear function)

Η ταυτοτική συνάρτηση είναι μια γραμμική συνάρτηση που παράγει έξοδο ανάλογη με τον γραμμικό μετασχηματισμό που δέχεται ως είσοδο από τον νευρώνα. Το πεδίο ορισμού και κατά συνέπεια το σύνολο τιμών της είναι όλοι οι πραγματικοί αριθμοί, δηλαδή η χρήση της δεν ικανοποιεί την ανάγκη για περιορισμένο εύρος τιμών στην έξοδο. Συνεπώς, δεν μπορεί να λύσει κανένα από τα δύο προβλήματα που αναφέρθηκαν στην αρχή της ενότητας. Η συγκεκριμένη συνάρτηση είναι παραγωγίσιμη, έχει ανοχή στο πρόβλημα των εξαφανισμένων κλίσεων, είναι κεντραρισμένη στο μηδέν και έχει χαμηλό υπολογιστικό κόστος. Ωστόσο, η ανικανότητα εφαρμογής οπισθοδιάδοσης, λόγω σταθερής παραγώγου, και η γραμμικότητα που την διέπει, την καθιστούν ικανοποιητική επιλογή αποκλειστικά σε γραμμικά διαχωρίσιμα προβλήματα.

$$f(x) = x$$



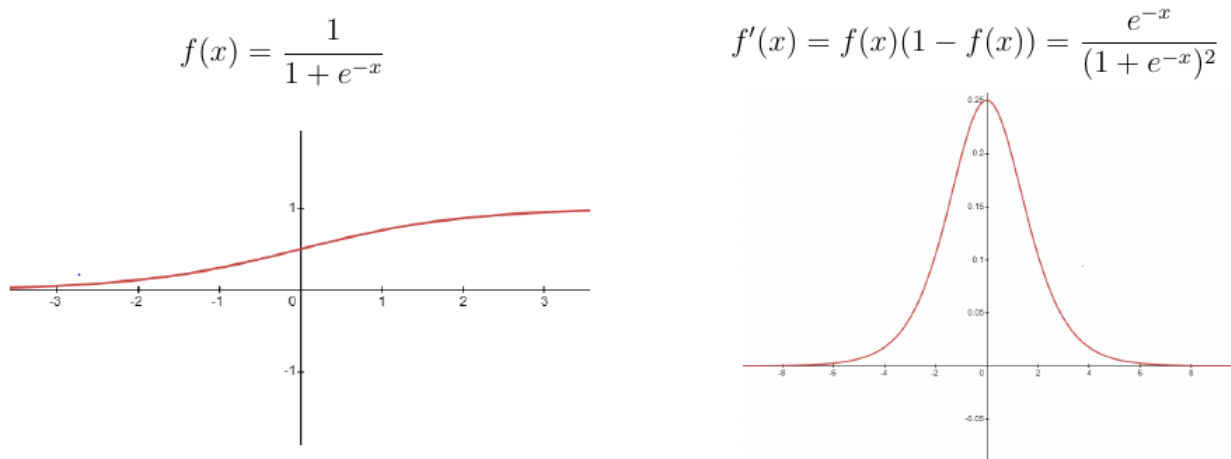
$$f'(x) = 1$$



Εικόνα 2.7: Απεικόνιση της γραμμικής συνάρτησης ενεργοποίησης και της παραγώγου της.

2.2.3 Λογιστική συνάρτηση (Logistic function)

Η λογιστική συνάρτηση ανήκει στην οικογένεια των σιγμοειδών συναρτήσεων η οποία περιέχει συναρτήσεις με χαρακτηριστικό σχήμα που μοιάζει με το αγγλικό γράμμα S. Επίσης, είναι μια μη γραμμική συνάρτηση η οποία μπορεί να παράξει ως έξοδο οποιονδήποτε πραγματικό αριθμό στο διάστημα $[0,1]$. Η συγκεκριμένη συνάρτηση δεν είναι κεντραρισμένη στο μηδέν, παράγοντας αποκλειστικά μη αρνητικές τιμές, όντας παράλληλα φραγμένη από τον αριθμό 1. Γι'αυτό το λόγο, όμως, το διάστημα τιμών της είναι ιδανικό για την μετατροπή της εισόδου ενός νευρώνα σε πιθανότητα. Το άθροισμα των πιθανοτήτων που παράγει ως έξοδο η λογιστική συνάρτηση δεν εξασφαλίζεται ότι αθροίζει στο ένα, καθιστώντας την προτιμότερη σε προβλήματα κατηγοριοποίησης πολλαπλών κατηγοριών (multi-label). Παρόλο που είναι παραγωγίσιμη, ο εκθετικός όρος που περιέχει καθιστά τον υπολογισμό της παραγώγου της ακριβό. Η παράγωγός της επιτρέπει την εφαρμογή της οπισθοδιάδοσης αλλά πάσχει από το πρόβλημα των εξαφανισμένων κλίσεων. Εν τέλει, λόγω της μη γραμμικότητας η συγκεκριμένη συνάρτηση θεωρητικά είναι ικανή να ανταπεξέλθει σε πολύπλοκα προβλήματα, ωστόσο εξαιτίας των προαναφερθέντων προβλημάτων που παρουσιάζει, δεν αποτελεί την πιο δημοφιλή επιλογή.



Εικόνα 2.8: Απεικόνιση της λογιστικής συνάρτησης ενεργοποίησης και της παραγώγου της.

2.2.4 Κανονικοποιημένη εκθετική συνάρτηση (Softmax)

Η κανονικοποιημένη εκθετική συνάρτηση αποτελεί μια γενίκευση της λογιστικής συνάρτησης σε πολλές διαστάσεις και επιφέρει τα ίδια θετικά και αρνητικά χαρακτηριστικά με εκείνη. Η κύρια διαφορά τους είναι ότι η κανονικοποιημένη εκθετική συνάρτηση εξασφαλίζει ότι το άθροισμα των εξόδων της ισούται πάντα με ένα, σε αντίθεση με την λογιστική συνάρτηση όπου κάθε έξοδος είναι ανεξάρτητη. Το γεγονός αυτό καθιστά την κανονικοποιημένη εκθετική συνάρτηση να προτιμάται περισσότερο σε προβλήματα κατηγοριοποίησης πολλαπλών κλάσεων (multiclass classification).

Η συνάρτηση ενεργοποίησης Softmax εκφράζεται ως

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}}, \quad 0 \leq i \leq K \text{ και } x = (x_1, \dots, x_k) \in \mathbb{R}^K$$

και η παράγωγός της ως

$$f'(x_i) = \begin{cases} f(x_i)(1 - f(x_i)), & i = j \\ -f(x_i)f(x_j), & \text{αλλιώς} \end{cases}, \quad 0 \leq i \leq K \text{ και } x = (x_1, \dots, x_k) \in \mathbb{R}^K$$

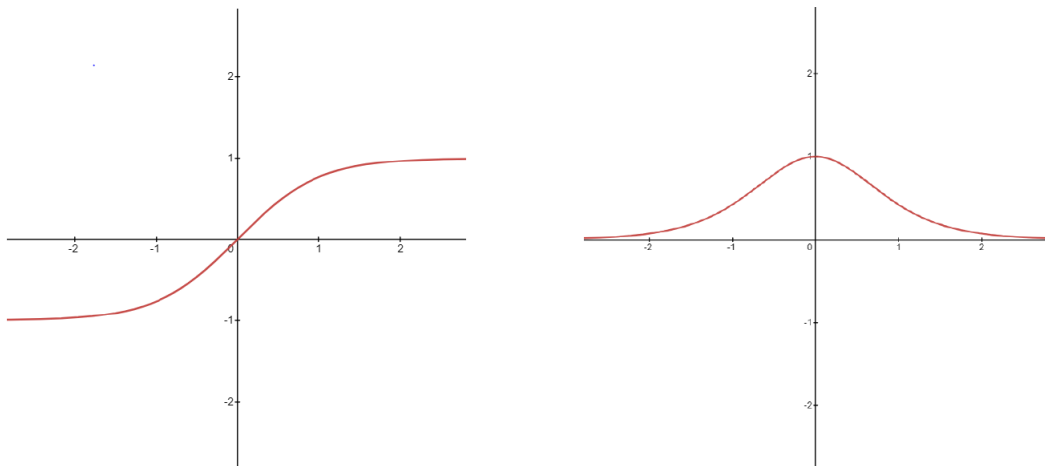
όπου K είναι ο αριθμός των εισόδων.

2.2.5 Υπερβολική εφαπτομένη (Tanh function)

Η υπερβολική εφαπτομένη, όπως μαρτυρά και το σχήμα 2.11, είναι παρόμοια με τη λογιστική. Παρά το γεγονός ότι τα περισσότερα χαρακτηριστικά τους, θετικά ή αρνητικά, είναι ταυτόσημα, υπάρχουν δύο διαφορές οι οποίες καθιστούν τη χρήση της υπερβολικής εφαπτομένης ορθότερη επιλογή. Συγκεκριμένα, η ίδια είναι κεντραρισμένη στο μηδέν καθώς το σύνολο τιμών της είναι το $[-1, 1]$ και παρέχει μεγαλύτερη ανοχή στο πρόβλημα των εξαφανισμένων κλίσεων διότι η παράγωγός της παρουσιάζει μεγαλύτερη κλίση από εκείνη της λογιστικής.

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

$$\tanh'(x) = 1 - \tanh^2(x) = 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2$$

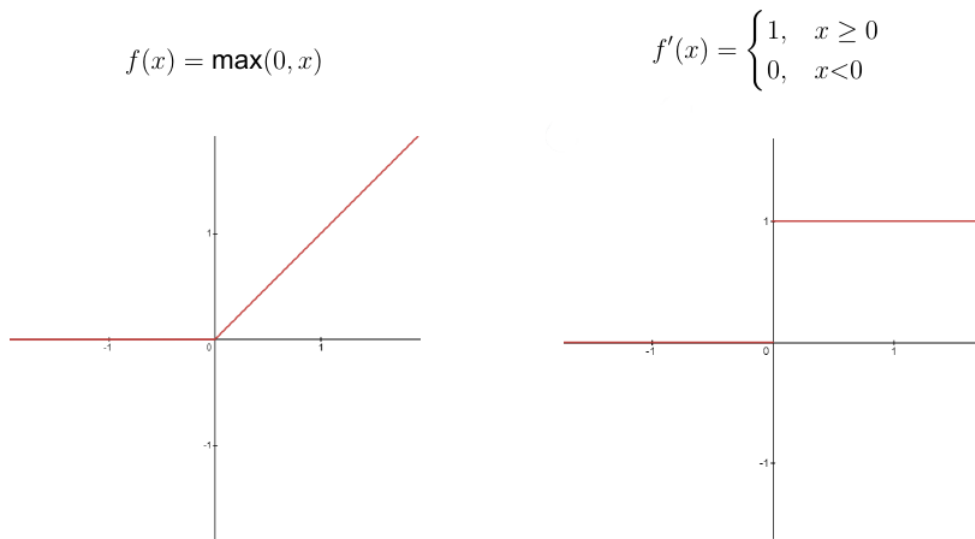


Εικόνα 2.9: Απεικόνιση της συνάρτησης ενεργοποίησης ReLU και της παραγώγου της.

2.2.6 Ανορθωμένη Γραμμική Μονάδα (Rectified Linear Unit – ReLU)

Η ReLU αποτελεί την πιο δημοφιλή επιλογή, καθώς η χρήση της έχει επιφέρει καλύτερα αποτελέσματα από τις υπόλοιπες σε πολλά προβλήματα. Ειδικότερα, η ίδια είναι μία μη-

γραμμική συνάρτηση η οποία όπως φαίνεται στην Εικόνα 2.13 μοιάζει πολύ σε γραμμικές συναρτήσεις. Για αυτό το λόγο συνδυάζει την απλότητα των γραμμικών συναρτήσεων που καθιστά τόσο τις ίδιες όσο και τις παραγώγους τους υπολογιστικά φθηνές, με τις ιδιότητες των μη-γραμμικών συναρτήσεων που είναι καταλληλότερες για την επίλυση δύσκολων προβλημάτων. Η ReLU παράγει τιμές που κυμαίνονται στο διάστημα $[0, \infty)$ και είναι παραγωγίσιμη σε κάθε σημείο εκτός του μηδέν, ενώ δεν πάσχει από το πρόβλημα των εξαφανισμένων κλίσεων. Ωστόσο, δεν είναι κεντραρισμένη στο μηδέν και πάσχει από το πρόβλημα dying ReLU, το οποίο βασίζεται στο γεγονός πως η ReLU μηδενίζει οποιαδήποτε αρνητική είσοδο, απενεργοποιώντας πιθανώς παραπάνω νευρώνες από όσους θα χρειαζόταν σε ένα πρόβλημα. Τέλος, δεν υπάρχει κανένα όριο στην έξοδο της οδηγώντας σε πιθανά υπολογιστικά προβλήματα.



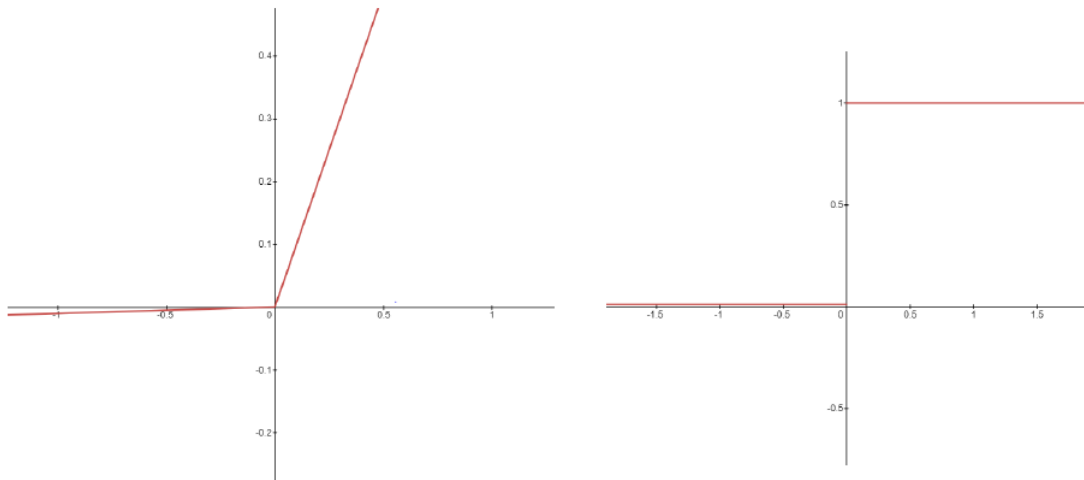
Εικόνα 2.10: Απεικόνιση της συνάρτησης ενεργοποίησης ReLU και της παραγώγου της.

2.2.7 Leaky ReLU

Η συνάρτηση ενεργοποίησης Leaky ReLU αποτελεί μία πιο τροποποιημένη έκδοση της ReLU η οποία εξασφαλίζει πως δε θα απενεργοποιηθούν όσοι νευρώνες δεχθούν αρνητική είσοδο. Αυτό επιτυγχάνεται πολλαπλασιάζοντας κάθε αρνητική είσοδο με μία πολύ μικρή σταθερά και συγκεκριμένα με 0.01.

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0.01x, & \text{αλλιώς} \end{cases}$$

$$f'(x) = \begin{cases} 1, & x \geq 0 \\ 0.01, & x < 0 \end{cases}$$



Εικόνα 2.11: Απεικόνιση της συνάρτησης Leaky ReLU και της παραγώγου της.

2.3 Στρώματα (Layers)

Μέχρι στιγμής έχει αναλυθεί ο τρόπος λειτουργίας ενός νευρώνα μεμονωμένα. Ωστόσο, για να είναι ικανό ένα νευρωνικό δίκτυο να ανταπεξέλθει σε απαιτητικά προβλήματα, είναι σημαντικό να αποτελείται από ένα αρκετά μεγαλύτερο αριθμό νευρώνων οι οποίοι αλληλεπιδρούν μεταξύ τους. Ο τρόπος που διασυνδέονται οι νευρώνες δημιουργεί τα λεγόμενα στρώματα τα οποία χωρίζονται σε τρεις κατηγορίες:

- **Στρώμα εισόδου (Input layer):** Κάθε νευρωνικό δίκτυο έχει ακριβώς ένα στρώμα εισόδου και ο αριθμός των νευρώνων που απαρτίζουν το συγκεκριμένο στρώμα εξαρτάται αποκλειστικά από τον αριθμό των χαρακτηριστικών (features) του συνόλου δεδομένων που εισέρχονται στο νευρωνικό δίκτυο. Η ιδιαιτερότητα του στρώματος εισόδου προέρχεται από το γεγονός ότι ο ρόλος του κάθε νευρώνα που ανήκει σε αυτό το στρώμα είναι η μεταβίβαση των τιμών που δέχεται στην είσοδο, στο πρώτο κρυφό στρώμα δίχως να τελέσει καμία επεξεργασία πάνω σε αυτές. Με άλλα λόγια, η λειτουργία ενός νευρώνα όπως αυτή έχει περιγραφεί δεν ισχύει για τους νευρώνες του στρώματος εισόδου. Αντιθέτως, λειτουργούν ως απλοί μεταβιβαστές πληροφορίας.
- **Κρυφό στρώμα (Hidden layer):** Σε ένα κρυφό στρώμα η λειτουργία των νευρώνων, σε αντίθεση με το στρώμα, ταυτίζεται με αυτή που περιγράφεται παραπάνω, ενώ ένα νευρωνικό δίκτυο ενδέχεται να έχει παραπάνω από ένα κρυφό στρώμα. Ωστόσο, αν ένα πρόβλημα είναι γραμμικά διαχωρίσιμο τότε δεν απαιτείται κανένα κρυφό στρώμα. Όσον αφορά τα μη γραμμικά διαχωρίσιμα προβλήματα, ένα κρυφό στρώμα είναι αρκετό για την επίλυση των περισσότερων από αυτά, ενώ σε μερικές περιπτώσεις απαιτείται η χρήση δύο κρυφών στρωμάτων. Μάλιστα, σε θεωρητικό επίπεδο, δεν υπάρχει λόγος χρησιμοποίησης παραπάνω από δύο κρυφά στρώματα [15]. Αντιθέτως, η άσκοπη πρόσθεση κρυφών στρωμάτων εγκυμονεί την πιθανότητα μειωμένης επίδοσης του νευρωνικού δικτύου. Συγκεκριμένα, τα

προβλήματα που ενδέχεται να εμφανιστούν είναι η υπερπροσαρμογή (overfitting) και η αυξημένη απαιτούμενη ώρα εκπαίδευσης του νευρωνικού δικτύου. Ομοίως, ο αριθμός των νευρώνων που θα απαρτίζει ένα κρυφό στρώμα αποτελεί μία πολύ σημαντική επιλογή καθώς έχει μεγάλη επίδραση στη συμπεριφορά του νευρωνικού δικτύου. Η επιλογή ενός μεγαλύτερου αριθμού νευρώνων από τον απαιτούμενο σε ένα κρυφό στρώμα έχει ακριβώς τις ίδιες πιθανές επιπτώσεις με την επιλογή παραπάνω κρυφών στρωμάτων που μόλις αναφέρθηκαν. Απεναντίας, η επιλογή λιγότερων στρωμάτων ή συνολικά νευρώνων οδηγεί στην υποπροσαρμογή (underfitting). Ωστόσο, ο κατάλληλος αριθμός νευρώνων ανά στρώμα δεν μπορεί να είναι γνωστός απο πριν, καθιστώντας τα παραπάνω φαινόμενα πολύ συχνά κατά την εκπαίδευση νευρωνικών δικτύων. Για αυτό ακριβώς το λόγο, εφαρμόζονται κάποιοι εμπειρικοί κανόνες που βοηθούν σε μία καλή αρχική προσέγγιση του επιθυμητού αριθμού νευρώνων στα κρυφά στρώματα. Από εκεί και πέρα, είναι ευθύνη του ανθρώπου που επιλέγει την αρχιτεκτονική του νευρωνικού δικτύου να καθορίσει τον ιδανικό αριθμό. Κάτι τέτοιο καθίσταται εφικτό αποκλειστικά με επαρκή πειραματισμό. Μερικοί εμπειρικοί κανόνες είναι οι εξής [15]:

- Ο αριθμός νευρώνων ανά κρυφό στρώμα πρέπει να βρίσκεται ανάμεσα από τον αριθμό νευρώνων στο στρώμα εισόδου και τον αριθμό νευρώνων στο στρώμα εξόδου.
 - Ο αριθμός νευρώνων ανά κρυφό στρώμα πρέπει να ισούται με τα $2/3$ του αριθμού νευρώνων στο στρώμα εισόδου συν τον αριθμό νευρώνων στο στρώμα εξόδου.
 - Ο αριθμός νευρώνων ανά κρυφό στρώμα πρέπει να είναι μικρότερος από το διπλάσιο του αριθμού των νευρώνων στο στρώμα εξόδου.
- **Στρώμα εξόδου (Output layer):** Το συγκεκριμένο στρώμα αφορά το τελικό αποτέλεσμα του νευρωνικού δικτύου και είναι μοναδικό. Ο αριθμός των νευρώνων που το απαρτίζουν, εξαρτάται από το εκάστοτε πρόβλημα. Αν το πρόβλημα είναι παλινδρόμηση (regression), τότε αρκεί μόνο ένας νευρώνας. Στην περίπτωση της δυαδικής κατηγοριοποίησης, αν η συνάρτηση ενεργοποίησης που χρησιμοποιείται είναι η λογιστική, τότε μπορεί και πάλι να χρησιμοποιηθεί ένας νευρώνας, διαφορετικά χρειάζονται δύο νευρώνες. Τέλος, στην κατηγοριοποίηση πολλαπλών κλάσεων, ο αριθμός των νευρώνων είναι ίσος με τον αριθμό των κλάσεων.

2.4 Συνάρτηση κόστους (Cost function)

Το βασικότερο κριτήριο κατά την αξιολόγηση ενός νευρωνικού δικτύου είναι η ικανότητά του να λαμβάνει σωστές αποφάσεις. Έναν τρόπο για την καταγραφή αυτής της ικανότητας αποτελεί η συνάρτηση κόστους η οποία προσφέρει ζωντανή ανατροφοδότηση καθώς το νευρωνικό δίκτυο εκπαιδεύεται. Ειδικότερα, κατά το πέρας κάθε εποχής, το νευρωνικό δίκτυο μέσω του στρώματος εξόδου παράγει προβλέψεις για τα δεδομένα με τα οποία εκπαιδεύεται και η συνάρτηση κόστους παράγει έναν πραγματικό αριθμό που προκύπτει από την σύγκριση των προβλεπόμενων τιμών με τις αληθινές τιμές των δεδομένων. Αξιοσημείωτο είναι το γεγονός ότι η συνάρτηση κόστους αφορά αποκλειστικά τη φάση εκπαίδευσης ενός νευρωνικού δικτύου, αποτελώντας μάλιστα καίριο συστατικό για την βελτιστοποίησή του. Η πρόβλεψη της απόδοσης ενός νευρωνικού δικτύου πάνω σε δεδομένα που δεν ανήκουν στο σύνολο δεδομένων εκμάθησης ονομάζεται πραγματικό ρίσκο (true risk) και η αδυναμία ακριβούς υπολογισμού του πραγματικού ρίσκου εξαιτίας

της άγνωστης κατανομής αυτών των δεδομένων αποτελεί την πηγή σημαντικότητας της συνάρτησης κόστους διότι μέσω εκείνης υπολογίζεται η επίδοση του νευρωνικού δικτύου στα δεδομένα εκμάθησης, ή αλλιώς εμπειρικό ρίσκο (empirical risk), από το οποίο μπορεί να γίνει μία ασφαλής πρόβλεψη για την μελλοντική επίδοση στα πραγματικά δεδομένα. Επίσης, η συνάρτηση κόστους δεν αποτελεί ένα αμιγές ποσοστό επιτυχουσών ή αποτυχουσών προβλέψεων αλλά μία ένδειξη της επίδοσης του νευρωνικού δικτύου στην εκμάθηση ενός δοσμένου συνόλου δεδομένων καθώς η σύγκριση αφορά τιμές που, όπως είδαμε στις συναρτήσεις ενεργοποίησης, πιθανώς να είναι πραγματικοί αριθμοί και όχι ακέραιοι. Σκοπός είναι η ελαχιστοποίηση της τιμής που παράγει η συνάρτηση εξόδου και για αυτό χρησιμοποιούνται συγκεκριμένοι αλγόριθμοι, όπως θα δούμε και παρακάτω. Η γενική μορφή της συνάρτησης κόστους δίνεται από τον τύπο:

$$C(f) = \frac{1}{N} \sum_{i=1}^N l(f(x_i), y_i)$$

όπου N είναι ο αριθμός των δεδομένων εκπαίδευσης, $f(x_i)$ είναι η πρόβλεψη για το δεδομένο εκπαίδευσης i , y_i είναι η πραγματική τιμή του δεδομένου εκπαίδευσης i και $l(f(x_i), y_i)$ είναι η συνάρτηση που παράγει την απόκλιση της πρόβλεψης για το δεδομένο εκπαίδευσης i από την πραγματική τιμή του, με βάση κάποιο κριτήριο, γνωστή ως συνάρτηση απώλειας (loss function). Παρακάτω παρατηρούνται οι πιο διαδεδομένες συναρτήσεις κόστους.

2.4.1 Μέσο απόλυτο σφάλμα (Mean absolute error)

Η συγκεκριμένη συνάρτηση υπολογίζει σε απόλυτη τιμή τη μέση απόκλιση των προβλέψεων από τις πραγματικές τιμές των δεδομένων και η χρήση της είναι συνήθης στα προβλήματα παλινδρόμησης. Η ύπαρξη απόλυτης τιμής καθιστά υπολογιστικά ακριβό τον υπολογισμό της παραγώγου, η οποία χρησιμοποιείται στη μέθοδο της κατηφορικής κλίσης (gradient descent). Εκφράζεται από τον τύπο:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - f(x_i)|$$

όπου N είναι ο αριθμός των δεδομένων εκπαίδευσης, y_i είναι οι προβλέψεις και x_i είναι οι πραγματικές τιμές των δεδομένων.

2.4.2 Μέσο τετραγωνικό σφάλμα (Mean squared error)

Το μέσο τετραγωνικό σφάλμα υπολογίζει τη μέση τετραγωνική απόκλιση των προβλέψεων από τις πραγματικές τιμές των δεδομένων και ομοίως αποτελεί εξαιρετική επιλογή για τα προβλήματα παλινδρόμησης. Στην πράξη η κύρια διαφορά του μέσου τετραγωνικού σφάλματος από το μέσο απόλυτο σφάλμα είναι η τάση της πρώτης να επιστρέφει ένα αρκετά υψηλότερο κόστος όταν στο σύνολο δεδομένων υπάρχουν προβλέψεις με μεγάλες αποκλίσεις από τις πραγματικές τιμές. Η ωφέλεια που παρέχει αυτή συμπεριφορά εξαρτάται αποκλειστικά από το στόχο που υπάρχει όταν εκπαιδεύεται ένα νευρωνικό δίκτυο. Συγκεκριμένα, εάν η ύπαρξη μερικών δεδομένων που έχουν μεγάλη απόκλιση από

το μέσο δεδομένο ενός συνόλου δεδομένων (outliers) είναι ανεκτή και δεν είναι μείζονος σημασίας η σωστή πρόβλεψη του νευρωνικού δικτύου για αυτά, τότε είτε η συνάρτηση κόστους που χρησιμοποιεί το μέσο τετραγωνικό σφάλμα πρέπει να αποφεύγεται, είτε πρέπει να αφαιρεθούν τα συγκεκριμένα δεδομένα καθώς θα έχουν λανθασμένα μεγάλη επιρροή στην αξιολόγηση της εκπαίδευσης. Αντιθέτως, εάν γνωρίζουμε πως δεν υπάρχουν αρκετά τέτοια δεδομένα μέσα σε ένα σύνολο δεδομένων ή εάν υπάρχουν και θέλουμε η συμπεριφορά του νευρωνικού δικτύου να γαλουχηθεί και μέσω αυτών, κάτι που βέβαια εγείρει τον κίνδυνο της υπερπροσαρμογής, τότε η χρήση της κρίνεται σκόπιμη. Τέλος, η απουσία απόλυτης τιμής καθιστά φθηνό τον υπολογισμό της παραγώγου. Το μέσο τετραγωνικό σφάλμα εκφράζεται ως

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

2.4.3 Ριζικό μέσο τετραγωνικό σφάλμα (Root mean squared error)

Το ριζικό μέσο τετραγωνικό σφάλμα αποτελεί μία τροποποίηση του μέσου τετραγωνικού σφάλματος και κατέχει τις ίδιες ιδιότητες με εκείνο. Αναφορικά με τη διαχείριση ακραίων τιμών δεδομένων, αποτελεί την καλύτερη επιλογή όταν αυτές είναι αρκετές σε ένα σύνολο δεδομένων, καθώς λόγω της ρίζας μετριάζει τις πολύ υψηλές τιμές που πιθανώς να οδηγήσουν σε λανθασμένα πεσιμιστικές αξιολογήσεις, ενώ παράλληλα, κρατά την ύψωση σε δύναμη η οποία δίνει περισσότερο βάρος στα δεδομένα με ακραίες τιμές, σε αντίθεση με το μέσο απόλυτο σφάλμα το οποίο αντιμετωπίζει κάθε δεδομένο ισάξια, αποφεύγοντας έτσι μία ενδεχόμενη λανθασμένα οπτιμιστική αξιολόγηση. Το μέσο τετραγωνικό σφάλμα εκφράζεται ως

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x_i))^2}$$

2.4.4 Δυαδική διασταυρωμένη εντροπία (Binary cross entropy)

Η δυαδική διασταυρωμένη εντροπία αποτελεί την πιο συνήθης επιλογή σε προβλήματα δυαδικής κατηγοριοποίησης. Για κάθε πρόβλεψη υπολογίζει τη διαφορά του λογαρίθμου αυτής της πρόβλεψης με την πραγματική τιμή, αντιμετωπίζοντας αυτές τις τιμές ως πιθανότητες γεγονός που σε πρακτικό επίπεδο καθιστά την συγκεκριμένη συνάρτηση αρκετά βολική. Στο τέλος επιστρέφει τον μέσο όρο όλων αυτών των διαφορών. Ο τύπος που εκφράζει τη δυαδική διασταυρωμένη εντροπία είναι ο εξής:

$$BCE = -\frac{1}{N} \sum_{i=1}^N (y_i * \log(f(x_i)) + (1 - y_i) * \log(1 - f(x_i)))$$

2.4.5 Κατηγορική διασταυρωμένη εντροπία (Categorical cross entropy)

Η κατηγορική διασταυρωμένη εντροπία είναι, αντίστοιχα, η δημοφιλέστερη συνάρτηση κόστους σε προβλήματα κατηγοριοποίησης πολλαπλών κλάσεων και ακολουθεί την

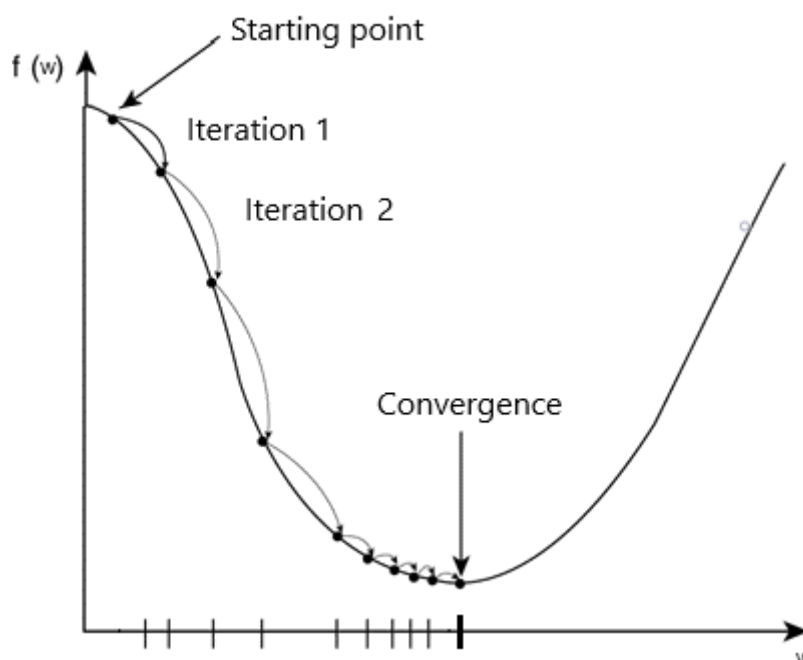
ίδια πιθανοτική προσέγγιση με την διαφορά ότι οι πιθανότητες είναι διαμοιρασμένες σε παραπάνω από δύο κλάσεις. Ο τύπος της είναι ο εξής:

$$CCE = -\frac{1}{N} \sum_{k=1}^M \sum_{i=1}^N (y_i^k * \log(f(x_i)))$$

όπου M είναι ο αριθμός των κλάσεων, N είναι ο αριθμός των δεδομένων εκπαίδευσης, y_i^k είναι η πραγματική τιμή του δεδομένου i το οποίο ανήκει στην κλάση k και $f(x_i)$ είναι η πρόβλεψη για το δεδομένο i .

2.5 Κάθοδος με βάση την κλίση (Gradient descent)

Η κάθοδος με βάση την κλίση είναι ένας προσεγγιστικός αλγόριθμος βελτιστοποίησης. Ένας αλγόριθμος βελτιστοποίησης στοχεύει στην εξεύρεση κατάλληλων εισόδων για μία αντικειμενική συνάρτηση έτσι ώστε αυτή να μεγιστοποιείται ή να ελαχιστοποιείται. Στην περίπτωση μας, η αντικειμενική συνάρτηση είναι η συνάρτηση κόστους και ο αλγόριθμος της καθόδου με βάση την κλίση στοχεύει στην ελαχιστοποίηση της τιμής που παράγεται από αυτή. Ο συνήθης τρόπος για την εύρεση τοπικών ελαχίστων μιας συνάρτησης είναι η επίλυση της εξίσωσης της παραγώγου μιας συνάρτησης με το μηδέν, ωστόσο σε πολύπλοκες συναρτήσεις με πολλές παραμέτρους, όπως είναι η συνάρτηση κόστους ενός νευρωνικού δικτύου, η επίλυση της συγκεκριμένης εξίσωσης καθίσταται εξαιρετικά δύσκολη ή ακόμα και αδύνατη, οδηγώντας έτσι στη χρήση προσεγγιστικών αλγορίθμων όπως αυτός της καθόδου με βάση την κλίση που προσφέρει ταχεία σύγκλιση. Όπως έχει αναφερθεί, κάθε νευρώνας πολλαπλασιάζει κάθε είσοδο με ένα βάρος και η τιμή κάθε βάρους μπορεί να μεταβάλλεται από το ίδιο το νευρωνικό δίκτυο. Η κατάλληλη μεταβολή των βαρών είναι εκείνη που οδηγεί στην ελαχιστοποίηση της συνάρτησης κόστους, συνεπώς αποτελεί βασική λειτουργία της καθόδου με βάση την κλίση.



Εικόνα 2.12: Απεικόνιση της σύγκλισης του αλγορίθμου της καθόδου με βάση την κλίση.

Το σημαντικότερο βήμα του αλγορίθμου είναι το εξής:

$$w_{jk}^{(t+1)} = w_{jk}^{(t)} - \eta * \frac{1}{N} \sum_{i=1}^N \nabla_w l(f(x_i), y_i) = w_{jk}^{(t)} - \eta * \frac{1}{N} \sum_{i=1}^N \nabla_w l(w_i^{(t)} * x_i, y_i) \quad (2.1)$$

όπου $w_{jk}^{(t)}$ είναι το βάρος που συνοδεύει την έξοδο του νευρώνα k , η οποία γίνεται είσοδος στον νευρώνα j κατά την επανάληψη t του αλγορίθμου της καθόδου με βάση την κλίση, $\nabla_w l(f(x_i), y_i)$ είναι η μερική παράγωγος της συνάρτησης απώλειας ως προς w , N είναι ο αριθμός των δεδομένων εκπαίδευσης και η είναι ο ρυθμός εκμάθησης του οποίου η χρησιμότητα θα εξηγηθεί λίγο παρακάτω. Περιγραφόμενος σε φυσική γλώσσα, ο αλγόριθμος σε κάθε επανάληψη υπολογίζει για κάθε πρόβλεψη τις επόμενες τιμές των βαρών που την συνοδεύουν αφαιρώντας την μερική παράγωγο της συνάρτησης κόστους ως προς w πολλαπλασιασμένη με τον ρυθμό εκμάθησης από τις τωρινές τιμές των βαρών της. Κατά το πέρας κάθε επανάληψης, η συνάρτηση κόστους αλλάζει και ελέγχεται εάν ικανοποιεί τα κριτήρια της ελαχιστοποίησης. Όμως, για τον υπολογισμό της μερικής παραγωγού της συνάρτησης κόστους χρειάζεται ο υπολογισμός της μερικής παραγωγού της συνάρτησης απώλειας κάθε δεδομένου εκπαίδευσης, γεγονός που καθιστά την κάθοδο με βάση την κλίση υπολογιστικά ακριβή.

Με σκοπό τη μείωση του υπολογιστικού κόστους, χρησιμοποιούνται παραλλαγές της καθόδου με βάση την κλίση. Μία από αυτές είναι η στοχαστική κάθοδος με βάση την κλίση (stochastic gradient descent), η οποία εκτελεί παρόμοια δουλειά με τη κάθοδο με βάση την κλίση έχοντας ως διαφορά ότι υπολογίζει τη μερική παράγωγο της συνάρτησης απώλειας μόνο ενός δεδομένου εκπαίδευσης το οποίο επιλέγεται τυχαία σε κάθε επανάληψη. Ουσιαστικά, η στοχαστική κάθοδος με βάση την κλίση κάνει μία προσέγγιση της μερικής παραγωγού της συνάρτησης κόστους, συνεπώς η ταχύτητα σύγκλισης του αλγορίθμου είναι περιορισμένη από το θόρυβο που επιφέρουν τα τυχαία επιλεγμένα δεδομένα εκπαίδευσης μέσω των οποίων γίνεται η προσέγγιση. Ωστόσο, στην πράξη παρουσιάζει εξαιρετικά αποτελέσματα. Η μεταβολή των βαρών στην στοχαστική κάθοδο με βάση την κλίση γίνεται ως εξής:

$$w_{jk}^{(t+1)} = w_{jk}^{(t)} - \eta * \nabla_w l(f(x_i), y_i) = w_{jk}^{(t)} - \eta * \nabla_w l(w_i^{(t)} * x_i, y_i) \quad (2.2)$$

όπου i αντιπροσωπεύει τον αριθμό του δεδομένου εκπαίδευσης που επιλέγεται τυχαία σε κάθε επανάληψη.

Μία άλλη αξιοσημείωτη παραλλαγή της καθόδου με βάση την κλίση αποτελεί η κάθοδος με βάση την κλίση μίνι-παρτίδας (mini-batch gradient descent). Ο συγκεκριμένος αλγόριθμος διαθέτει στοιχεία και από τους δύο προαναφερθείς αλγορίθμους καθώς σε κάθε επανάληψη επιλέγει τυχαία ένα υποσύνολο των δεδομένων εκπαίδευσης. Ο αριθμός των δεδομένων που θα επιλεχθούν είναι καθοριστικός στην συμπεριφορά του αλγορίθμου διότι όσο μικρότερος είναι τόσο πιο πολύ θα μοιάζει με τον αλγόριθμο της στοχαστικής καθόδου με βάση την κλίση, ενώ όσο πιο πολλά δεδομένα επιλέγονται ανά επανάληψη η συμπεριφορά του θα τείνει προς εκείνη της καθόδου με βάση την κλίση. Ένας από τους κύριους λόγους που χρησιμοποιείται η κάθοδος με βάση την κλίση κλίση μίνι-παρτίδας είναι η δυνατότητα εκμετάλλευσης της παραλληλίας ενός υπολογιστή, όταν ο αριθμός των δεδομένων που επιλέγεται για τον υπολογισμό των μερικών παραγωγών τους, είναι ανάλογος με τους πόρους του. Η μεταβολή των βαρών στην κάθοδο με βάση την κλίση

μίνι-παρτίδας γίνεται ως εξής:

$$w_{jk}^{(t+1)} = w_{jk}^{(t)} - \eta * \frac{1}{|S^{(t)}|} \sum_{i=1}^{|S^{(t)}|} \nabla_w l(f(x_i), y_i) = w_{jk}^{(t)} - \eta * \frac{1}{|S^{(t)}|} \sum_{i=1}^{|S^{(t)}|} \nabla_w l(w_i^{(t)} * x_i, y_i) \quad (2.3)$$

όπου $|S^{(t)}|$ είναι ο πληθικός αριθμός του υποσυνόλου των δεδομένων εκπαίδευσης που επιλέχθηκε στην επανάληψη t του αλγορίθμου. Και οι τρεις αλγόριθμοι έχουν αρκετά κοινά χαρακτηριστικά μεταξύ τους, ένα εκ των οποίων είναι ο ρυθμός εκμάθησης. Ο ρυθμός εκμάθησης αποτελεί μία υπερπαράμετρο, δηλαδή είναι παράμετρος που ελέγχει τη διαδικασία εκμάθησης. Δίνοντας μικρή τιμή στον ρυθμό εκμάθησης αυξάνεται η πιθανότητα ικανοποιητικής προσέγγισης του ελαχίστου της συνάρτησης κόστους αλλά μειώνεται η ταχύτητα σύγκλισης του αλγορίθμου. Αντιθέτως, δίνοντας μεγάλη τιμή στο ρυθμό εκμάθησης αυξάνεται η ταχύτητα σύγκλισης του αλγορίθμου και μειώνεται η πιθανότητα ικανοποιητικής προσέγγισης του ελαχίστου της συνάρτησης κόστους. Τέλος, οι εξισώσεις (2.1), (2.2) και (2.3) αναδεικνύουν τη σημαντικότητα του προβλήματος των εξαφανισμένων κλίσεων καθώς όσο μικρότερη είναι η κλίση, τόσο μικρότερες μεταβολές γίνονται στα βάρη οδηγώντας σε αργή εκπαίδευση του νευρωνικού δικτύου. Μάλιστα, εάν η κλίση είναι μηδενική τότε δε θα υπάρξει καμία μεταβολή στα βάρη και το νευρωνικό δίκτυο θα πάψει να εκπαιδεύεται πρόωρα.

2.6 Οπισθοδιάδοση (Backpropagation)

Η οπισθοδιάδοση είναι ένας αλγόριθμος που επιτρέπει σε αποδοτικό υπολογισμό παραγώγων. Όπως είδαμε προηγουμένως, ο υπολογισμός κλίσεων ή ισοδύναμα ο υπολογισμός παραγώγων, αποτελεί αναπόσπαστο κομμάτι της εκμάθησης ενός νευρωνικού δικτύου, κάνοντας την χρήση της οπισθοδιάδοσης μία εξαιρετικά σύννηθη επιλογή. Ο λόγος που καθιστά την οπισθοδιάδοση αποδοτική στον υπολογισμό παραγώγων είναι η χρήση του κανόνα αλυσίδας (chain rule). Ο κανόνας της αλυσίδας δίνει τη δυνατότητα μετατροπής μιας συνάρτησης σε μία ισοδύναμη μορφή η οποία πολλές φορές προσφέρει ευκολότερο υπολογισμό της παραγώγου. Η υπολογισμός παραγώγου σύνθετης συνάρτησης είναι μία περίπτωση που αναδεικνύει την χρησιμότητα του κανόνας της αλυσίδας. Για παράδειγμα, έστω η συναρτήση:

$$H(x) = f \circ g = f(g(x))$$

όπου f, g είναι δύο παραγωγίσιμες συναρτήσεις. Χρησιμοποιώντας τον κανόνα αλυσίδας, η παράγωγος της συνάρτησης μπορεί να εκφραστεί ως:

$$H'(x) = (f' \circ g)g' = f'(g(x))g'(x)$$

Εναλλακτικά, χρησιμοποιώντας τον συμβολισμό Leibniz, για $y = f(u)$ και $u = g(x)$ η παράγωγος της συνάρτησης εκφράζεται ως:

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

Έχοντας πλέον κατανοήσει μία απλή εφαρμογή του κανόνα αλυσίδας σε σύνθετες συναρτήσεις μπορούμε να την γενικεύσουμε. Έστω η συνάρτηση

$$H(x) = f_1 \circ (f_2 \circ \dots (f_{n-1} \circ f_k))$$

όπου κάθε συνάρτηση f_i , $1 \leq i \leq k$ είναι παραγωγίσιμη. Χρησιμοποιώντας τον κανόνα αλυσίδας η παράγωγός της εκφράζεται ως:

$$\frac{df_1}{dx} = \frac{df_1}{df_2} \frac{df_2}{df_3} \cdots \frac{df_n}{dx}$$

Η συνάρτηση απώλειας είναι και αυτή μία σύνθετη συνάρτηση. Συνδυάζοντας όλα όσα έχουμε πει για τους νευρώνες, μπορούμε να εκφράσουμε την έξοδο οποιουδήποτε νευρώνα μέσα σε ένα νευρωνικό δίκτυο, εκτός φυσικά από εκείνους που ανήκουν στο στρώμα εισόδου, ως

$$a_j^{(L)} = f_j^{(L)} \left(\underbrace{\sum_{k=1}^N w_{jk}^{(L)} a_k^{(L-1)} + b_j^{(L)}}_{z_j^{(L)}} \right) \quad (2.4)$$

όπου L είναι οποιοδήποτε στρώμα του νευρωνικού δικτύου εκτός του στρώματος εισόδου, j είναι οποιοσδήποτε νευρώνας ανήκει στο στρώμα L , f είναι η συνάρτηση ενεργοποίησης του νευρώνα j , είναι ο αριθμός των νευρώνων που η έξοδος του αποτελεί είσοδο για τον νευρώνα j , $w_{jk}^{(L)}$ είναι το βάρος του νευρώνα j που αντιστοιχεί στην έξοδο του νευρώνα k η οποία είναι είσοδος για τον νευρώνα j , $a_k^{(L-1)}$ είναι η έξοδος του νευρώνα k που ανήκει στο στρώμα $L - 1$ και $b_j^{(L)}$ είναι η μεροληψία του νευρώνα j . Διαισθητικά γίνεται αντιληπτό ότι οποιαδήποτε αλλαγή στα βάρη ενός νευρώνα, επιφέρει αλλαγή στην είσοδο της συνάρτησης ενεργοποίησης του, η οποία με τη σειρά της εγείρει αλλαγή στην έξοδο του νευρώνα, γεγονός που εν τέλει αλλάζει την τιμή της συνάρτησης απώλειας του νευρώνα και κατ' επέκταση τη συνάρτηση κόστους του νευρωνικού δικτύου. Σκεπτόμενοι αφηρημένα θα μπορούσαμε να πούμε πως η συγκεκριμένη συμπεριφορά προσομοιώνει μία αλυσίδα. Βασισμένη σε αυτή τη λογική γίνεται και ο υπολογισμός κλίσεων που χρησιμοποιούνται στην κάθοδο με βάση την κλίση. Ο υπολογισμός της κλίσης της συνάρτησης κόστους ενός νευρώνα ως προς ένα βάρος $w_{ji}^{(L)}$ για ένα δεδομένο εκπαιδευσης δίνεται από την εξίσωση

$$\frac{\partial C_m}{\partial w_{ji}^{(L)}} = \frac{\sum_{k=1}^N \partial(w_{jk}^{(L)} a_k^{(L-1)} + b_j^{(L)})}{\partial w_{ji}^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_m}{\partial a_j^{(L)}} \quad (2.5)$$

με τον συμβολισμό να είναι ο ίδιος με την εξίσωση (2.4). Ωστόσο ο υπολογισμός του τρίτου όρου της εξίσωσης διαφέρει ανάλογα με το αν ένας νευρώνας ανήκει στο στρώμα εξόδου ή σε κρυφό στρώμα. Όταν ένας νευρώνας ανήκει στο στρώμα εξόδου, η συνάρτηση κόστους ως προς ένα βάρος $w_{ji}^{(L)}$ για ένα δεδομένο εκπαιδευσης ταυτίζεται με την συνάρτηση απώλειάς του, άρα υπολογίζοντας τις μερικές παραγώγους κάθε όρου της εξίσωσης (2.5) έχουμε ισοδύναμα ότι

$$\frac{\partial C_m}{\partial w_{ji}^{(L)}} = a_j^{(L-1)} f_j'^{(L)}(z_j^{(L)}) l'(z_j^{(L)}) \quad (2.6)$$

Όπως αναφέρεται, οι εξισώσεις (2.5) και (2.6) αφορούν ένα δεδομένο κάθε φορά, οπότε συνολικά ο υπολογισμός της κλίσης της συνάρτησης κόστους ως προς ένα βάρος $w_{ji}^{(L)}$

δίνεται από την εξίσωση

$$\frac{\partial C}{\partial w_{ji}^{(L)}} = \frac{1}{N} \sum_{k=1}^N \frac{\partial C_k}{\partial w_{ji}^{(L)}} \quad (2.7)$$

Απο την άλλη πλευρά, όταν ένας νευρώνας ανήκει σε κρυφό στρώμα, η συνάρτηση κόστους ως προς ένα βάρος $w_{ji}^{(L)}$ για ένα δεδομένο εκπαίδευσης εξαρτάται από τα κόστη όλων των νευρώνων οι οποίοι δέχονται ως είσοδο την έξοδο του. Με άλλα λόγια, η διαφορά είναι ότι ένας νευρώνας που ανήκει σε κρυφό στρώμα έχει, δυνητικά, παραπάνω από έναν παραλήπτες, σε αντίθεση με έναν νευρώνα που ανήκει στο στρώμα εξόδου ο οποίος έχει αποκλειστικά έναν παραλήπτη, τη συνάρτηση απώλειας. Συνεπώς, ο τρίτος όρος της εξίσωσης (2.5) εκφράζεται ως

$$\frac{\partial C_m}{\partial a_j^{(L)}} = \sum_{k=1}^N \frac{\partial z_k^{(L+1)}}{\partial a_j^{(L)}} \frac{\partial a_k^{(L+1)}}{\partial z_k^{(L+1)}} \frac{\partial C_m}{\partial a_k^{(L+1)}} = \sum_{k=1}^N w_{kj}^{(L+1)} f_k'^{(L+1)}(z_k^{(L+1)}) l'(z_k^{(L+1)}) \quad (2.8)$$

Ωστόσο, η εξίσωση (2.8) αφορά ένα δεδομένο εκπαίδευσης, συνεπώς για να ληφθούν υπόψη όλα τα δεδομένα εκπαίδευσης χρησιμοποιείται η εξίσωση

$$\frac{\partial C}{\partial a_j^{(L)}} = \frac{1}{N} \sum_{k=1}^N \frac{\partial C_k}{\partial a_j^{(L)}} \quad (2.9)$$

Επίσης, υπολογίζοντας τις μερικές παραγώγους κάθε όρου της εξίσωσης (2.5) έχουμε ισοδύναμα ότι όταν ένας νευρώνας ανήκει σε κρυφό στρώμα, η συνάρτηση κόστους ως προς ένα βάρος $w_{ji}^{(L)}$ για ένα δεδομένο εκπαίδευσης εκφράζεται ως

$$\frac{\partial C_m}{\partial w_{ji}^{(L)}} = a_j^{(L-1)} f'^{(L)}(z_j^{(L)}) \sum_{k=1}^N w_{kj}^{(L+1)} f_k'^{(L+1)}(z_k^{(L+1)}) l'(z_k^{(L+1)})$$

ενώ για όλα τα δεδομένα εκπαίδευσης εκφράζεται ως

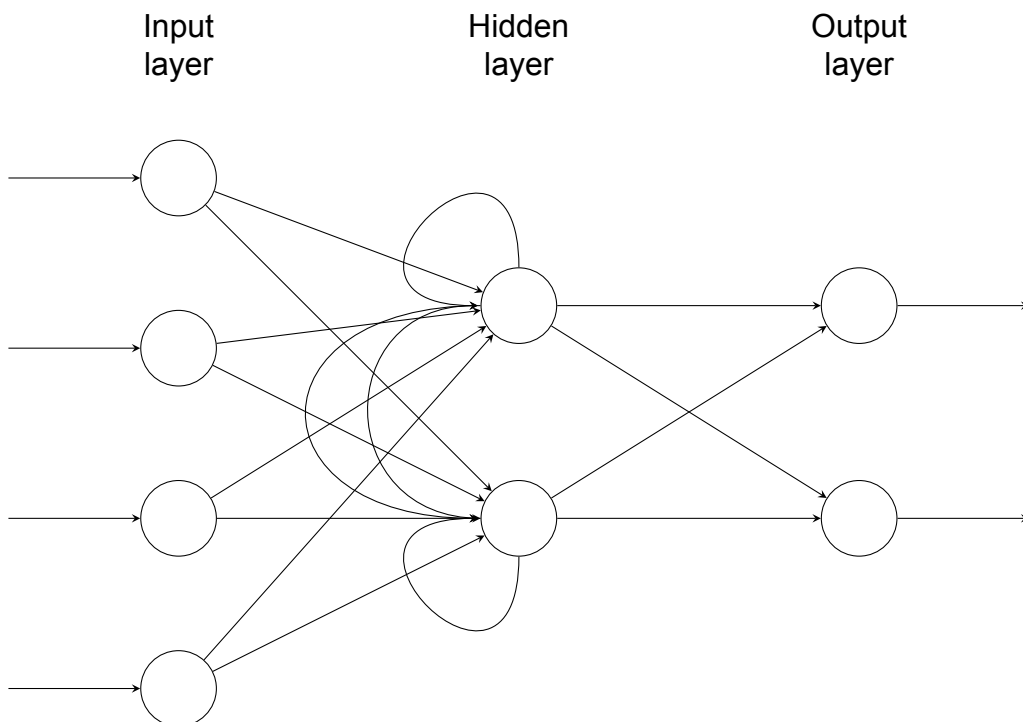
$$\frac{\partial C_k}{\partial w_{ji}^{(L)}} = \frac{1}{N} \sum_{k=1}^N a_j^{(L-1)} f'^{(L)}(z_j^{(L)}) \frac{\partial C}{\partial a_j^{(L)}}$$

Εν τέλει, όπως είδαμε η διαδικασία της οπισθοδιάδοσης χρησιμοποιεί την έξοδο της συνάρτησης ενεργοποίησης. Συνεπώς, για να έχει νόημα η διαδικασία της οπισθοδιάδοσης, είναι απαραίτητο η έξοδος της συνάρτησης ενεργοποίησης να μην είναι ανεξάρτητη από την είσοδο. Σε διαφορετική περίπτωση, είναι αδύνατο να γίνουν αλλαγές στα βάρη με σκοπό την ελαχιστοποίηση της συνάρτησης κόστους καθώς οποιαδήποτε τιμή των βαρών θα επιφέρει την ίδια έξοδο της συνάρτησης ενεργοποίησης.

2.7 Αναδρομικό νευρωνικό δίκτυο (Recurrent neural network – RNN)

Το αναδρομικό νευρωνικό δίκτυο, γνωστό και ως RNN, αποτελεί μία συγκεκριμένη αρχιτεκτονική νευρωνικών δικτύων η οποία αρχικά σχεδιάστηκε με στόχο την αποτελεσματική αντιμετώπιση σειριακών δεδομένων (sequential data), δηλαδή

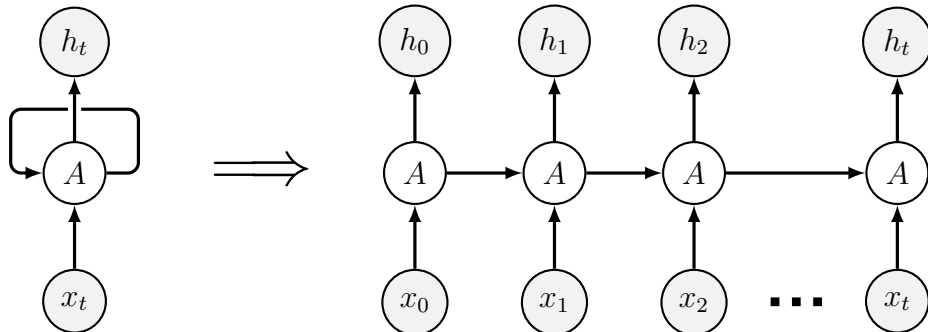
δεδομένων που η σειρά με την οποία αντιμετωπίζονται διαδραματίζει σημαντικό ρόλο στο τελικό αποτέλεσμα. Παρόλο τον αρχικό σχεδιασμό, έχει διαπιστωθεί ανά τα χρόνια ότι τα αναδρομικά νευρωνικά δίκτυα συμπεριφέρονται εξαιρετικά και σε μη σειριακά δεδομένα καθιστώντας τα, εν τέλει, εφαρμόσιμα σε πάρα πολλούς τομείς όπως είναι η αναγνώριση λόγου, η παραγωγή κειμένου και η αναγνώριση αντικειμένων σε εικόνες και βίντεο. Η ειδοποιός διαφορά της συγκεκριμένης αρχιτεκτονικής από τις υπόλοιπες είναι ο μηχανισμός που διαθέτει ο οποίος λειτουργεί κατά κάποιον τρόπο ως μνήμη, δημιουργώντας εξαρτήσεις μεταξύ των δεδομένων και εισάγοντας την έννοια του χρόνου στο νευρωνικό δίκτυο, σε αντίθεση με τις υπόλοιπες αρχιτεκτονικές όπου κάθε δεδομένο αντιμετωπίζεται ανεξάρτητα. Η πληροφορία που έχει αποθηκευμένη ένας νευρώνας κάθε δεδομένη χρονική στιγμή ονομάζεται εσωτερική κατάσταση (internal state). Για την επίτευξη της λειτουργίας αυτού του μηχανισμού δίνεται η δυνατότητα στους νευρώνες των κρυφών στρωμάτων να δημιουργούν κύκλους, όσον αφορά τον τρόπο με τον οποίο μεταβιβάζουν πληροφορία. Αυτό σημαίνει ότι κάθε νευρώνας έχει τη δυνατότητα να μεταβιβάζει την έξοδό του και σε νευρώνες που ανήκουν σε προηγούμενο ή στο ίδιο στρώμα με αυτόν, συμπεριλαμβανομένου και του εαυτού τους. Διαισθητικά, θα μπορούσαμε να πούμε ότι ο τρόπος με τον οποίο λειτουργούν τα αναδρομικά νευρωνικά δίκτυα έχει εμπνευστεί από την επιρροή που ασκούν στην σκέψη των ανθρώπων οι παλιές εμπειρίες και οι μνήμες τους. Η Εικόνα 2.13 απεικονίζει μία πιθανή δομή ενός αναδρομικού νευρωνικού δικτύου, όπου κάθε κόμβος του κρυφού στρώματος μεταβιβάζει την έξοδό του, όχι μόνο στους κόμβους του στρώματος εξόδου, αλλά και στον αντίστοιχο άλλο κόμβο του κρυφού στρώματος και στον εαυτό του.



Εικόνα 2.13: Παράδειγμα δομής αναδρομικού νευρωνικού δικτύου.

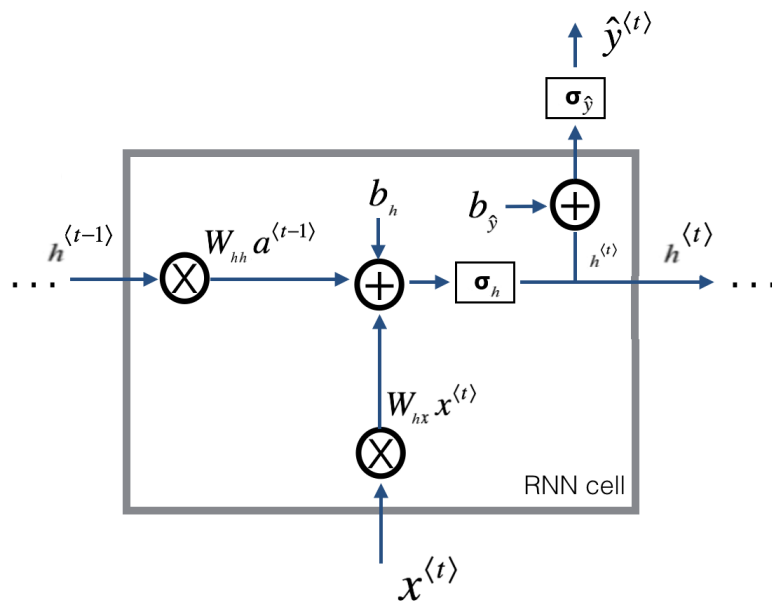
Η Εικόνα 2.14 εστιάζει στην λειτουργία ενός νευρώνα μεμονωμένα και απεικονίζει δύο πολύ συχνά χρησιμοποιούμενες αναπαράστασεις αυτής. Συγκεκριμένα, στο αριστερό μέρος της συνεπαγωγής παρατηρείται μια απλοϊκή αναπαράσταση, η οποία κάνει ξεκάθαρο το γεγονός ότι ο νευρώνας συμπεριφέρεται πανομοιότυπα για κάθε είσοδο x_t , δηλαδή ανανεώνει την εσωτερική κατάστασή του s_t , και παράγει μία έξοδο y_t .

Όταν επιλέγεται αυτή η αναπαράσταση, λέμε ότι το αναδρομικό νευρωνικό δίκτυο είναι διπλωμένο (folded). Στο δεξί μέρος της συνεπαγωγής παρατηρείται μία πιο λεπτομερής αναπαράσταση καθώς εμφανίζονται όλες οι διαδοχικές εσωτερικές καταστάσεις s_i και οι εξόδοι y_i , για κάθε είσοδο x_i , όπου $0 \leq i \leq t$. Σε τούτη την αναπαράσταση, λέμε ότι το αναδρομικό νευρωνικό δίκτυο είναι ξεδιπλωμένο (unfolded).



Εικόνα 2.14: Απεικόνιση δύο αναπαραστάσεων ενός νευρώνα της αρχιτεκτονικής RNN.

Τέλος, η Εικόνα 2.15 αναδεικνύει τον ακριβή τρόπο με τον οποίο ένας νευρώνας πραγματοποιεί την διαδικασία που περιγράφηκε παραπάνω. Ο νευρώνας, αρχικά, πολλαπλασιάζει την είσοδο $x^{(t)}$ και την προηγούμενη εσωτερική κατάσταση του $h^{(t-1)}$ με τα βάρη W_{hx} και W_{hh} αντίστοιχα. Στη συνέχεια, στο άθροισμα των δύο πολλαπλασιασμών προστίθενται η πόλωση που αφορά την εσωτερική κατάσταση του νευρώνα b_h και το συνολικό αποτέλεσμα αποτελεί είσοδο για την συνάρτηση ενεργοποίησης της εσωτερικής κατάστασης σ_h . Ύστερα, η έξοδος της συνάρτησης ενεργοποίησης ακολουθεί δύο μονοπάτια. Στο ένα μονοπάτι, αποτελεί αυτούσια την νέα εσωτερική κατάσταση του νευρώνα, ενώ στο άλλο προστίθεται η πόλωση της εξόδου b_y και στη συνέχεια τροφοδοτείται στη συνάρτηση ενεργοποίησης της εξόδου, η οποία παράγει, εν τέλει, την έξοδο $\hat{y}^{(t)}$.



Εικόνα 2.15: Απεικόνιση ενός κελιού RNN.

Συνοψίζοντας την παραπάνω συμπεριφορά, καταλήγουμε ότι για κάθε δεδομένη χρονική

στιγμή t , η εσωτερική κατάσταση ενός νευρώνα εκφράζεται ως

$$h^{(t)} = \sigma_h(W_{hx}x^{(t)} + W_{hh}h^{(t-1)} + b_h)$$

όπου τα βάρη και οι πολώσεις παραμένουν ίδιες σε κάθε ανανέωση της εσωτερικής κατάστασης, και η έξοδος του εκφράζεται ως

$$\hat{y}^{(t)} = \sigma_h(W_{yh}h^{(t)} + b_y)$$

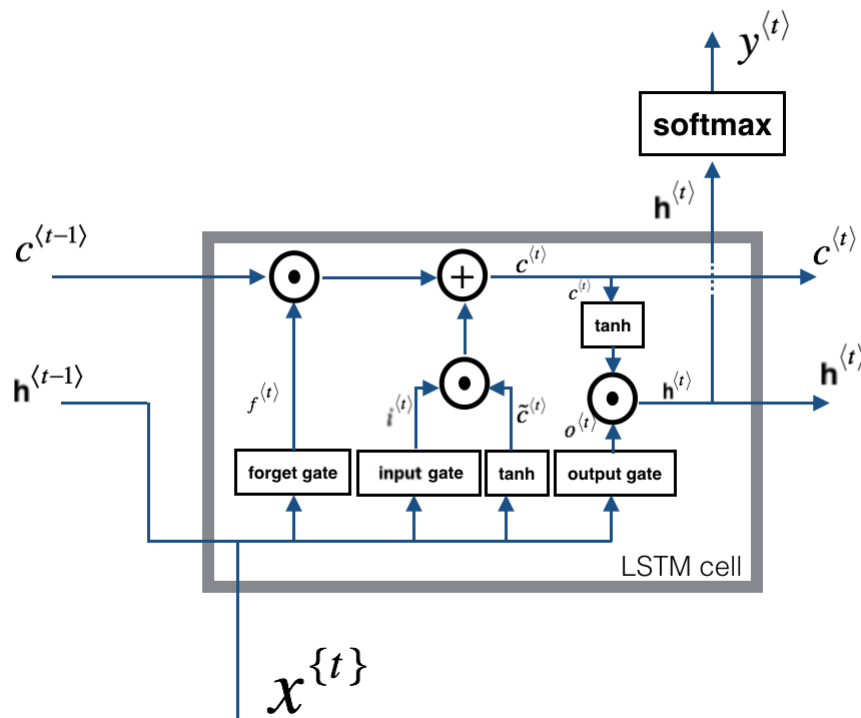
Παρόλο που τα αναδρομικά νευρωνικά δίκτυα είναι ικανά να μάθουν εξαρτήσεις που υπάρχουν μεταξύ των δεδομένων, η ικανότητα αυτή είναι δυσανάλογη του χρόνου που εκπαιδεύεται το νευρωνικό δίκτυο, ή αλλιώς, των φορών που εκτελείται η οπισθοδιάδοση, παρουσιάζοντας αδυναμία στην εκμάθηση μακροχρόνιων εξαρτήσεων (long-term dependencies). Ο λόγος που συμβαίνει αυτό ανάγεται στην επιρρέπεια της συγκεκριμένης αρχιτεκτονικής στο πρόβλημα των εξαφανισμένων κλίσεων αλλά και στο πρόβλημα των εκρηγνυόμενων κλίσεων (exploding gradient problem), το οποίο αποτελεί το αντίθετο πρόβλημα από αυτό των εξαφανισμένων κλίσεων, δηλαδή οι κλίσεις λαμβάνουν πολύ υψηλές τιμές. Αυτό, πέρα από πιθανά υπολογιστικά προβλήματα, επιφέρει μεγάλες και απότομες αλλαγές στα βάρη καθιστώντας το νευρωνικό δίκτυο ασταθές και ανίκανο να εκπαιδευτεί έτσι ώστε να ελαχιστοποιηθεί το σφάλμα. Τέλος, η οπισθοδιάδοση που εφαρμόζεται στα αναδρομικά νευρωνικά δίκτυα, τυπικά, ονομάζεται οπισθοδιάδοση στον χρόνο (backpropagation through time) και η εφαρμογή της απαιτεί το ξεδίπλωμα του νευρωνικού δικτύου όπως είδαμε στην Εικόνα 2.19.

2.8 Νευρωνικό δίκτυο μακροχρόνιας και βραχυχρόνιας μνήμης (Long short-term memory - LSTM)

Το νευρωνικό δίκτυο βραχυχρόνιας και μακροχρόνιας μνήμης, γνωστό και ως LSTM, αποτελεί μία τροποποιημένη εκδοχή των αναδρομικών νευρωνικών δικτύων που προτάθηκε για πρώτη φορά το 1997 από τους Sepp Hochreiter και Jürgen Schmidhuber. Το πλεονέκτημα του σε σχέση με το κλασικό αναδρομικό νευρωνικό δίκτυο, όπως αυτό περιγράφεται παραπάνω, είναι το γεγονός ότι καταφέρνει να ξεπεράσει τα προβλήματα που αφορούν την εκπαίδευσή του, έχοντας έτσι την δυνατότητα να αφομοιώσει μακροχρόνιες εξαρτήσεις μεταξύ των δεδομένων, κάτι που κάνει το νευρωνικό δίκτυο να συμπεριφέρεται ακόμα καλύτερα στα προβλήματα που του ανατίθενται. Ο λόγος για τον οποίο αυτό καθίσταται εφικτό είναι η περίπλοκη αρχιτεκτονική που ακολουθείται, της οποίας η ειδοποιός διαφορά με την αρχιτεκτονική RNN είναι η ύπαρξη ενός δομικού στοιχείου που ονομάζεται κατάσταση κελιού (cell state). Η κατάσταση κελιού είναι αυτή που, εν τέλει, επιτρέπει στο νευρωνικό δίκτυο να αφομοιώνει τις μακροχρόνιες εξαρτήσεις. Φυσικά, αυτή η συμπεριφορά είναι αποτέλεσμα ενός συνδυασμού μηχανισμών οι οποίοι θα εξηγηθούν παρακάτω. Προτού γίνει αυτό, κρίνεται σκόπιμο να αναφερθούν οι λειτουργίες που επιτελεί ένας νευρώνας της αρχιτεκτονικής LSTM σε υψηλό επίπεδο:

1. Ξεχνάει όσα μέρη πληροφορίας της προηγούμενης εσωτερικής κατάστασής του δεν είναι πλέον χρήσιμα.
2. Υπολογίζει ποια μέρη πληροφορίας της εισόδου είναι χρήσιμα.
3. Ανανεώνει την εσωτερική κατάστασή του.

4. Παράγει μία έξοδο.



Εικόνα 2.16: Απεικόνιση ενός κελιού LSTM.

Η Εικόνα 2.16 απεικονίζει τα συστατικά ενός νευρώνα της αρχιτεκτονικής LSTM. Αρχικά, κάθε νευρώνας αποτελείται από τρεις πύλες: την πύλη λήθης (forget gate), την πύλη εισόδου (input gate) και την πύλη εξόδου (output gate). Η δουλειά κάθε πύλης είναι να κρίνει ποια μέρη της πληροφορίας που συνοδεύει την είσοδό της είναι χρήσιμα για τον νευρώνα. Ο τρόπος που κάθε πύλη ελέγχει το μέρος της πληροφορίας που εισέρχεται είναι διοχετεύοντας την είσοδό της στην σιγμοειδή συνάρτηση η οποία, όπως έχουμε δει, επιστρέφει έναν αριθμό μεταξύ του μηδέν και του ένα. Όσο μεγαλύτερη τιμή παράξει η σιγμοειδής συνάρτηση, τόσο σημαντικότερη κρίνεται η πληροφορία που περιέχει η είσοδος της πύλης, αποκόπτοντας λιγότερα μέρη της. Η προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ και η είσοδος $x^{(t)}$ του νευρώνα, αποτελούν κοινή είσοδο και για τις τρεις πύλες. Η διαφοροποίηση των τριών πηγάζει από την αντιστοίχιση αυτών των εισόδων με διαφορετικά βάρη σε κάθε πύλη, από την ύπαρξη διαφορετικής πόλωσης σε κάθε πύλη και από την διαφορετική αντιμετώπιση που έχει στη συνέχεια η έξοδος κάθε πύλης. Επίσης, η οριζόντια γραμμή στο πάνω μέρος αντιπροσωπεύει την κατάσταση κελιού η οποία σε πρώτη φάση πολλαπλασιάζει το αποτέλεσμα που προκύπτει από την πύλη λήθης $f^{(t)}$ με την προηγούμενη κατάσταση κελιού $c^{(t-1)}$ και στη συνέχεια προσθέτει το αποτέλεσμα της πύλης εισόδου. Αυτό που προκύπτει είναι η νέα κατάσταση κελιού $c^{(t)}$ η οποία έχει αποθηκευμένες τις πληροφορίες που κρίθηκαν χρήσιμες από τις εισόδους που έχει δεχθεί κατά τη διάρκεια των t χρονικών βημάτων ο νευρώνας. Παρακάτω περιγράφονται αναλυτικότερα οι λειτουργίες κάθε πύλης:

1. Πύλη λήθης (Forget gate): Η πύλη λήθης είναι υπεύθυνη να κρίνει ποια μέρη πληροφορίας δεν είναι πλέον χρήσιμα και να τα αποκόψει. Η έξοδος της εκφράζεται από την εξίσωση

$$f^{(t)} = \sigma(W_{fx}x^{(t)} + W_{fh}h^{(t-1)} + b_f)$$

όπου σ είναι η σιγμοειδής συνάρτηση, W_{fx} είναι τα βάρη που αντιστοιχίζουν την είσοδο $x^{(t)}$ με την πύλη λήθης, W_{fh} είναι τα βάρη που αντιστοιχίζουν την προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ με την πύλη λήθης και b_f είναι η πόλωση της πύλης λήθης. Μετά την παραγωγή της εξόδου εφαρμόζεται πολλαπλασιασμός στοιχείο προς στοιχείο (element-wise multiplication) μεταξύ αυτής και της προηγούμενης κατάστασης κελιού.

2. Πύλη εισόδου (Input gate): Η πύλη εισόδου είναι υπεύθυνη να κρίνει ποια μέρη πληροφορίας είναι χρήσιμα και πρέπει να προστεθούν στη νέα κατάσταση κελιού. Η έξοδος της εκφράζεται από την εξίσωση

$$i^{(t)} = \sigma(W_{ix}x^{(t)} + W_{ih}h^{(t-1)} + b_i)$$

όπου σ είναι η σιγμοειδής συνάρτηση, W_{ix} είναι τα βάρη που αντιστοιχίζουν την είσοδο $x^{(t)}$ με την πύλη εισόδου, W_{ih} είναι τα βάρη που αντιστοιχίζουν την προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ με την πύλη εισόδου και b_i είναι η πόλωση της πύλης εισόδου. Παράλληλα, η είσοδος και η προηγούμενη κρυφή κατάσταση του νευρώνα, μαζί με τα αντίστοιχα βάρη, παρέχονται ως είσοδος στην συνάρτηση ενεργοποίησης υπερβολικής εφαπτομένης της οποίας η έξοδος εκφράζεται από την εξίσωση

$$\hat{c}^{(t)} = \tanh(W_{cx}x^{(t)} + W_{ch}h^{(t-1)} + b_c)$$

όπου \tanh είναι η υπερβολική εφαπτομένη, W_{cx} είναι τα βάρη που αντιστοιχίζουν την είσοδο $x^{(t)}$ με την υπερβολική εφαπτομένη, W_{ch} είναι τα βάρη που αντιστοιχίζουν την προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ με την υπερβολική εφαπτομένη και b_c είναι η πόλωση της πύλης εισόδου. Στη συνέχεια, στην έξοδο της πύλης εισόδου και της υπερβολικής εφαπτομένης εφαρμόζεται πολλαπλασιασμός στοιχείο προς στοιχείο. Το αποτέλεσμα αυτής της πράξης προστίθεται στο αποτέλεσμα του πολλαπλασιασμού της πύλης λήθης με την προηγούμενη κατάσταση κελιού. Αυτή η συμπεριφορά εκφράζεται από την εξίσωση

$$c^{(t)} = \hat{c}^{(t)} \odot i^{(t)} + c^{(t-1)} \odot f^{(t)}$$

3. Πύλη εξόδου (Output gate):

Η πύλη εξόδου είναι υπεύθυνη για τα μέρη της πληροφορίας που θα περαστούν στην έξοδο του νευρώνα $y^{(t)}$ και στην κρυφή κατάστασή του $h^{(t)}$. Η έξοδος της συγκεκριμένης πύλης εκφράζεται από την εξίσωση

$$o^{(t)} = \sigma(W_{ox}x^{(t)} + W_{oh}h^{(t-1)} + b_o) \quad (2.10)$$

όπου σ είναι η σιγμοειδής συνάρτηση, W_{ox} είναι τα βάρη που αντιστοιχίζουν την είσοδο $x^{(t)}$ με την πύλη εξόδου, W_{oh} είναι τα βάρη που αντιστοιχίζουν την προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ με την πύλη εξόδου και b_o είναι η πόλωση της πύλης εξόδου. Στη συνέχεια, η έξοδος της συγκεκριμένης πύλης πολλαπλασιάζεται στοιχείο προς στοιχείο με το αποτέλεσμα της εφαρμογής της εφαπτομένης συνάρτησης ενεργοποίησης στην κατάσταση κελιού. Ο κυριότερος λόγος που χρησιμοποιείται η εφαπτομένη ως συνάρτηση ενεργοποίησης στις πύλες εισόδου και εξόδου είναι η ανοχή της στο πρόβλημα των εξαφανισμένων και

εκρηγνυόμενων κλίσεων. Το αποτέλεσμα αυτών των πράξεων αποτελεί την κρυφή κατάσταση $h(t)$ η οποία εκφράζεται ως

$$h^{(t)} = o^{(t)} \odot \sigma(c^{(t)})$$

Τέλος, η έξοδος του νευρώνα $y^{(t)}$ προκύπτει διοχετεύοντας την κρυφή κατάσταση $h(t)$ στη συνάρτηση ενεργοποίησης Softmax.

2.9 Συνελικτικό νευρωνικό δίκτυο (Convolutional neural network – CNN)

Το συνελικτικό νευρωνικό δίκτυο, γνωστό και ως CNN ή ConvnNet, αποτελεί άλλη μία αρχιτεκτονική η οποία χρησιμοποιείται ευρέως για την επίλυση πολλών διαφορετικών προβλημάτων. Η συγκεκριμένη αρχιτεκτονική σχεδιάστηκε για την αντιμετώπιση προβλημάτων που αφορούν την ανάλυση εικόνας. Ακόμα και σήμερα, αποτελεί την δημοφιλέστερη επιλογή για την αντιμετώπιση τέτοιου είδους προβλημάτων, ενώ παράλληλα έχει παρατηρηθεί ότι συμπεριφέρεται εξαιρετικά και σε διαφορετικές οικογένειες προβλημάτων. Ενδεικτικά, κάποιες εφαρμογές του συνελικτικού νευρωνικού δικτύου είναι η αυτόνομη οδήγηση, η πρόγνωση μέσω της ανάλυσης ιατρικών εικόνων και η ανάλυση κειμένων. Χαρακτηριστικό στοιχείο του συνελικτικού νευρωνικού δικτύου είναι η ικανότητά του να εκμεταλλεύεται την χωρική πληροφορία (spatial information) που συνοδεύει την είσοδό του. Για παράδειγμα, ας θεωρήσουμε ότι η είσοδος ενός νευρωνικού δικτύου αποτελείται από δισδιάστατες εικόνες. Γενικά, οι δισδιάστατες εικόνες αναπαριστώνται ως δισδιάστατοι πίνακες όπου κάθε στοιχείο τους αντιστοιχεί σε ένα εικονοστοιχείο. Όσο μία εικόνα αναπαρίσταται κατά αυτόν τον τρόπο, η χωρική πληροφορία της παραμένει ακέραιη. Ωστόσο, για την επεξεργασία αυτών των εικόνων, πολλές αρχιτεκτονικές θα απαιτούσαν τη δημιουργία τόσων νευρώνων στο στρώμα εισόδου όσος και ο αριθμός των εικονοστοιχείων της εικόνας. Αυτό θα σήμαινε ότι ο δισδιάστατος πίνακας που περιγράφει κάθε εικόνα θα έπρεπε να μετατραπεί σε έναν μονοδιάστατο, ο οποίος πλέον δε θα αποτελούσε απεικόνιση καμίας εικόνας, χάνοντας με αυτόν τον τρόπο σημαντικό όγκο της χωρικής πληροφορίας. Από την άλλη πλευρά, η αρχιτεκτονική του συνελικτικού νευρωνικού δικτύου εισάγει δύο πράξεις οι οποίες συνεισφέρουν στην αποτροπή της παραπάνω συμπεριφοράς, ονόματι συνέλιξη (convolution) και συγκέντρωση (pooling). Η πράξη της συνέλιξης λειτουργεί ως εξαγωγέας χαρακτηριστικών (feature extractor) της εισόδου. Ο τρόπος με τον οποίο καταφέρνει να εξάγει τα χαρακτηριστικά είναι αρκετά απλός. Αρχικά ορίζεται ένα φίλτρο το οποίο εφαρμόζεται στον πίνακα μιας εικόνας. Οι φορές που θα εφαρμοστεί το φίλτρο συνολικά στην εικόνα και ο τρόπος που το φίλτρο θα ομαδοποιήσει τα στοιχεία του πίνακα εξαρτώνται τόσο από το μέγεθος του φίλτρου όσο και από το μέγεθος της μετατόπισης του φίλτρου πάνω στον πίνακα της εικόνας μετά από κάθε εφαρμογή του, γνωστό και ως βήμα (stride). Η Εικόνα 2.17 δείχνει ένα παράδειγμα εφαρμογής της συνέλιξης, όπου ο πίνακας εισόδου I είναι μεγέθους 7×7 , το φίλτρο K είναι μεγέθους 3×3 και ο χάρτης χαρακτηριστικών είναι μεγέθους 5×5 . Φυσικά, το μέγεθος του χάρτη χαρακτηριστικών υποδεικνύει ότι το βήμα που χρησιμοποιήθηκε είναι ίσο με ένα. Αν είχε χρησιμοποιηθεί βήμα ίσο με δύο τότε ο τελικός υποπίνακας θα ήταν μεγέθους 3×3 .

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 & 4 & 1 \\ 1 & 2 & 4 & 3 & 3 \\ 1 & 2 & 3 & 4 & 1 \\ 1 & 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 & 0 \end{pmatrix}$$

I K $I * K$

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 & 4 & 1 \\ 1 & 2 & 4 & 3 & 3 \\ 1 & 2 & 3 & 4 & 1 \\ 1 & 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 & 0 \end{pmatrix}$$

I K $I * K$

Εικόνα 2.17: Παράδειγμα εφαρμογής της συνέλιξης.

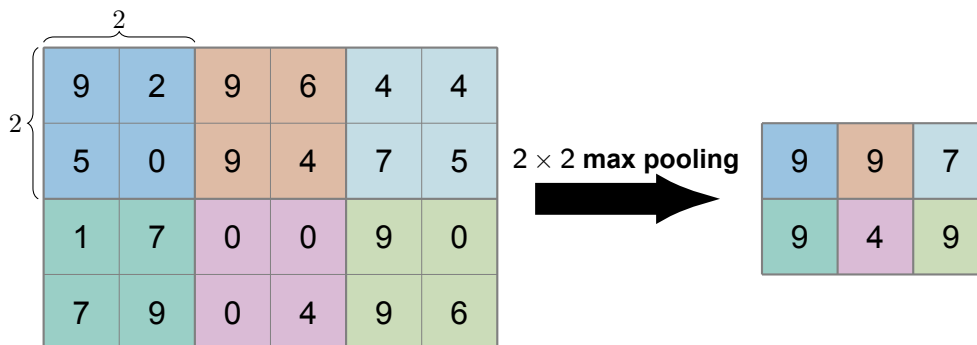
Έχοντας πλέον εξηγήσει την πράξη της συνέλιξης, μπορούμε να αναλύσουμε τον τρόπο με τον οποίο την εκμεταλλεύεται η συγκεκριμένη αρχιτεκτονική. Συγκεκριμένα, κάθε στοιχείο του υποπίνακα που προκύπτει αντιστοιχεί σε έναν νευρώνα του νευρωνικού δικτύου, συνεπώς αφότου εκτελεστεί η πράξη της συνέλιξης, προστίθεται σε κάθε νευρώνα η πόλωση και στη συνέχεια κάθε νευρώνας ενεργοποιείται μέσω μιας συνάρτησης ενεργοποίησης. Το αποτέλεσμα αυτών των διαδοχικών πράξεων ονομάζεται χάρτης χαρακτηριστικών (feature map). Κάθε στοιχείο του χάρτη χαρακτηριστικών είναι ένας νευρώνας ο οποίος, λόγω της εφαρμογής του φίλτρου, αντιστοιχίζεται με ένα σύνολο από γειτονικά στοιχεία του πίνακα στον οποίο εφαρμόστηκε το φίλτρο. Κάθε ένα από αυτά τα γειτονικά στοιχεία αποτελεί είσοδο για τον νευρώνα, ενώ οι τιμές του φίλτρου αποτελούν τα βάρη για τα οποία εκπαιδεύεται το νευρωνικό δίκτυο και είναι υπεύθυνο να βρει τις κατάλληλες τιμές με σκοπό την εξαγωγή των επιθυμητών χαρακτηριστικών. Με βάση τα παραπάνω μπορούμε να εκφράσουμε την έξοδο οποιουδήποτε νευρώνα ανήκει σε έναν χάρτη χαρακτηριστικών ως

$$y = \sigma \left(\sum_{i=1}^N \sum_{j=1}^M w_{ij} x_{i+p, j+q} + b \right)$$

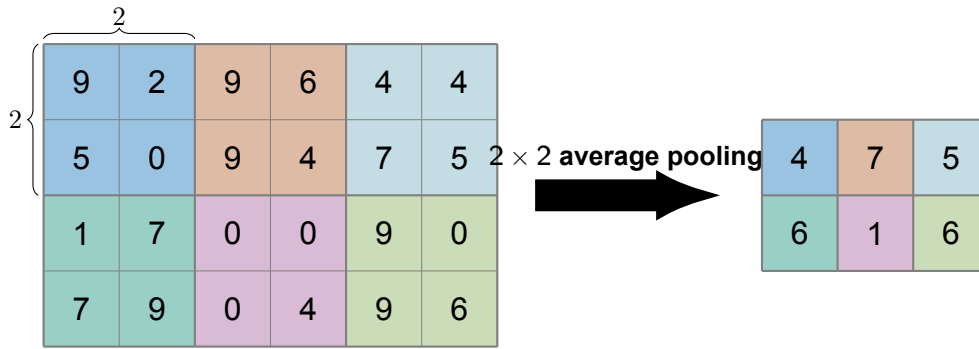
όπου σ είναι η συνάρτηση ενεργοποίησης, N και M είναι τα μεγέθη του φίλτρου σε κάθε διάσταση, w_{ij} είναι το βάρος στην αντίστοιχη θέση ij του φίλτρου, $x_{i+p, j+q}$ είναι η είσοδος του νευρώνα ο οποίος βρίσκεται στη θέση (p, q) του χάρτη χαρακτηριστικών και b είναι η πόλωση του νευρώνα. Συνεπώς, παράλληλα με τη συγκράτηση της χωρικής πληροφορίας, η συγκεκριμένη πράξη μειώνει σταδιακά το μέγεθος των παραμέτρων του νευρωνικού δικτύου καθιστώντας ευκολότερη την εκπαίδευσή του. Όπως είπαμε, οι τιμές των στοιχείων του φίλτρου είναι αυτές που καθορίζουν τα χαρακτηριστικά που θα εξαχθούν από το δεδομένο στο οποίο θα εφαρμοστεί το φίλτρο. Άρα, αλλάζοντας αυτές τις τιμές μπορούμε στοχευμένα να εξάγουμε το χαρακτηριστικό που μας ενδιαφέρει. Για παράδειγμα, το Λαπλασιανό φίλτρο (Laplacian filter) εφαρμοσμένο σε μία εικόνα

λειτουργεί ως ανιχνευτής ακμών (edge detector), ενώ το φίλτρο Sharpen αυξάνει την οξύτητα μιας εικόνας. Ωστόσο, για να μπορέσει το νευρωνικό δίκτυο να ανταπεξέλθει σε σύνθετα προβλήματα πρέπει να μπορεί να εξάγει πολλά χαρακτηριστικά. Το γεγονός αυτό καθιστά τη χρήση πολλών διαφορετικών φίλτρων μονόδρομο. Αυτή την ανάγκη εξυπηρετεί η ίδια η αρχιτεκτονική η οποία επιτρέπει την ύπαρξη πολλών χαρτών χαρακτηριστικών. Ένα σύνολο από χάρτες χαρακτηριστικών απαρτίζει ένα νέο στρώμα το οποίο είναι μέρος του κρυφού στρώματος και ονομάζεται συνελικτικό στρώμα (convolutional layer). Πέρα από τη δυνατότητα ύπαρξης πολλών χαρτών χαρακτηριστικών μέσα σε ένα συνελικτικό στρώμα, υπάρχει και η δυνατότητα ύπαρξης πολλών συνελικτικών στρωμάτων. Ο λόγος που δίνεται αυτή η δυνατότητα εξαρτάται από την πράξη της συγκέντρωσης και θα εξηγηθεί παρακάτω. Ένα συνελικτικό στρώμα δέχεται ως είσοδο χάρτες χαρακτηριστικών, ή το δεδομένο εισόδου αν είναι το πρώτο συνελικτικό στρώμα, και παράγει άλλους χάρτες χαρακτηριστικών ως έξοδο. Ο αριθμός των χαρτών που παράγει ισούται με τον αριθμό των φίλτρων που χρησιμοποιούνται στο συγκεκριμένο στρώμα. Σε πολλές περιπτώσεις, όπως όπως ισχύει και στο παράδειγμα της Εικόνας 2.17, τα φίλτρα παρουσιάζονται ως δισδιάστατοι πίνακες, ωστόσο τυπικά τα φίλτρα είναι τριών διαστάσεων και συγκεκριμένα $w \times h \times d$, όπου w, h, d είναι το πλάτος, ύψος και βάθος αντίστοιχα. Το βάθος των φίλτρων σε ένα συνελικτικό στρώμα είναι ίσο με τον αριθμό των χαρτών χαρακτηριστικών του προηγούμενου στρώματος.

Η πράξη της συγκέντρωσης εφαρμόζεται σε πίνακα τον οποίο χωρίζει σε υποπίνακες όπου από τον κάθε ένα εξάγει ένα στοιχείο με βάση κάποιο κριτήριο. Τα στοιχεία που επιλέγονται, συνθέτουν τον τελικό πίνακα ο οποίος είναι μικρότερης διάστασης από τον αρχικό. Σε παρόμοια λογική με τη συνέλιξη, το τελικό αποτέλεσμα επηρεάζεται από το μέγεθος των υποπινάκων και από το βήμα. Οι συχνότερες μέθοδοι εξαγωγής στοιχείων από τους υποπίνακες είναι δύο: η μέση συγκέντρωση (average pooling) η οποία εξάγει ως νέο στοιχείο τον μέσο όρο όλων των στοιχείων του υποπίνακα στον οποίο εφαρμόζεται και η μέγιστη συγκέντρωση (max pooling) η οποία εξάγει το στοιχείο με τη μέγιστη τιμή. Η Εικόνα 2.18 και η Εικόνα 2.19 απεικονίζουν ένα παράδειγμα του τρόπου λειτουργίας της μέσης και μέγιστης συγκέντρωσης αντίστοιχα.



Εικόνα 2.18: Παράδειγμα εφαρμογής της μέγιστης συγκέντρωσης.



Εικόνα 2.19: Παράδειγμα εφαρμογής της μέσης συγκέντρωσης.

Στην αρχιτεκτονική του συνελκτικού νευρωνικού δικτύου, η συγκέντρωση εφαρμόζεται στους χάρτες χαρακτηριστικών των συνελκτικών στρωμάτων. Μάλιστα, το αποτέλεσμα της εφαρμογής της συγκέντρωσης σε κάθε χάρτη χαρακτηριστικών ενός συνελκτικού στρώματος, απαρτίζει ένα νέο στρώμα το οποίο, όπως και το συνελκτικό στρώμα, είναι μέρος του κρυφού στρώματος, και ονομάζεται στρώμα συγκέντρωσης. Οι σημαντικότεροι λόγοι ύπαρξης των στρωμάτων συγκέντρωσης είναι δύο. Ο πρώτος λόγος είναι η αμεταβλητότητα που προσφέρουν στο νευρωνικό δίκτυο για δεδομένα που τα εικονοστοιχεία τους έχουν μετατοπιστεί προς μία κατεύθυνση (translational invariance). Με άλλα λόγια, με την ύπαρξη στρωμάτων συγκέντρωσης το νευρωνικό δίκτυο μπορεί να αναγνωρίσει εάν δύο ή παραπάνω εικόνες απεικονίζουν ουσιαστικά το ίδιο αντικείμενο ανεξάρτητα με το αν αυτό είναι μετατοπισμένο προς μία διαφορετική κατεύθυνση κάθε φορά. Η Εικόνα 2.20 δείχνει δύο πιθανές εισόδους για ένα νευρωνικό δίκτυο, οι οποίες απεικονίζουν το ίδιο αντικείμενο, ένα χειρόγραφο πέντε μετατοπισμένο αριστερά και δεξιά αντίστοιχα.



Εικόνα 2.20: Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο μετατοπισμένο σε διαφορετικές κατευθύνσεις.

Ιδανικά, θέλουμε το νευρωνικό δίκτυο να παρουσιάζει αμεταβλητότητα σε δεδομένα που παρουσιάζουν το ίδιο αντικείμενο με διαφορετική περιστροφή (rotational invariance), όπως για παράδειγμα απεικονίζεται στην Εικόνα 2.21, και σε δεδομένα που παρουσιάζουν το ίδιο αντικείμενο με διαφορετικό μέγεθος (size invariance), όπως για παράδειγμα απεικονίζονται στην Εικόνα 2.22. Εντούτοις, η πράξη της συνέλιξης δεν είναι ικανή να του προσδώσει αυτές τις ιδιότητες, συνεπώς για τον συγκεκριμένο σκοπό χρησιμοποιούνται διαφορετικές τεχνικές, όπως αυτή της επαύξησης δεδομένων (data augmentation).



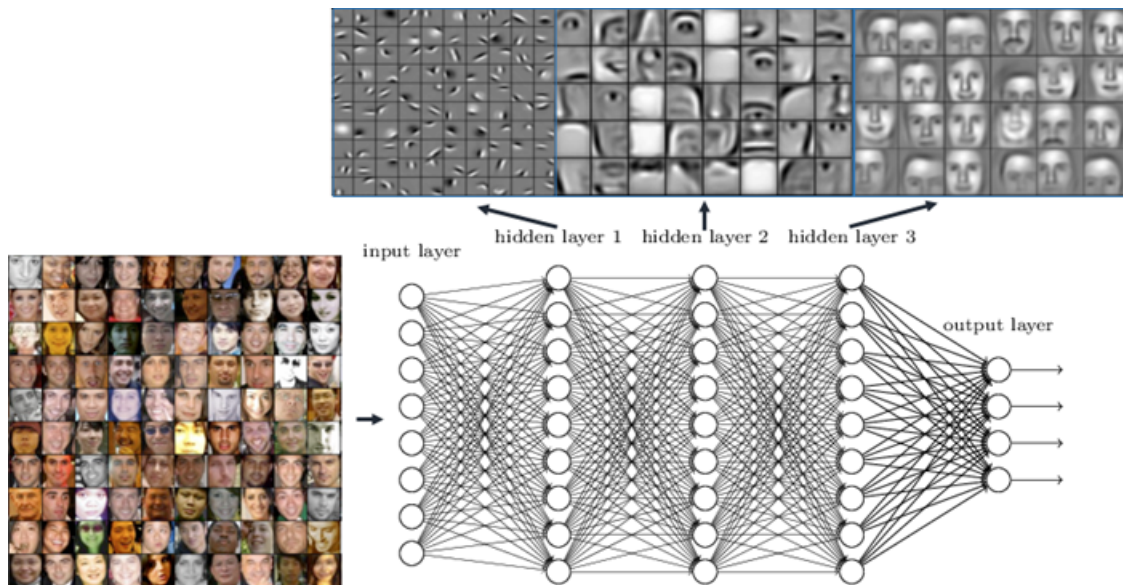
Εικόνα 2.21: Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο με διαφορετική περιστροφή.



Εικόνα 2.22: Παράδειγμα εισόδων που απεικονίζουν το ίδιο αντικείμενο με διαφορετικό μέγεθος.

Ο δεύτερος και σημαντικότερος λόγος είναι η μείωση της διάστασης των δεδομένων (dimensionality reduction), η οποία οδηγεί σε ταχύτερη εκμάθηση και αποφυγή της υπερπροσαρμογής καθώς μειώνει σημαντικά τον αριθμό των χαρακτηριστικών προς εκμάθηση. Συγκεκριμένα, η διάσταση των δεδομένων μειώνεται στο $\frac{1}{s^2}$ της αρχικής, όταν $s > 1$, όπου s είναι το βήμα.

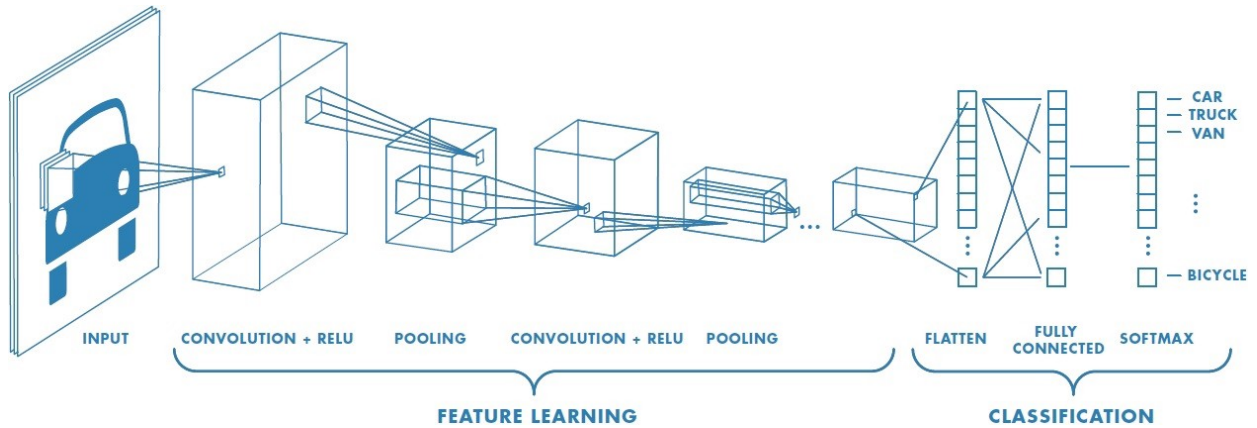
Ο συνδυασμός των συνελκτικών στρωμάτων και των στρωμάτων συγκέντρωσης αποτελεί το πρώτο μέρος ενός συνελκτικού νευρωνικού δικτύου το οποίο είναι γνωστό ως εκμάθηση χαρακτηριστικών (feature learning). Το είδος των χαρακτηριστικών που εξάγονται από μία είσοδο διαφέρει ανάλογα με τη θέση κάθε συνελκτικού στρώματος στη συνολική αρχιτεκτονική. Συγκεκριμένα, τα πρώτα στρώματα εξάγουν χαμηλότερου επιπέδου χαρακτηριστικά, για παράδειγμα διάφορες γραμμές οι οποίες συντελούν κάποια σχήματα, ενώ τα στρώματα που βρίσκονται πιο βαθιά εξάγουν υψηλότερου επιπέδου χαρακτηριστικά, για παράδειγμα πρόσωπα και αντικείμενα. Η Εικόνα 2.23 παρουσιάζει ένα παράδειγμα αυτής της λειτουργίας.



Εικόνα 2.23: Παράδειγμα του τρόπου εξαγωγής χαρακτηριστικών ανά συνελκτικό στρώμα.

Το δεύτερο μέρος εκμεταλλεύεται τα χαρακτηριστικά που έχουν εξαχθεί για να επιτευχθεί ο αρχικός στόχος, ο οποίος μπορεί να είναι κατηγοριοποίηση, εντοπισμός αντικειμένου(object detection), σημασιολογική κατάτμηση (semantic segmentation) ή κάτι άλλο παρεμφερές. Ειδικότερα, η έξοδος του τελευταίου στρώματος, είτε αυτό είναι συνελκτικό είτε συγκέντρωσης, μετατρέπεται σε έναν μονοδιάστατο πίνακα ο οποίος δίνεται ως είσοδος σε ένα πλήρως συνδεδεμένο στρώμα (fully connected layer, dense layer). Το πλήρως συνδεδεμένο στρώμα δεν είναι κάτι άλλο παρά ένα εμπροσθοτροφοδοτούμενο νευρωνικό δίκτυο (feedforward neural network) όπου κάθε νευρώνας που ανήκει σε αυτό δέχεται ως είσοδο όλες τις εξόδους του προηγούμενου

στρώματος. Η Εικόνα 2.24 παρουσιάζει ένα παράδειγμα της αρχιτεκτονικής συνελικτικού νευρωνικού δικτύου. Στο συγκεκριμένο παράδειγμα, κάθε συνελικτικό στρώμα ακολουθείται από ένα στρώμα συγκέντρωσης, κάτι που αποτελεί μία αρκετά συχνή προσέγγιση, ωστόσο ο κάθε ένας είναι ελεύθερος να στήσει το νευρωνικό δίκτυο διαφορετικά, ανάλογα με τις ανάγκες του. Το ποια δομή ανταπεξέρχεται καλύτερα σε ένα πρόβλημα, πολλές φορές δεν μπορεί να είναι γνωστό εκ των προτέρων και ο μόνος τρόπος για να μαθευτεί είναι οι διαρκείς πειραματισμοί.



Εικόνα 2.24: Παράδειγμα αρχιτεκτονικής συνελικτικού νευρωνικού δικτύου.

3. ΣΥΝΑΦΕΙΣ ΕΡΓΑΣΙΕΣ

3.1 Μέθοδοι βασισμένες στην κατωφλίωση (Thresholding)

Η κατωφλίωση αποτελεί μία δημοφιλή και ταυτόχρονα απλή τεχνική κατάτμησης εικόνων με αποχρώσεις του γκρι (grayscale) κατά την οποία η κατάτμηση της εικόνας επιτυγχάνεται συγκρίνοντας την τιμή κάθε εικονοστοιχείου της εικόνας, γνωστό και ως ένταση (intensity), με ένα κατώφλι έντασης. Συνολικά για μία εικόνα, ορίζονται είτε ένα καθολικό κατώφλι είτε περισσότερα τοπικά κατώφλια όπου το κάθε ένα αντιστοιχεί σε μία συγκεκριμένη περιοχή της εικόνας. Όποια ένταση είναι μικρότερη από το κατώφλι που της αντιστοιχεί μηδενίζεται, παράγοντας ένα μαύρο εικονοστοιχείο, διαφορετικά λαμβάνει την τιμή ένα παράγοντας ένα άσπρο εικονοστοιχείο. Το τελικό αποτέλεσμα είναι μία δυαδική εικόνα διαιρεμένη σε δύο τμήματα: στο προσκήνιο (foreground) και στο φόντο (background). Η κατωφλίωση καθαυτή διαιρεί την τελική εικόνα σε ακριβώς σε δύο τμήματα, ωστόσο για την διαίρεση σε περισσότερα τμήματα η τεχνική ονομάζεται πολυεπίπεδη κατάτμηση. Ανά τα χρόνια, έχουν καθιερωθεί διάφορες μέθοδοι, οι οποίες έχουν αυτοματοποιήσει πλήρως τη διαδικασία επιλογής κατωφλίου. Κρίνοντας από ποικίλες δημοσιεύσεις, η χρήση της κατωφλίωσης μεμονωμένα με σκοπό τον εντοπισμό όγκου στον εγκέφαλο, δεν αποτελεί αξιόπιστη επιλογή καθώς, στην πληθώρα αυτών, η κατωφλίωση αποτελεί αρχικό στάδιο ακολουθούμενο από μία σειρά διαφόρων βημάτων επεξεργασίας εικόνας, στη διαδικασία της κατάτμησης. Ένα από τα καθιερωμένα είδη μεθόδων κατωφλίωσης αποτελεί η εύρεση καθολικού κατωφλίου με τη χρήση ιστογραμμάτων τα οποία απεικονίζουν τις εντάσεις των εικόνων. Το επιστημονικό άρθρο [33] πραγματεύεται τη διαδικασία εντοπισμού όγκου στον εγκέφαλο από εικόνα μαγνητικής τομογραφίας, με τη χρήση ιστογραμμάτων. Βέβαια, στο συγκεκριμένο άρθρο η κατωφλίωση χρησιμοποιείται ως αρχικό και μη μοναδικό στάδιο επεξεργασίας της εικόνας. Παρόμοια δουλειά παρατηρείται και στο άρθρο [26], με την κύρια διαφορά να ανάγεται στο γεγονός ότι κάθε εικόνα χωρίζεται σε ισομεγέθη τμήματα, υπολογίζοντας το ιστόγραμμα για κάθε ένα από αυτά τα τμήματα. Ωστόσο, η τελική κατάτμηση των εικόνων δεν κρίνεται ικανοποιητική. Αξιοσημείωτο είναι το άρθρο [17] στο οποίο χρησιμοποιείται ένα καθολικό κατώφλι και η τιμή του προέρχεται από τον υπολογισμό του μέσου όρου όλων των μοναδικών μη μηδενικών εντάσεων, πετυχαίνοντας ικανοποιητικά αποτελέσματα. Μία άλλη μέθοδος, της οποίας η χρησιμότητα αναδεικνύεται στο [6] είναι η μέθοδος του Otsu. Τέλος, μία παραλλαγή της μεθόδου του Otsu προτείνεται στο άρθρο [13], η οποία επιτρέπει την πολυεπίπεδη κατάτμηση εικόνων. Γενικά, η τεχνική της κατωφλίωσης δεν αποτελεί επαρκής για την αξιόπιστη κατάτμηση εικόνων. Το κυριότερο μειονέκτημά της είναι ότι λαμβάνει υπόψη αποκλειστικά την ένταση των εικόνων, αγνοώντας άλλες σημαντικές παραμέτρους, όπως αυτή της σχέσης μεταξύ των εικονοστοιχείων. Με άλλα λόγια, στην παραγωγή της δυαδικής εικόνας πιθανό σενάριο αποτελεί η λανθασμένη συμπερίληψη μερών της αρχικής εικόνας, που είναι εντελώς ανεξάρτητα με τον όγκο, φυσικά ως άσπρα εικονοστοιχεία πλέον, οδηγώντας σε λανθασμένη διάγνωση.

3.2 Μέθοδοι βασισμένες σε περιοχές της εικόνας (Region-based)

Την αδυναμία εκμετάλλευσης της πιθανής ομογένειας μεταξύ γειτονικών εικονοστοιχείων των εικονοστοιχείων που διακατέχουν τις μεθόδους βασισμένες στην κατωφλίωση, διορθώνουν οι μέθοδοι βασισμένες σε περιοχές της εικόνας. Συγκεκριμένα, αυτές οι μέθοδοι διαιρούν την εικόνα σε περιοχές και η διαίρεση γίνεται με βάση κριτήρια

ομοιογένειας πάνω σε γειτονικά εικονοστοιχεία. Η διαίρεση γίνεται με τέτοιο τρόπο ώστε κάθε εικονοστοιχείο να ανήκει μόνο σε μία περιοχή, τα κριτήρια ομοιογένειας να πληρούνται για κάθε εικονοστοιχείο που ανήκει σε μία περιοχή, ενώ αν δύο περιοχές της εικόνας ενωθούν σε μία, τότε τα κριτήρια ομοιογένειας θα παύουν να πληρούνται για όλα τα εικονοστοιχεία. Μία εξαιρετικά δημοφιλής μέθοδος είναι η αυτή της ανάπτυξης περιοχών (region growing). Η συγκεκριμένη μέθοδος ξεκινά με ένα η περισσότερα εικονοστοιχεία της εικόνας ως αρχικά σημεία (seeds) όπου το κάθε ένα σηματοδοτεί μία περιοχή της εικόνας. Στη συνέχεια για κάθε περιοχή ελέγχει την ομοιογένεια των γειτονικών εικονοστοιχείων με βάση τα κριτήρια που έχουν οριστεί και το κάθε εικονοστοιχείο γίνεται πλέον κτήμα της περιοχής της οποίας πληροί τα κριτήρια, αναπτύσσοντας με αυτόν τον τρόπο τις περιοχές. Η προαναφερθείσα διαδικασία επαναλαμβάνεται έως ότου δεν υπάρχει κανένα εικονοστοιχείο που να μπορεί να προστεθεί σε κάποια περιοχή. Στο άρθρο [10] υλοποιείται η μέθοδος ανάπτυξης περιοχών, με τα αρχικά σημεία να διαλέγονται μέσω ενός αλγορίθμου που βασίζεται στην πυκνότητα των εικονοστοιχείων και κάθε γειτονικό εικονοστοιχείο μιας περιοχής να εισέρχεται σε αυτή μόνο αν η διαφορά, σε απόλυτη τιμή, της έντασης του εικονοστοιχείου με κάθε άλλο εικονοστοιχείο που ανήκει σε αυτήν την περιοχή είναι μικρότερη από μία προκαθορισμένη σταθερά. Αν δεν ικανοποιείται η συγκεκριμένη συνθήκη, τότε το γειτονικό εικονοστοιχείο εντάσσεται σε μία καινούργια περιοχή. Η διαδικασία συνεχίζεται μέχρις ότου όλα τα εικονοστοιχεία να έχουν ενταχθεί σε μία περιοχή. Στο άρθρο [32] προτείνεται μία τροποποιημένη εκδοχή η οποία επιλέγει ορθότερα τα εικονοστοιχεία που αποτελούν τα όρια το όγκου κατά το εντοπισμό του και της οποίας η υπερίσχυση σε σχέση με την κλασική μέθοδο αναλύεται στο άρθρο [31]. Στο άρθρο [19] υλοποιείται μία ημι-αυτόματη εκδοχή της μεθόδου ανάπτυξης περιοχών, όπου δίνεται η δυνατότητα στον χρήστη βλέποντας την εικόνα να διαλέξει αυτός τα αρχικά σημεία. Βέβαια, αυτή τη φορά η συγκεκριμένη μέθοδος αποτελεί το τελευταίο βήμα από μία σειρά μεθόδων επεξεργασίας της εικόνας.

3.3 Μέθοδοι βασισμένες στη στατιστική (Statistics-based)

Η χρήση μεθόδων βασισμένων στη στατιστική αποτελεί μία απλούστερη στην υλοποίηση προσέγγιση η οποία είναι αρκετά διαδεδομένη για τον εντοπισμό όγκου στον εγκέφαλο. Δύο εξαιρετικά σημαντικές μέθοδοι είναι τα Μαρκοβιανά Τυχαία Πεδία (Markov Random Fields) και το Γκαουσιανό Μοντέλο Ανάμειξης (Gaussian Mixture Model).

3.3.1 Μαρκοβιανά Τυχαία Πεδία (Markov Random Fields)

Τα Μαρκοβιανά Τυχαία Πεδία εφαρμοσμένα σε μία εικόνα, αντιμετωπίζουν τα εικονοστοιχεία αυτής ως τυχαίες μεταβλητές που πληρούν τη Μαρκοβιανή ιδιότητα (Markov property) και εκφράζονται συνολικά από ένα μη κατευθυνόμενο γράφο του οποίου οι ακμές συνδέουν τα γειτονικά εικονοστοιχεία. Η χρησιμότητα αυτής της μεθόδου ανάγεται στην ικανότητά της να αναγνωρίζει τις ομοιότητες που διακρίνουν τα γειτονικά εικονοστοιχεία και να συμπεριφέρεται ικανοποιητικά σε δεδομένα με θόρυβο. Στο άρθρο [16] περιγράφεται η χρήση της συγκεκριμένης μεθόδου για την κατάτμηση εικόνων εγκεφάλου λαμβάνοντας υπόψη χαρακτηριστικά όπως τη μη-παραμετρική κατανομή των εντάσεων που αφορούν το κομμάτι του εγκεφάλου, συσχετίσεις γειτονιών (neighbourhood correlations) και ανομοιογένειες σήματος (signal inhomogeneities). Τα δεδομένα που χρησιμοποιήθηκαν αφορούσαν υγιής εγκεφάλους, συνεπώς δεν υπήρχε εντοπισμός κάποιου όγκου, ωστόσο η εικόνα κατατμήθηκε με γνώμονα άλλα μορφολογικά

χαρακτηριστικά του εγκεφάλου, αναδεικνύοντας με αυτόν τον τρόπο τις δυνατότητες αυτής της μεθόδου. Αντίθετα, πάνω σε μη υγιείς εγκεφάλους δοκιμάστηκε η προσέγγιση που προτείνεται στο άρθρο [35] η οποία εφαρμόζει τα Μαρκοβιανά Τυχαία Πεδία μόνο σε εικόνες εγκεφάλων που έχουν κατηγοριοποιηθεί ως μη υγιή από έναν κατηγοριοποιητή μηχανικής μάθησης σε προηγούμενο στάδιο, αξιοποιώντας τα πλεονεκτήματα που προσφέρουν τα Μαρκοβιανά Τυχαία Πεδία με σκοπό τον καλύτερο δυνατό διαχωρισμό του όγκου από το υπόλοιπο υγιές μέρος του εγκεφάλου. Παρόμοια δουλειά περιγράφεται και στο άρθρο [36], με τη διαφορά ότι για κάθε εικόνα εφαρμόζεται επαναληπτικά η μέθοδος των Μαρκοβιανών Τυχαίων Πεδίων σε δύο διαφορετικά επίπεδα τα οποία αλληλεπιδρούν μεταξύ τους, έως ότου υπάρξει ικανοποιητική κατάτμηση.

3.3.2 Γκαουσιανό μοντέλο ανάμειξης (Gaussian mixture model)

Το Γκαουσιανό Μοντέλο Ανάμειξης είναι ένα πιθανοτικό μοντέλο που σκοπεύει στο διαχωρισμό των δεδομένων σε συστάδες. Αυτό το επιτυγχάνει κάνοντας την υπόθεση ότι όλα τα δεδομένα παράγονται από ένα σύνολο Γκαουσιανών κατανομών με άγνωστες παραμέτρους. Ο αριθμός των κατανομών του συνόλου αποτελεί και τον αριθμό των συστάδων στις οποίες ανατίθενται τα δεδομένα. Σε αναλογία με το πρόβλημά μας, τα δεδομένα προς συσταδοποίηση είναι τα εικονοστοιχεία της εικόνας και ο σκοπός είναι η ανάθεση όλων των εικονοστοιχείων που απεικονίζουν τον όγκο στον εγκέφαλο σε μία κοινή συστάδα. Στο άρθρο [21] χρησιμοποιείται το Γκαουσιανό Μοντέλο Ανάμειξης ως το ενδιάμεσο στάδιο κατά την κατηγοριοποίηση εγκεφάλων, έχοντας προηγηθεί μία πολυεπίπεδη κατωφλίωση, ενώ η τελική απόφαση για το αν υπάρχει όγκος σε έναν εγκέφαλο ή όχι λαμβάνεται από μία μέθοδο μηχανικής μάθησης. Παρόμοια δουλειά παρουσιάζεται και στο άρθρο [7]. Ως αρχικό στάδιο σε μία επεξεργασία εικόνας τριών σταδίων χρησιμοποιείται στο άρθρο [39] το Γκαουσιανό Μοντέλο Ανάμειξης, διαχωρίζοντας τα εικονοστοιχεία σε 5 διαφορετικές συστάδες. Επιπροσθέτως, μία παραλλαγή της μεθόδου η οποία παρουσιάζει ικανοποιητικά αποτελέσματα διατυπώνεται στο άρθρο. Συχνά το Γκαουσιανό Μοντέλο Ανάμειξης προτιμάται να συνδυάζεται με άλλες μεθόδους στην αντιμετώπιση του προβλήματός μας, δίνοντας τη αίσθηση ότι δεν είναι ικανή να ανταπεξέλθει αυτούσια.

3.4 Μέθοδοι βασισμένες στη βαθιά μάθηση (Deep learning-based)

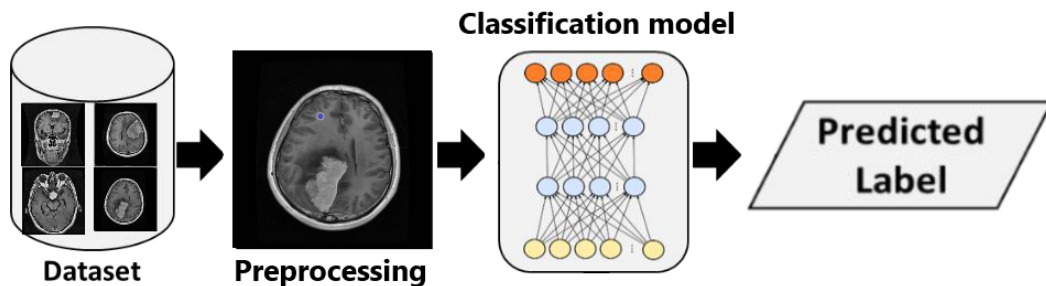
Κατά κύριο λόγο η βαθιά μάθηση, σε συναφή προβλήματα, χρησιμοποιείται για την κατάτμηση και κατηγοριοποίηση των όγκων εγκεφάλου. Στο άρθρο [27] επιτυγχάνεται κατάτμηση του όγκου εφαρμόζοντας κατηγοριοποίηση στα εικονοστοιχεία της εικόνας. Συγκεκριμένα, χρησιμοποιούνται τέσσερα, ίδιας αρχιτεκτονικής, συνελκτικά νευρωνικά δίκτυα όπου το κάθε ένα δέχεται μία ελαφρώς διαφορετική εικόνα του ίδιου εγκεφάλου. Στη συνέχεια κάθε συνελκτικό νευρωνικό δίκτυο κατηγοριοποιεί κάθε εικονοστοιχείο σε μία από πέντε πιθανές κλάσεις που αφορούν χαρακτηριστικά των εγκεφαλικών όγκων. Στο τέλος ενώνονται τα τελευταία κρυφά στρώματα των τεσσάρων δικτύων και χρησιμοποιούνται ως εξαγωγείς χαρακτηριστικών τροφοδοτούνται σε έναν κατηγοριοποιητή μηχανικής μάθησης ονόματι Τυχαία Δάση (Random Forests) που είναι αυτός ο οποίος, εν τέλει, κατηγοριοποιεί τα εικονοστοιχεία. Άλλη μία μέθοδος που συνδυάζει κατάτμηση και κατηγοριοποίηση προτείνεται στο άρθρο [3] στην οποία χρησιμοποιείται ένα ημι-αυτόματο λογισμικό κατάτμησης εγκεφαλικών όγκων [4] και στη συνέχεια εφαρμόζεται κατηγοριοποίηση, με τη χρήση ενός συνελκτικού νευρωνικού

δικτύου, στην κατάσταση του μοριακού χαρακτηριστικού 1p/19q το οποίο αποτελεί ένδειξη για τον σωστότερο τρόπο αντιμετώπισης του όγκου στο μέλλον. Ενδιαφέρον έχει η μέθοδος του άρθρου [38] η οποία αφορά επεξεργάζεται τρισδιάστατες εικόνες εγκεφάλων. Ειδικότερα, εφαρμόζεται κατάτμηση των όγκων με τη χρήση της αρχιτεκτονικής M-Net [25] και τα αποτελέσματά της τροφοδοτούνται σε ένα ήδη εκπαιδευμένο μοντέλο VGG-16 με σκοπό την κατηγοριοποίηση τριών διαφορετικών τύπων όγκου. Στο άρθρο [18] προτείνεται μια τροποποιημένη εκδοχή της αρχιτεκτονικής U-Net [29] με σκοπό τη μεγιστοποίηση της απόδοσής της σε προβλήματα κατάτμησης εγκεφαλικών όγκων. Τέλος, στο άρθρο [5] εφαρμόζεται κατάτμηση των όγκων μέσω κατωφλίωσης και στη συνέχεια τους κατηγοριοποιεί ως καλοήθεις ή κακοήθεις, συνδυάζοντας τις εξόδους των δύο προ-εκπαιδευμένων μοντέλων, GoogleNet [37] και AlexNet [20], τα οποία δέχθηκαν ως είσοδο τους κατατμημένους όγκους. Τέλος, στο άρθρο [2] προτείνεται ένα LSTM μοντέλο με σχετικά λίγες παραμέτρους για τον εντοπισμό όγκων εγκεφάλου.

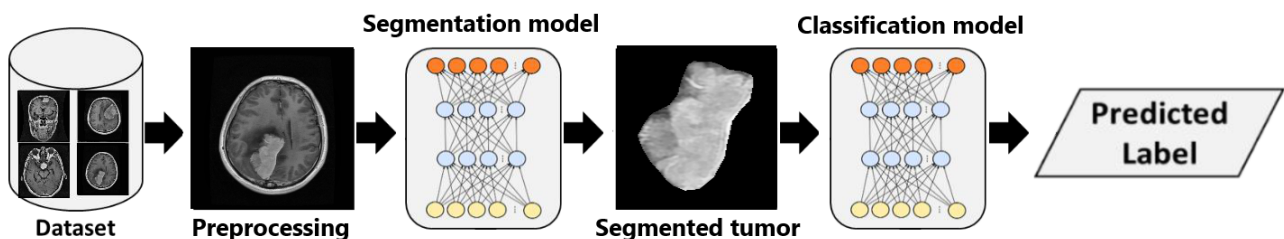
4. ΠΡΟΤΕΙΝΟΜΕΝΗ ΠΡΟΣΕΓΓΙΣΗ

4.1 Περιγραφή προβλήματος (Problem description)

Στόχος αυτής της εργασίας είναι η προσομοίωση της διαδικασίας εξαγωγής ιατρικών αποτελεσμάτων που ακολουθεί ένας ειδικός από τη στιγμή που αρχίζει να εξετάζει τις εικόνες που απεικονίζουν τον εγκέφαλο ενός ασθενή και έχουν παραχθεί από την ακτινολογική μέθοδο της μαγνητικής τομογραφίας, χρησιμοποιώντας τεχνικές βαθιάς μάθησης. Η διαδικασία αυτή μπορεί να αναχθεί σε δύο κύρια βήματα: εντοπισμός των περιοχών που υπάρχει όγκος και κατηγοριοποίηση του όγκου. Στο συγκεκριμένο πρόβλημα ο εντοπισμός και η κατηγοριοποίηση εφαρμόζονται σε τρεις διαφορετικούς τύπους όγκου: μηνιγγίωμα (meningioma), γλοίωμα (glioma) και όγκος υπόφυσης (pituitary). Για την εκπλήρωση αυτού του στόχου και με γνώμονα πάντα τη σωστή κατηγοριοποίηση των όγκων, ακολουθούνται δύο διαφορετικές προσεγγίσεις. Στην πρώτη προσέγγιση δεν υλοποιείται κάποιο μοντέλο που να εκτελεί μεμονωμένα το βήμα του εντοπισμού. Αντιθέτως, η εξαγωγή χαρακτηριστικών και η κατηγοριοποίηση των όγκων εκτελείται από ένα μοντέλο. Στη δεύτερη προσέγγιση χρησιμοποιούνται δύο μοντέλα, με το πρώτο να εκτελεί κατάτμηση του όγκου και το δεύτερο να κατηγοριοποιεί τον κατατμημένο όγκο. Και στις δύο περιπτώσεις τα δεδομένα υπόκεινται σε προεπεξεργασία προτού αποτελέσουν είσοδοι των μοντέλων. Η Εικόνα 4.1 και η Εικόνα 4.2 απεικονίζουν διαγραμματικά αυτές τις δύο προσεγγίσεις αντίστοιχα.



Εικόνα 4.1: Διάγραμμα της πρώτης προσέγγισης.

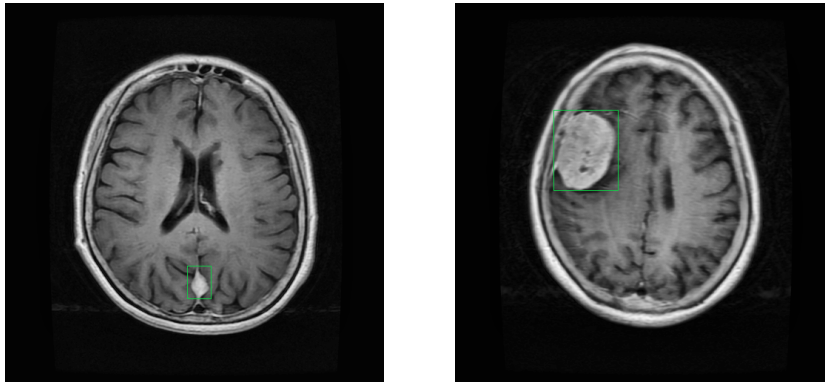


Εικόνα 4.2: Διάγραμμα της δεύτερης προσέγγισης.

4.2 Προκλήσεις Προβλήματος

- **Μορφολογική αβεβαιότητα:** Η τοποθεσία που βρίσκεται ένας όγκος στο χώρο του εγκεφάλου και τα μορφολογικά χαρακτηριστικά αυτού, όπως είναι το

μέγεθος και το σχήμα του, ενδέχεται να διαφέρουν σημαντικά από ασθενή σε ασθενή, επιβαρύνοντας την, έτσι κι αλλιώς, εξαιρετικά απαιτητική δουλειά της κατηγοριοποίησης των όγκων στον εγκέφαλο. Για παράδειγμα, η Εικόνα 4.3 απεικονίζει δύο διαφορετικούς εγκεφάλους στους οποίους παρατηρείται ύπαρξη όγκου. Παρόλο που το σχήμα, το μέγεθος και η τοποθεσία των δύο όγκων δε συμπίπτουν καθόλου, το είδος στο οποίο ανήκουν είναι το ίδιο.



Εικόνα 4.3: Παράδειγμα δύο ίδιου είδους όγκων με διαφορετικά χαρακτηριστικά.

- **Ανεπάρκεια δεδομένων:** Η ανεπάρκεια των δεδομένων είναι ένα φαινόμενο που παρατηρείται συχνά σε ιατρικά δεδομένα και οφείλεται, κυρίως, στο ιατρικό απόρρητο. Συνήθως, από τα σετ δεδομένων που διατίθενται ελεύθερα προς το ευρύ κοινό, αυτά που παρέχουν ετικέτες (labels) είναι μικρά στον αριθμό, ενώ τα σετ με επαρκή αριθμό δεδομένων δεν παρέχουν ετικέτες. Στην πρώτη περίπτωση, εγκυμονεί ο κίνδυνος ανεπαρκούς εκπαίδευσης του νευρωνικού δικτύου, ενώ στην δεύτερη περίπτωση για να μπορέσει να εκπαιδευτεί ένα νευρωνικό δίκτυο, απαιτείται η χειροκίνητη ερμηνεία των δεδομένων, γεγονός που απαιτεί χρόνο και ιατρικές γνώσεις.
- **Ποιότητα δεδομένων:** Η ποιότητα των φωτογραφιών είναι συχνά χαμηλής ποιότητας καθώς πιθανώς να έχουν θόρυβο και χαμηλή αντίθεση επιβαρύνοντας τη διαδικασία εντοπισμού των χωρικών ορίων του όγκου.
- **Μεροληψία στα δεδομένα:** Η τελική μορφή της παραγόμενης φωτογραφίας από την μαγνητική τομογραφία, εξαρτάται από πολλούς παράγοντες, όπως είναι το ίδιο το μηχάνημα που εκτελεί την μαγνητική τομογραφία αλλά και η μέθοδος που επιλέγεται. Ενδεικτικά, δύο γνωστές μέθοδοι μαγνητικής τομογραφίας είναι η T1-σταθμισμένη και η T2-σταθμισμένη. Αυτοί οι παράγοντες ενδέχεται να διαφέρουν από κλινική σε κλινική, επιφέροντας με αυτόν τον τρόπο μεροληψία σε τεχνικά χαρακτηριστικά της φωτογραφίας, για παράδειγμα στην ένταση.
- **Υπολογιστικό κόστος:** Κατά γενική ομολογία, το υπολογιστικό κόστος των νευρωνικών δικτύων είναι ακριβό και οι βαθύτερες αρχιτεκτονικές αυτών, με εκατομμύρια εκπαιδευσιμα βάρη, όπως αυτές που εφαρμόστηκαν σε αυτό το πρόβλημα, απαιτούν υλικό (hardware) υψηλών προδιαγραφών για να μπορέσει να γίνει η εκπαίδευση σε εύλογο χρονικό διάστημα.

4.3 Μεταφορά μάθησης (Transfer learning)

Η μεταφορά μάθησης αποτελεί μία εξαιρετικά δημοφιλής τεχνική με εφαρμογές στην βαθιά μάθηση. Η ιδέα της μεταφοράς μάθησης είναι η επίλυση ενός προβλήματος χρησιμοποιώντας γνώση που έχει αποκτηθεί από την επίλυση άλλων παρεμφερών προβλημάτων. Αυτή η ιδέα εφαρμόζεται καθημερινά και στην ανθρώπινη ζωή καθώς αποτελεί καίριο συστατικό του τρόπου με τον οποίο οι άνθρωποι αποκτούν νέες ικανότητες. Για παράδειγμα, έστω ότι δίνεται σε δύο ανθρώπους το ίδιο χρονικό διάστημα για να μάθουν να παίζουν για πρώτη φορά πινγκ πονγκ, και ο πρώτος είναι επαγγελματίας τενίστας ενώ ο άλλος δεν έχει παίξει ποτέ τένις ούτε παρόμοια αθλήματα. Κατα μέσο όρο, ο πρώτος άνθρωπος, θα εκμεταλλευτεί, συνειδητά ή μη, τις γνώσεις που έχει αποκτήσει πάνω στο τένις και θα καταφέρει να φτάσει σε υψηλότερο επίπεδο από τον δεύτερο άνθρωπο, σε αυτό το δοσμένο χρονικό διάστημα. Αντιστοιχίζοντας το προηγούμενο παράδειγμα με τη βαθιά μάθηση, μπορούμε να θεωρήσουμε ότι οι δύο άνθρωποι αντιπροσωπεύουν δύο νευρωνικά δίκτυα και το χρονικό διάστημα αντιπροσωπεύει ένα σετ δεδομένων. Το πρώτο νευρωνικό δίκτυο, εκμεταλλευόμενο το γεγονός ότι στο παρελθόν έχει εκπαιδευτεί σε πολλά παρεμφερή δεδομένα με αυτά που βρίσκονται στο νέο σετ δεδομένων και έχει καταφέρει να αποκτήσει την ικανότητα της γενίκευσης θα έχει, κατά μέσο όρο, καλύτερη επίδοση από το δεύτερο νευρωνικό δίκτυο το οποίο θα αναγκαστεί να εκπαιδευτεί εξ ολοκλήρου μόνο με το καινούργιο σετ δεδομένων. Η θετική επίδραση που μπορεί να έχει η εφαρμογή της μεταφοράς μάθησης στην επίλυση ενός προβλήματος γίνεται εμφανέστερη όσο μικρότερο είναι το σετ δεδομένων του συγκεκριμένου προβλήματος. Επίσης, πολύ σημαντικό πλεονέκτημα της μεταφοράς μάθησης είναι η δυνατότητα μείωσης του απαιτούμενου υπολογιστικού κόστους. Όπως έχει αναφερθεί και παραπάνω, το σετ δεδομένων αυτού του προβλήματος δεν είναι όσο μεγάλο όσο θα ήταν ιδανικά επιθυμητό και το υπολογιστικό κόστος του συγκεκριμένου προβλήματος είναι αρκετά υψηλό. Συνεπώς, για την επίλυσή του η τεχνική της μεταφοράς μάθησης.

Συνολικά, οι κύριες στρατηγικές εφαρμογής της μεταφοράς μάθησης στη βαθιά μάθηση είναι δύο. Στην πρώτη στρατηγική, το ήδη εκπαιδευμένο νευρωνικό δίκτυο χρησιμοποιείται ως εξαγωγέας χαρακτηριστικών. Για να συμβεί αυτό, τα βάρη του προ-εκπαιδευμένου νευρωνικού δικτύου παραμένουν άθικτα και αφαιρείται κατ' ελάχιστον το στρώμα εξόδου του. Φυσικά, υπάρχει η δυνατότητα αφαίρεσης περισσότερων στρωμάτων, ξεκινώντας πάντα την αφαίρεση σειριακά από το στρώμα εξόδου στα προηγούμενα στρώματα. Συνήθως αφαιρείται είτε μόνο το στρώμα εξόδου είτε το στρώμα εξόδου και τα πλήρως συνδεδεμένα στρώματα που προηγούνται. Σε κάθε περίπτωση, το τελευταίο στρώμα από τα εναπομείναντα είναι αυτό που κατέχει όλα τα χαρακτηριστικά που εξάγονται μετά την είσοδο των δεδομένων στο ήδη εκπαιδευμένο νευρωνικό δίκτυο. Από εκεί και πέρα αυτό το στρώμα είναι εκμεταλλεύσιμο με ποικίλους τρόπους. Για παράδειγμα, το στρώμα αυτό θα μπορούσε να αποτελεί είσοδο για ένα άλλο αντίστοιχο νευρωνικό δίκτυο το οποίο εκτελεί περαιτέρω εξαγωγή χαρακτηριστικών ή απλούστερα να συνδεθεί με ένα στρώμα εξόδου με κατάλληλο αριθμό εξόδων για το σετ δεδομένων που χρησιμοποιείται, με σκοπό την κατηγοριοποίηση των εισόδων.

Στην δεύτερη στρατηγική, γνωστή και ως fine-tuning, πέρα από την αφαίρεση των στρωμάτων, γίνεται και εκπαίδευση σε ορισμένα από τα εναπομείναντα στρώματα. Όπως έχει τονιστεί στην ενότητα 2.9, τα αρχικά στρώματα μαθαίνουν γενικά χαρακτηριστικά τα οποία είναι πιο πιθανό να είναι κοινά σε πολλά διαφορετικά προβλήματα, ενώ τα πιο βαθιά στρώματα μαθαίνουν πιο συγκεκριμένα χαρακτηριστικά που αφορούν κυρίως το σετ δεδομένων του προβλήματος. Συνεπώς, η επιλογή του αριθμού στρωμάτων που θα

εκπαιδευτούν, αποτελεί μια παράμετρο που εξαρτάται τόσο από τον αριθμό όσο και από το είδος των χαρακτηριστικών που θέλουμε να εξάγει το ήδη εκπαιδευμένο νευρωνικό δίκτυο από το σετ δεδομένων. Φυσικά, όσα λιγότερα στρώματα εκπαιδευτούν, τόσο μικρότερο είναι το υπολογιστικό κόστος. Ο τρόπος με τον οποίο εφαρμόζεται η μεταφορά μάθησης στο συγκεκριμένο πρόβλημα αναλύεται στην ενότητα 4.10.

4.4 Σετ δεδομένων (Dataset)

Το σετ δεδομένων που χρησιμοποιείται [8] αποτελείται από 3064 T1-σταθμισμένες (T1-weighted) εικόνες οι οποίες, συλλογικά, απεικονίζουν τρεις διαφορετικούς είδους όγκου στον εγκέφαλο: μηνιγγίωμα, γλοίωμα και όγκος υπόφυσης. Το μηνιγγίωμα αποτελεί τον πιο συχνό πρωτοπαθή όγκο εγκεφάλου και αναπτύσσεται σε μία από τις μεμβράνες που περιβάλλουν τον εγκέφαλο, γνωστές ως μήνιγγες. Συχνά συμπτώματα του συγκεκριμένου όγκου είναι επιληψία, προβλήματα όρασης και κινητικά προβλήματα. Περίπου το 80% των μηνιγγιωμάτων είναι καλοήθης όγκοι και αφαιρούνται πλήρως με χειρουργική επέμβαση. Το γλοίωμα αποτελεί και αυτό έναν από τους συχνότερους πρωτοπαθείς όγκους εγκεφάλου ο οποίος αναπτύσσεται από τα νευρογλοιακά κύτταρα που περιβάλλουν τα νευρικά κύτταρα του εγκεφάλου και συχνά είναι κακοήθης. Περίπου το 30% όλων των εγκεφαλικών όγκων και το 80% όλων των κακοήθων όγκων είναι γλοιώματα. Τέλος, ο όγκος υπόφυσης είναι κατά κύριο λόγο καλοήθης και βρίσκεται στην περιοχή που είναι γνωστή ως τουρκικός εφίππιος ο οποίος είναι πίσω από τη μύτη. Η παρουσία ενός όγκου υπόφυσης ενδέχεται να δημιουργήσει οπτικά προβλήματα ή περίσσια παραγωγή ορμονών που επιφέρουν ποικίλα συμπτώματα με τη σειρά τους. Ωστόσο οι περισσότεροι όγκοι υπόφυσης δεν εμφανίζουν συμπτώματα.

Κάθε εικόνα, μαζί με κάποιες, δυνητικά, χρήσιμες πληροφορίες που την αφορούν, βρίσκεται αποθηκευμένη σε ένα ξεχωριστό δυαδικό αρχείο MATLAB. Αναλυτικότερα, κάθε αρχείο αυτού του σετ δεδομένων περιέχει τα δεδομένα της T1-σταθμισμένης εικόνας, τον τύπο όγκου που παρατηρείται στην εικόνα, τις συντεταγμένες που απαρτίζουν τα χωρικά όρια του όγκου, τα δεδομένα μιας δυαδικής εικόνας που αντιπροσωπεύει την μάσκα του όγκου και ένα μοναδικό αντιπροσωπευτικό αριθμό του ασθενή που ανήκει η εικόνα. Για το συγκεκριμένο πρόβλημα χρησιμοποιήθηκε ένα υποσύνολο αυτών των πληροφοριών, καθώς για την κατηγοριοποίηση χρειάστηκαν τα δεδομένα της T1-σταθμισμένης εικόνας και ο τύπος του όγκου, ενώ για την κατάτμηση χρειάστηκαν τα δεδομένα της T1-σταθμισμένης εικόνας και η μάσκα του όγκου.

Το συγκεκριμένο σετ δεδομένων είναι μη ισορροπημένο (imbalanced) καθώς ο αριθμός των εικόνων με γλοίωμα (1426 εικόνες) είναι σημαντικά μεγαλύτερος από αυτούς των εικόνων με μηνιγγίωμα (708 εικόνες) εικόνες και με όγκο υπόφυσης (930 εικόνες) καθιστώντας ακόμα πιο δύσκολη την επίτευξη της γενίκευσης του νευρωνικού δικτύου κατά την εκπαίδευσή του για το πρόβλημα της κατηγοριοποίησης. Το σετ δεδομένων διασπάστηκε τυχαία σε τρία υποσύνολα: σετ εκπαίδευσης (training set), σετ επικύρωσης (validation set) και σετ αξιολόγησης (test set). Το σετ εκπαίδευσης αποτελείται από 1964 εικόνες, δηλαδή περίπου από το 64% του αρχικού σετ δεδομένων και τα δεδομένα χρησιμοποιούνται αποκλειστικά για την εκπαίδευση του νευρωνικού δικτύου. Το σετ επικύρωσης αποτελείται από 500 εικόνες, δηλαδή περίπου από το 16% του αρχικού σετ δεδομένων και τα δεδομένα του χρησιμοποιούνται αποκλειστικά για την αξιολόγηση του νευρωνικού δικτύου, ενόσω εκπαιδεύεται και συγκεκριμένα στο πέρας κάθε εποχής. Η χρήση του σετ επικύρωσης συνεισφέρει στην αποφυγή του φαινομένου της υπερπροσαρμογής η οποία καθίσταται εξαιρετικά πιθανή όταν η αξιολόγηση κατά τη

διάρκεια της εκπαίδευσης γίνεται με τα δεδομένα που εκπαιδεύεται το νευρωνικό δίκτυο. Τέλος, το σετ αξιολόγησης αποτελείται από 600 εικόνες, δηλαδή περίπου από το 20% του αρχικού σετ δεδομένων και τα δεδομένα του χρησιμοποιούνται μετά το πέρας της εκπαίδευσης του νευρωνικού δικτύου καθώς αποσκοπεί στην αξιολόγηση της επίδοσης του σε δεδομένα τα οποία προσομοιώνουν τα δεδομένα του πραγματικού κόσμου.

4.5 Προεπεξεργασία δεδομένων (Data preprocessing)

Τα δεδομένα του πραγματικού κόσμου, κατά κανόνα, περιέχουν περιττές πληροφορίες και διακατέχονται από ιδιότητες, όπως ο θόρυβος, οι οποίες μπορούν να επηρεάσουν σημαντικά την επίδοση των νευρωνικών δικτύων.

Συχνό φαινόμενο στις μαγνητικές τομογραφίες εγκεφάλων είναι η χαμηλή αντίθεσή τους. Η χαμηλή αντίθεση προκύπτει όταν οι τιμές έντασης που ανήκουν σε ένα μεγάλο ποσοστό των εικονοστοιχείων αποτελούν ένα μικρό ποσοστό του πλήρους εύρους τιμών που μπορεί να λάβει η ένταση. Η τεχνική της Προσαρμοστικής Εξίσωσης Ιστογράμματος Περιορισμένης Αντίθεσης (Contrast Limited Adaptive Histogram Equalization – CLAHE) στοχεύει σε μία πιο ομοιόμορφη κατανομή των τιμών έντασης που λαμβάνουν τα εικονοστοιχεία της εικόνας. Αρχικά, χωρίζει την εικόνα σε μικρότερα τμήματα και για το κάθε ένα υπολογίζει το ιστόγραμμα του. Στη συνέχεια αποκόπτει τα μέρη του ιστογράμματος που ξεπερνούν ένα προκαθορισμένο όριο και μετά μετασχηματίζει κάθε τιμή του ιστογράμματος εφαρμόζοντας σε κάθε μία τη συνάρτηση

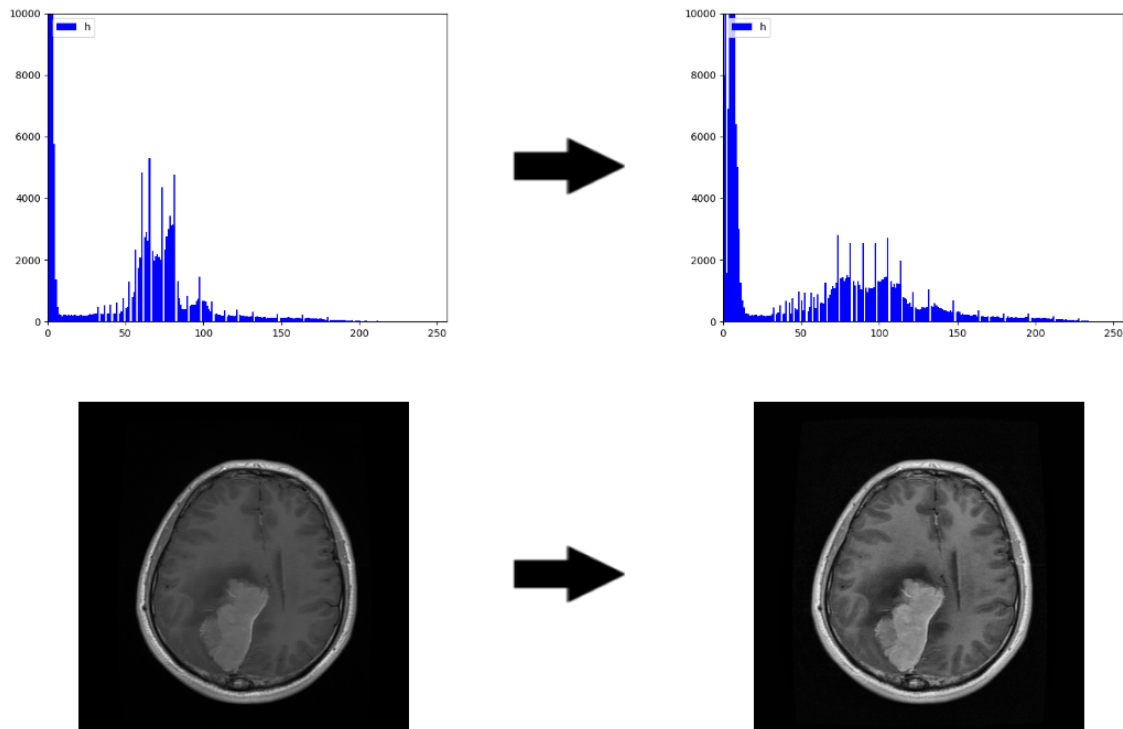
$$T(k) = (L - 1) \sum_{i=0}^k p_i$$

όπου L είναι ο αριθμός όλων των πιθανών τιμών έντασης που μπορεί να λάβει ένα εικονοστοιχείο και p_n εκφράζεται από η εξίσωση

$$p_i = \frac{n_i}{n}$$

όπου n_i είναι ο αριθμός των εικονοστοιχείων με ένταση i και n ο συνολικός αριθμός των εικονοστοιχείων.

Στο τέλος, τα εικονοστοιχεία που έχουν τις τιμές που αποκόπηκαν από το ιστόγραμμα λαμβάνουν νέα τιμή με ομοιόμορφο τρόπο. Για το συγκεκριμένο πρόβλημα η τιμή ορίου αποκοπής που επιλέχθηκε ισούται με δύο και κάθε εικόνα χωρίζεται σε περιοχές μεγέθους 8×8 . Η Εικόνα 4.4 δείχνει ένα παράδειγμα εφαρμογής αυτής της τεχνικής σε μία εικόνα του σετ δεδομένων.

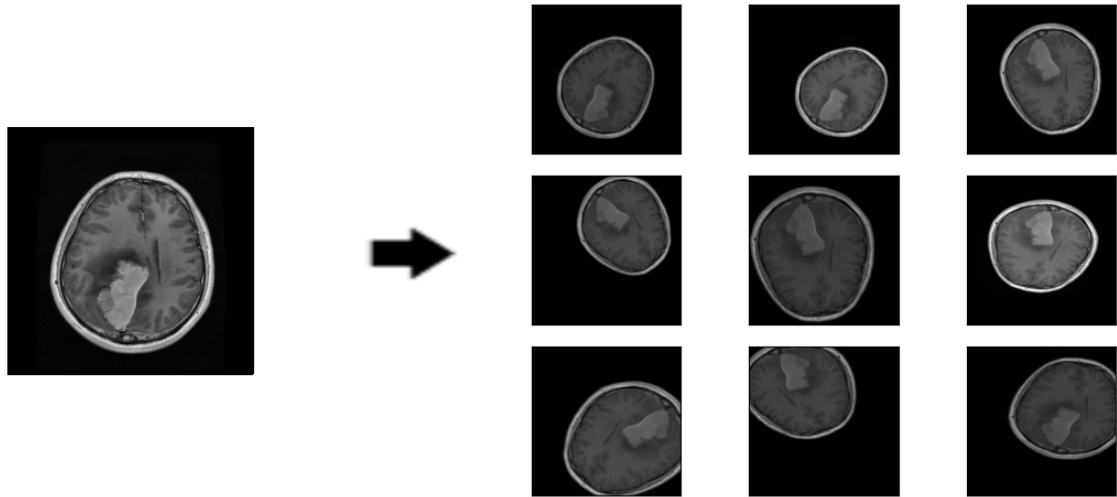


Εικόνα 4.4: Παράδειγμα εφαρμογής CLAHE.

4.6 Επαύξηση δεδομένων

Η επαύξηση δεδομένων είναι μια τεχνική η οποία αυξάνει το μέγεθος ενός σετ δεδομένων παράγοντας νέα ελαφρώς τροποποιημένα αντίγραφα των δεδομένων που το απαρτίζουν. Σπάνιες είναι οι φορές που αυτή η τεχνική δε θα προσδώσει κάποιο όφελος στην εκπαίδευση ενός νευρωνικού δικτύου, ανεξάρτητα από το μέγεθος και τη μορφή των δεδομένων. Ωστόσο, η χρησιμότητά της αναδεικνύεται ακόμα περισσότερο σε μικρά σετ δεδομένων όπου ο κίνδυνος υπερπροσαρμογής είναι υψηλός. Η συγκεκριμένη τεχνική εφαρμόζεται και στις δύο προσεγγίσεις, με μικρές διαφοροποιήσεις από την μία στην άλλη. Συγκεκριμένα, στην πρώτη προσέγγιση κάθε εικόνα του σετ εκπαίδευσης υπόκεινται σε:

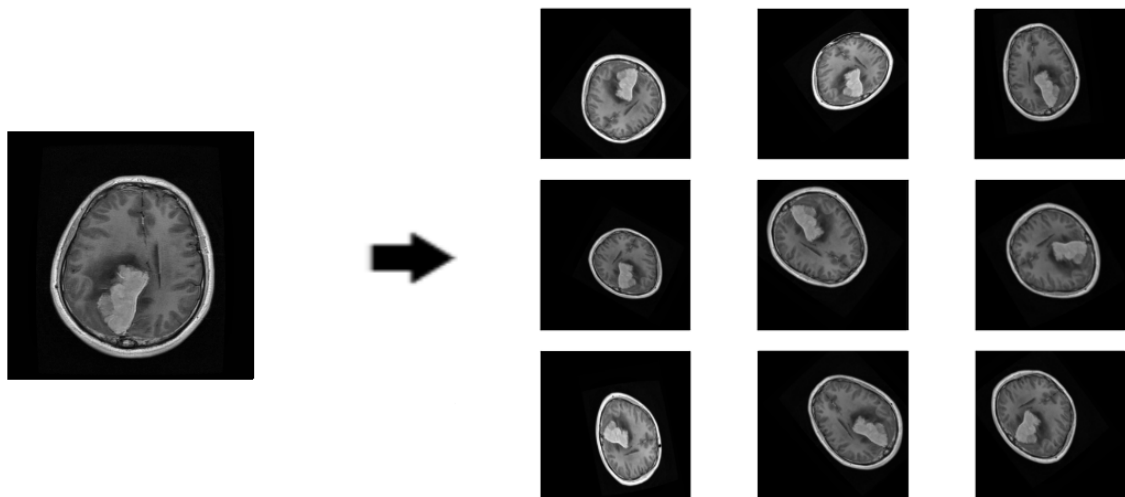
- Τυχαία περιστροφή (rotation) έως και 90 μοιρών.
- Τυχαία οριζόντια μετατόπιση (horizontal translation) κατά έως και 20% του πλάτους της.
- Τυχαία κατακόρυφη μετατόπιση (vertical translation) κατά έως και 20% του ύψους της.
- Τυχαίο οριζόντια αναστροφή (horizontal flip).
- Τυχαίο κατακόρυφη αναστροφή (vertical flip).
- Τυχαία μεγέθυνση (zoom in) ή σμικρυνση (zoom out) κατά έως και 20%.
- Τυχαία πρόσθεση ή αφαίρεση φωτεινότητας κατά έως και 50%.



Εικόνα 4.5: Παράδειγμα επαύξησης των δεδομένων στην πρώτη προσέγγιση.

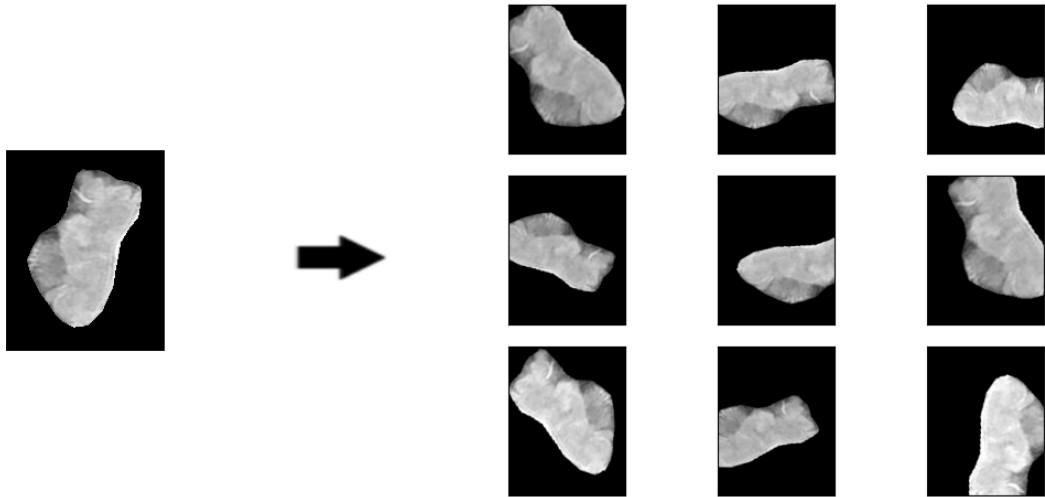
Στην δεύτερη προσέγγιση η τεχνική της επαύξησης δεδομένων εφαρμόζεται δύο φορές: μία φορά για το μοντέλο κατάτμησης και μία φορά για το μοντέλο κατηγοριοποίησης. Αναφορικά με το μοντέλο κατάτμησης, εφαρμόζονται:

- Τυχαία περιστροφή έως και 90 μοιρών
- Τυχαία μεγέθυνση ή σμίκρυνση κατά έως και 30%.
- Τυχαία οριζόντια μετατόπιση κατά έως και 15% του πλάτους της.
- Τυχαία κατακόρυφη μετατόπιση κατά έως και 15% του ύψους της.
- Τυχαίο οριζόντια αναστροφή.
- Τυχαίο κατακόρυφη αναστροφή.
- Τυχαίο ψαλίδισμα (shearing) κατά έως και 8 μοίρες.
- Τυχαία πρόσθεση φωτεινότητας κατά έως και 50% ή αφαίρεση φωτεινότητας κατά έως και 20%.
- Τυχαία εφαρμογή Γκαουσιανού φίλτρου με τυπική απόκλιση από 0 έως 5.



Εικόνα 4.6: Παράδειγμα επαύξησης των δεδομένων για το μοντέλο κατάτμησης της δεύτερης προσέγγισης.

Για το μοντέλο κατηγοριοποίησης εφαρμόζεται ό,τι εφαρμόστηκε και στο μοντέλο κατάτμησης της πρώτης προσέγγισης με τη μόνη διαφορά ότι αφαιρείται ή προστίθεται, τυχαία, έως και 20% της αρχικής φωτεινότητας αντί για 50%.



Εικόνα 4.7: Παράδειγμα επαύξησης δεδομένων για το μοντέλο κατηγοριοποίησης της δεύτερης προσέγγισης.

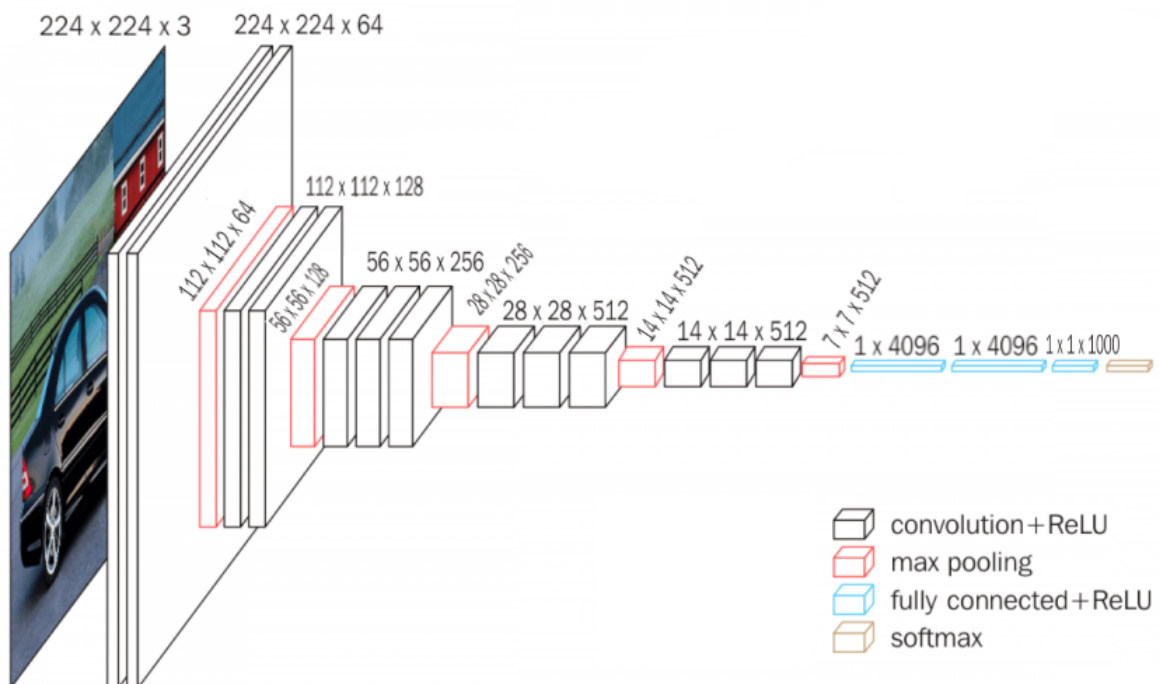
4.7 VGGNet

Η αρχιτεκτονική VGGNet [34] είναι δημιούργημα του γκρουπ οπτικής γεωμετρίας (Visual Geometry Group) του Πανεπιστημίου της Οξφόρδης και αποτελεί μια πολύ γνωστή επιλογή για προβλήματα κατηγοριοποίησης εικόνων. Κατά τη δημιουργία αυτής της αρχιτεκτονικής, αρχικός σκοπός τους ήταν να δείξουν ότι η αύξηση του βάθους ενός νευρωνικού δικτύου επιφέρει και αύξηση στις επιδόσεις του. Στο άρθρο που παρουσιάζεται η συγκεκριμένη αρχιτεκτονική, προτείνονται εννέα διαφορετικές παραμετροποιήσεις της που διαφέρουν αποκλειστικά στο βάθος, το οποίο διαδοχικά αυξάνεται από 11 έως 19. Για την συγκεκριμένη αρχιτεκτονική, ως βάθος ορίζεται ο αριθμός των στρωμάτων που έχουν εκπαιδευσιμα βάρη. Για αυτό το πρόβλημα, χρησιμοποιούνται οι δύο γνωστότερες παραμετροποιήσεις: VGG-16 και VGG-19 με βάθος 16 και 19 αντίστοιχα. Οι δύο αυτές παραμετροποιήσεις θα αποτελέσουν τα μοντέλα κατηγοριοποίησης και στα δύο προβλήματα.

Η είσοδος περνάει από 5 σετ συνελικτικών στρωμάτων με κάθε σετ να ακολουθείται από ένα στρώμα μέγιστης συγκέντρωσης. Τα πρώτα δύο σετ είναι ίδια για τις VGG-16, VGG-19 παραμετροποιήσεις καθώς κάθε σετ αποτελείται από δύο συνελικτικά στρώματα. Τα επόμενα τρία σετ συνελικτικών στρωμάτων είναι εκείνα που διαφοροποιούν τις δύο παραμετροποιήσεις, διότι στην παραμετροποίηση VGG-16 κάθε σετ αποτελείται από τρία συνελικτικά στρώματα, ενώ στην VGG-19 κάθε σετ αποτελείται από τέσσερα. Μετά τα σετ των συνελικτικών στρωμάτων, ανεξαρτήτως παραμετροποίησης, ακολουθούν τρία πλήρως συνδεδεμένα στρώματα και στο τέλος το στρώμα εξόδου με συνάρτηση ενεργοποίησης Softmax. Αξιοσημείωτο χαρακτηριστικό της συγκεκριμένης αρχιτεκτονικής είναι το ασυνήθιστα μικρό μέγεθος φίλτρου που χρησιμοποιείται για την πράξη της συνέλιξης, το οποίο επιλέχθηκε με γνώμονα την μείωση των εκπαιδευσιμων βαρών, που αναπόφευκτα λόγω βάθους είναι πολλά.

Πίνακας 4.1: Λεπτομέρειες της παραμετροποίησης VGG-16.

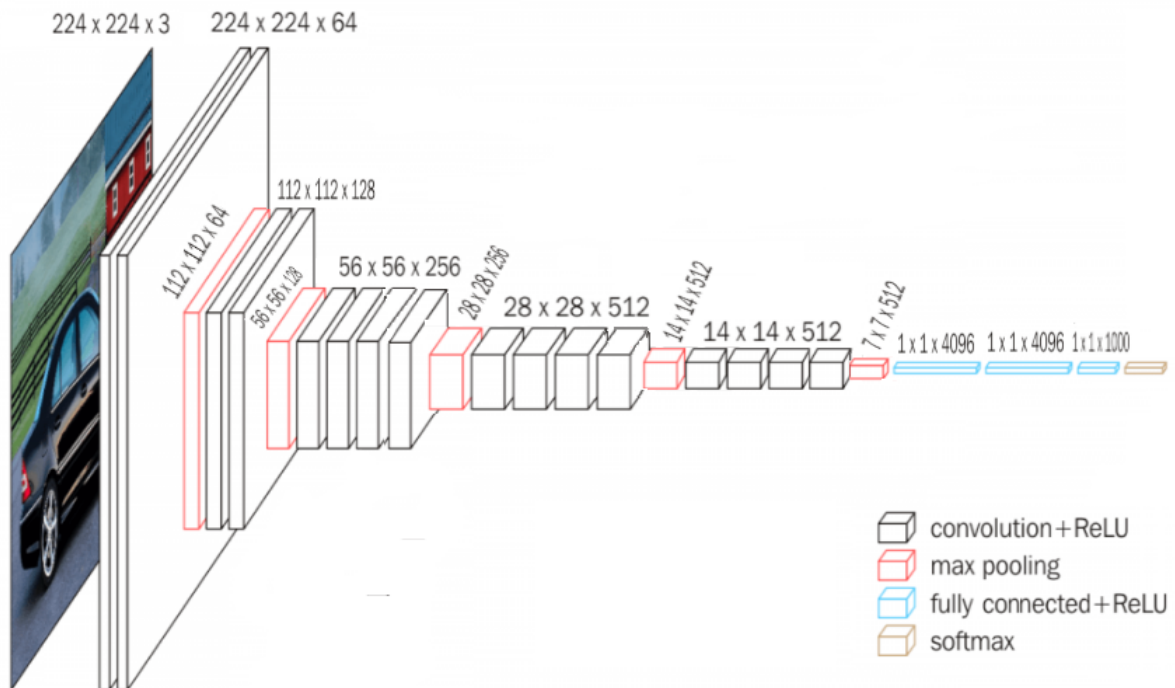
| Βάθος | Στρώμα | Χάρτες χαρακτηριστικών | Μέγεθος | Μέγεθος φίλτρου | Βήμα | Συνάρτηση ενεργοποίησης |
|-------|-----------------------|------------------------|-----------------------------|-----------------|------|-------------------------|
| - | Εισόδου | 1 | $224 \times 224 \times 3$ | - | - | - |
| 1 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Μέγιστης συγκέντρωσης | 64 | $112 \times 112 \times 64$ | 2×2 | 2 | - |
| 3 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 4 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 4 | Μέγιστης συγκέντρωσης | 128 | $56 \times 56 \times 128$ | 2×2 | 2 | - |
| 4 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 5 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Μέγιστης συγκέντρωσης | 256 | $28 \times 28 \times 256$ | 2×2 | 2 | - |
| 7 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 8 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 9 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 9 | Μέγιστης συγκέντρωσης | 512 | $14 \times 14 \times 512$ | 2×2 | 2 | - |
| 10 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 11 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 12 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 12 | Μέγιστης συγκέντρωσης | 512 | $7 \times 7 \times 512$ | 2×2 | 2 | - |
| 14 | Πλήρως συνδεδεμένο | - | 4096 | - | - | ReLU |
| 15 | Πλήρως συνδεδεμένο | - | 4096 | - | - | ReLU |
| 16 | Εξόδου | - | 1000 | - | - | Softmax |



Εικόνα 4.8: Σχηματική απεικόνιση της παραμετροποίησης VGG-16.

Πίνακας 4.2: Λεπτομέρειες της παραμετροποίησης VGG-19.

| Βάθος | Στρώμα | Χάρτες χαρακτηριστικών | Μέγεθος | Μέγεθος φίλτρου | Βήμα | Συνάρτηση ενεργοποίησης |
|-------|-----------------------|------------------------|-----------------------------|-----------------|------|-------------------------|
| - | Εισόδου | 1 | $224 \times 224 \times 3$ | - | - | - |
| 1 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Μέγιστης συγκέντρωσης | 64 | $112 \times 112 \times 64$ | 2×2 | 2 | - |
| 3 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 3 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 3 | Μέγιστης συγκέντρωσης | 128 | $56 \times 56 \times 128$ | 2×2 | 2 | - |
| 4 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 5 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Μέγιστης συγκέντρωσης | 256 | $28 \times 28 \times 256$ | 2×2 | 2 | - |
| 7 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 8 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 9 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 512 | $28 \times 28 \times 256$ | 3×3 | 1 | ReLU |
| 9 | Μέγιστης συγκέντρωσης | 512 | $14 \times 14 \times 512$ | 2×2 | 2 | - |
| 10 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 11 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 12 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 512 | $14 \times 14 \times 256$ | 3×3 | 1 | ReLU |
| 12 | Μέγιστης συγκέντρωσης | 512 | $7 \times 7 \times 512$ | 2×2 | 2 | - |
| 14 | Πλήρως συνδεδεμένο | - | 4096 | - | - | ReLU |
| 15 | Πλήρως συνδεδεμένο | - | 4096 | - | - | ReLU |
| 16 | Εξόδου | - | 1000 | - | - | Softmax |

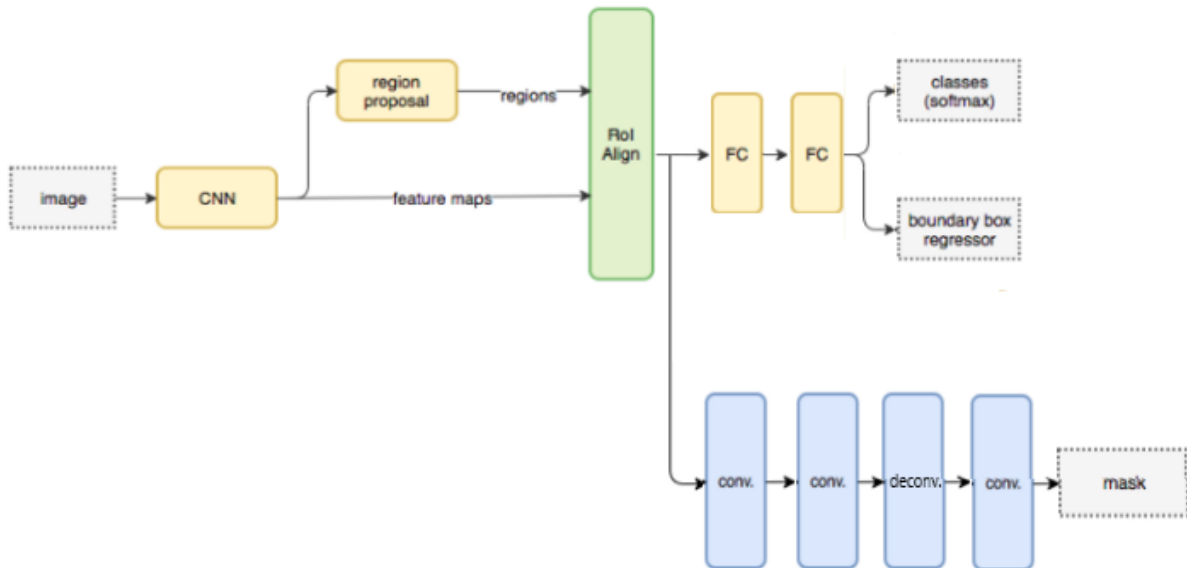


Εικόνα 4.9: Σχηματική απεικόνιση της παραμετροποίησης VGG-19.

4.8 Mask R-CNN

Η αρχιτεκτονική Mask R-CNN [14] είναι αφοσιωμένη στην επίλυση του προβλήματος της κατάτμησης στιγμιότυπου (instance segmentation) και αποτελεί επέκταση της αρχιτεκτονικής Faster R-CNN [28]. Η αρχιτεκτονική Faster R-CNN αντίστοιχα αποτελεί επέκταση της αρχιτεκτονικής Fast R-CNN [11] η οποία με τη σειρά της επεκτείνει την αρχιτεκτονική R-CNN [12].

Η συγκεκριμένη αρχιτεκτονική δέχεται ως είσοδο εικόνες, ανιχνεύει υποψήφια αντικείμενα σε αυτές και για κάθε υποψήφιο αντικείμενο παράγει τρεις εξόδους: την κατηγορία στην οποία ανήκει, το οριοθετικό πλαίσιο (bounding box) του και την μάσκα του. Η παραγωγή της μάσκας αποτελεί το κυριότερο πλεονέκτημα της αρχιτεκτονικής Mask R-CNN σε σχέση με την αρχιτεκτονική Faster R-CNN καθώς η τελευταία παράγει μόνο τις δύο πρώτες εξόδους, όπως αυτές αναφέρονται. Η παραγόμενη μάσκα είναι και ο λόγος επιλογής τη συγκεκριμένης αρχιτεκτονικής για αυτό το πρόβλημα, καθώς σε σχέση με το οριοθετικό πλαίσιο προσφέρει ακριβέστερη κατάτμηση των όγκων. Η Εικόνα 4.10 απεικονίζει τα κυριότερα βήματα της αρχιτεκτονικής Mask R-CNN, τα οποία ανήκουν συνολικά σε δύο στάδια.

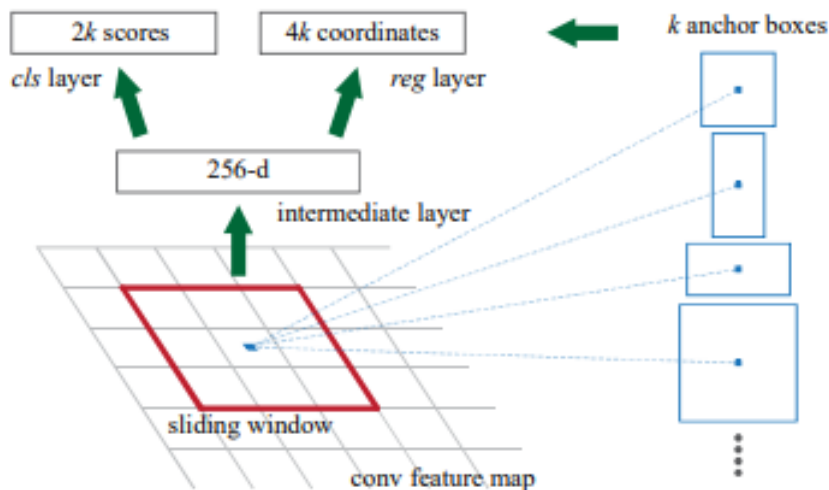


Εικόνα 4.10: Διάγραμμα κύριων βημάτων αρχιτεκτονικής Mask R-CNN.

Το πρώτο στάδιο δημιουργεί τα οριοθετικά πλαίσια, όπου το κάθε πλαίσιο εμπεριέχει ένα υποψήφιο αντικείμενο της εικόνας. Ουσιαστικά, επιδεικνύει στο δεύτερο στάδιο, το οποίο θα αναλυθεί παρακάτω, τις περιοχές τις οποίες πρέπει να επεξεργαστεί, μειώνοντας με αυτόν τον τρόπο το υπολογιστικό κόστος. Συγκεκριμένα, η εικόνα αρχικά εισέρχεται σε ένα συνελκτικό νευρωνικό δίκτυο το οποίο λειτουργεί ως εξαγωγέας χαρακτηριστικών. Για το συγκεκριμένο πρόβλημα, αυτό το συνελκτικό νευρωνικό δίκτυο είναι βασισμένο στην αρχιτεκτονική FPN [22].

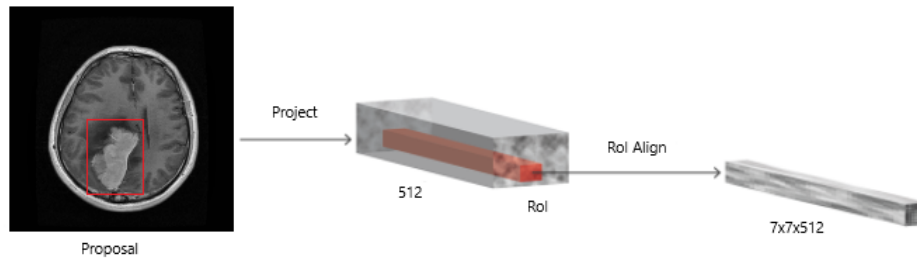
Στη συνέχεια, η έξοδος του τελευταίου συνελκτικού στρώματος η οποία είναι ένα σύνολο από χάρτες χαρακτηριστικών, αποτελεί είσοδο για ένα δίκτυο γνωστό ως Δίκτυο Πρότασης Περιοχών (Region Proposal Network – RPN) το οποίο είναι αυτό που, εν τέλει, καθορίζει τις περιοχές που θα επεξεργαστούν από το δεύτερο στάδιο. Όπως απεικονίζεται στην Εικόνα 4.11, το Δίκτυο Πρότασης Περιοχών χρησιμοποιεί ένα κυλιόμενο παράθυρο (sliding window) το οποίο σταδιακά μεταφέρεται σε όλα τα σημεία του χάρτη χαρακτηριστικών. Σε κάθε νέα θέση του κυλιόμενου παραθύρου παράγονται k νέα οριοθετικά πλαίσια, με το σχήμα κάθε πλαισίου να είναι διαφορετικό από τα υπόλοιπα $k - 1$ πλαίσια, αλλά σταθερό ανά θέση κυλιόμενου παραθύρου. Αυτά τα πλαίσια ονομάζονται άγκυρες (anchors) και το σχήμα τους είναι ένα ορθογώνιο με μοναδικό συνδυασμό κλίμακας (scale) και λόγου διαστάσεων (aspect ratio). Η ποικιλία στο σχήμα των πλαισίων επιτρέπει

στον εντοπισμό διαφορετικών αντικειμένων. Για έναν χάρτη χαρακτηριστικών μεγέθους $W \times H$ παράγονται $W \times H \times k$ άγκυρες συνολικά. Ως προεπιλογή χρησιμοποιούνται τρεις κλίμακες και τρεις λόγοι διαστάσεων που δημιουργούν εννέα διαφορετικές άγκυρες σε κάθε θέση του κυλιόμενου παραθύρου. Κατόπιν, στο ενδιάμεσο στρώμα (intermediate layer), εφαρμόζεται συνέλιξη με 512 φίλτρα και μέγεθος φίλτρου 3×3 σε κάθε k -άδα που παράχθηκε σε κάθε σημείο του κυλιόμενου παραθύρου. Το αποτέλεσμα αυτής της πράξης τροφοδοτείται σε δύο παράλληλα, πλήρως συνδεδεμένα στρώματα, όπου το ένα εκτελεί κατηγοριοποίηση και το άλλο παλινδρόμηση. Το στρώμα παλινδρόμησης παράγει τέσσερις εξόδους για κάθε άγκυρα που αναπαριστούν το ύψος, το πλάτος και τις κεντρικές συντεταγμένες της, δηλαδή συνολικά παράγει $4 \times k$ εξόδους. Αυτό το πετυχαίνει εφαρμόζοντας στην είσοδό του συνέλιξη με $4 \times k$ φίλτρα με μέγεθος φίλτρου 1×1 . Το στρώμα της κατηγοριοποίησης παράγει δύο εξόδους οι οποίες εκφράζουν την πιθανότητα ένα αντικείμενο να βρίσκεται ή να μην βρίσκεται, μέσα σε μία άγκυρα. Συνεπώς, συνολικά παράγει $2k$ εξόδους. Αντίστοιχα, αυτό το πετυχαίνει εφαρμόζοντας στην είσοδό του συνέλιξη με $4 \times k$ φίλτρα με μέγεθος φίλτρου 1×1 και συνάρτηση ενεργοποίησης Softmax.



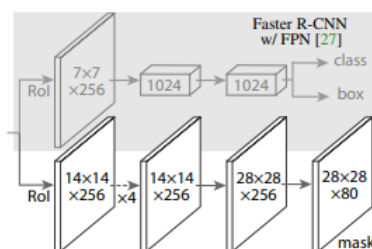
Εικόνα 4.11: Αρχιτεκτονική Δικτύου Πρότασης Περιοχών.

Το πέρας της παραπάνω διαδικασίας αποτελεί και το τέλος του πρώτου σταδίου της αρχιτεκτονικής Mask R-CNN. Όπως είδαμε, το πρώτο στάδιο προτείνει περιοχές της εικόνας σε μορφή ορθογωνίων που, εν δυνάμει, περιέχουν αντικείμενα που μας ενδιαφέρουν, χωρίς όμως να κατηγοριοποιεί τα ίδια τα αντικείμενα. Το δεύτερο βήμα, λοιπόν, έχει τρεις στόχους: να κατηγοριοποιήσει το περιεχόμενο των προτεινόμενων περιοχών (κατηγοριοποίηση ως φόντο σε περίπτωση απουσίας αντικειμένου στην προτεινόμενη περιοχή), να βελτιώσει την τοποθεσία των οριοθετικών πλαισίων που παρήχθησαν από το Δίκτυο Πρότασης Περιοχών με σκοπό την ακριβέστερο διαχωρισμό κάθε αντικειμένου και τέλος να παράξει τη μάσκα κάθε αντικειμένου. Οι περιοχές περικλυόμενες από οριοθετικά πλαίσια είναι γνωστές και ως περιοχές ενδιαφέροντος (regions of interest – RoI). Το δεύτερο στάδιο ξεκινάει εξάγοντας έναν μικρό και σταθερού μεγέθους ($7 \times 7 \times 512$) χάρτη χαρακτηριστικών από κάθε περιοχή ενδιαφέροντος. Αυτή τη φορά η εξαγωγή χαρακτηριστικών δε γίνεται με τη χρήση νευρωνικού δικτύου αλλά με μία πράξη που ονομάζεται RoIAlign.



Εικόνα 4.12: Απεικόνιση εφαρμογής RoI Align, σε υψηλό επίπεδο.

Στη συνέχεια, ο χάρτης χαρακτηριστικών ακολουθεί δύο διαφορετικά μονοπάτια, τα οποία εν τέλει παράγουν τρεις διαφορετικές εξόδους. Στο πρώτο μονοπάτι ο ίδιος διοχετεύεται σε δύο διαδοχικά πλήρως συνδεδεμένα στρώματα με το αποτέλεσμα του δεύτερου να τροφοδοτείται σε δύο παράλληλα πλήρως συνδεδεμένα στρώματα παρόμοια με αυτά που είδαμε στο Δίκτυο Πρότασης Περιοχών. Το ένα παράλληλο πλήρως συνδεδεμένο στρώμα, κατηγοριοποιεί το είδος του αντικειμένου που βρίσκεται σε κάθε περιοχή ενδιαφέροντος. Η κατηγοριοποίηση γίνεται μέσω της συνάρτησης ενεργοποίησης Softmax η οποία παράγει $K + 1$ πιθανότητες που αντιστοιχούν σε K είδη αντικειμένων και στο φόντο στην περίπτωση που η περιοχή ενδιαφέροντος δεν περιέχει κανένα αντικείμενο που ανήκει σε κάποιο από τα K είδη. Το δεύτερο στρώμα, βελτιώνει την τοποθεσία των οριοθετικών πλαισίων εκτελώντας παλινδρόμηση και παράγοντας ανανεωμένες τις τέσσερις τιμές που τα αναπαριστούν, όπως αυτές παρουσιάζονται στο Δίκτυο Πρότασης Περιοχών. Στο δεύτερο μονοπάτι για κάθε αντικείμενο των περιοχών ενδιαφέροντος, παράγεται η αντίστοιχη μάσκα του η οποία, για το συγκεκριμένο πρόβλημα, είναι χρήσιμη για τη δημιουργία εικόνων με απομονωμένους όγκους, οι οποίες αποτελούν την είσοδο των μοντέλων κατηγοριοποίησης στην δεύτερη προσέγγιση. Για την παραγωγή της μάσκας ο χάρτης χαρακτηριστικών αρχικά εισέρχεται σε τέσσερα διαδοχικά συνελκτικά στρώματα με μέγεθος φίλτρου 3×3 , στη συνέχεια εισέρχεται ένα στρώμα αποσυνέλιξης (deconvolution) με μέγεθος φίλτρου 2×2 και βήμα 2, το οποίο το οποίο αυξάνει τις χωρικές διαστάσεις και στο τέλος εισέρχεται σε ένα συνελκτικό στρώμα με μέγεθος φίλτρου 1×1 .



Εικόνα 4.13: Λεπτομερέστερη αναπαράσταση των τελικών βημάτων παραγωγής των τριών εξόδων της αρχιτεκτονικής Mask R-CNN.

Γενικά, ο αριθμός των συνελκτικών στρωμάτων και οι παράμετροι τους, που χρησιμοποιούνται για την εξαγωγή της μάσκας δεν είναι σταθεροί και εξαρτώνται από το ποιο συνελκτικό νευρωνικό δίκτυο έχει χρησιμοποιηθεί ως εξαγωγέας χαρακτηριστικών στο πρώτο στάδιο της αρχιτεκτονικής Mask R-CNN. Συνολικά υπάρχουν ποικίλες παραμετροποιήσεις του μοντέλου Mask R-CNN που διαφοροποιούν τόσο τον τρόπο με τον οποίο εξάγεται η μάσκα όσο και τα προηγούμενα στάδια που περιγράφηκαν

παραπάνω. Για το συγκεκριμένο πρόβλημα έχουν γίνει οι επιλογές που περιγράφηκαν παραπάνω και είναι βασισμένες στο [1], όπου υπάρχουν και περισσότερες λεπτομέρειες.

4.9 Μετρικές (Metrics)

Εξαιρετικά σημαντικό στάδιο κατά την ανάπτυξη ενός νευρωνικού δικτύου αποτελεί η αξιολόγηση της επίδοσής του σε δεδομένα που δεν έχει συναντήσει ποτέ στο παρελθόν. Μία τέτοια αξιολόγηση αποτελεί αξιόπιστη ένδειξη για την αντίστοιχη επίδοση του νευρωνικού δικτύου στα δεδομένα του πραγματικού κόσμου και πραγματοποιείται με τη χρήση μετρικών.

4.9.1 Μετρικές μοντέλων κατηγοριοποίησης

Πολλές μετρικές γενικά, καθώς και οι μετρικές που χρησιμοποιούνται για την αξιολόγηση των μοντέλων κατηγοριοποίησης, βασίζονται στον πίνακα σύγχυσης (confusion matrix). Ο πίνακας σύγχυσης είναι ένας πίνακας μεγέθους $N \times N$, όπου N είναι ο αριθμός των κλάσεων που ανήκουν συνολικά τα δεδομένα, ο οποίος συνδυάζει τις πραγματικές τιμές των δεδομένων με τις προβλεπόμενες. Ένας πίνακας σύγχυσης απαρτίζεται από τέσσερις έννοιες που εκφράζουν το είδος της πρόβλεψης:

- **Αληθώς Θετική (True Positive – TP):** Η πραγματική τιμή του δεδομένου και η προβλεπόμενη τιμή είναι θετικές.
- **Αληθώς Αρνητική (True Negative – TN):** Η πραγματική τιμή του δεδομένου και η προβλεπόμενη τιμή είναι αρνητικές.
- **Ψευδώς Θετική (False Positive – FP):** Η πραγματική τιμή του δεδομένου είναι αρνητική, ενώ η προβλεπόμενη τιμή είναι θετική.
- **Ψευδώς Αρνητική (False Negative – FN):** Η πραγματική τιμή του δεδομένου είναι θετική, ενώ η προβλεπόμενη τιμή είναι αρνητική.

Οι παραπάνω έννοιες παραπέμπουν σε δυαδική κατηγοριοποίηση, ωστόσο οι ίδιες μπορούν να γενικευτούν και για την περίπτωση κατηγοριοποίησης πολλαπλών κλάσεων. Η γενική περίπτωση προϋποθέτει τον υπολογισμό των τεσσάρων ειδών πρόβλεψης για κάθε κλάση. Συνεπώς για μία συγκεκριμένη κλάση, έστω A , οι τέσσερις αυτές έννοιες μπορούν να εκφραστούν ως εξής:

- **Αληθώς Θετική:** Το δεδομένο ανήκει στην κλάση A και η πρόβλεψη το ανέθεσε στην κλάση A .
- **Αληθώς Αρνητική:** Το δεδομένο ανήκει σε μια άλλη κλάση και η πρόβλεψη το ανέθεσε στην κλάση που ανήκει.
- **Ψευδώς Θετική:** Το δεδομένο ανήκει σε μια άλλη κλάση και η πρόβλεψη το ανέθεσε στην κλάση A .
- **Ψευδώς Αρνητική:** Το δεδομένο ανήκει στην κλάση A και η πρόβλεψη το ανέθεσε σε άλλη κλάση.

Για παράδειγμα έστω η κλάση B της Εικόνας 4.14. Ο αριθμός των αληθώς θετικών προβλέψεων, δηλαδή όσα δεδομένα ανατέθηκαν σωστά στην κλάση B, είναι 10. Ο αριθμός των αληθώς αρνητικών προβλέψεων είναι ίσος με το άθροισμα των αληθώς θετικών προβλέψεων των άλλων κλάσεων, δηλαδή οκτώ σωστές προβλέψεις για την κλάση A και έξι για την κλάση C. Ο αριθμός των ψευδώς θετικών προβλέψεων είναι ίσος με το άθροισμα των δεδομένων που ανήκουν στις άλλες δύο κλάσεις αλλά η πρόβλεψη τους ανέθεσε στην κλάση B, δηλαδή κανένα δεδομένο της κλάσης C και τρία δεδομένα της κλάσης A. Τέλος, οι ψευδώς αρνητικές προβλέψεις ισούνται με τα δεδομένα που ανήκουν στην κλάση B αλλά ανατέθηκαν στις κλάσεις A και C, δηλαδή κανένα δεδομένο στην κλάση A και δύο δεδομένα στην κλάση C. Με τον ίδιο τρόπο υπολογίζεται ο αριθμός κάθε είδους πρόβλεψης και για τις υπόλοιπες κλάσεις.

| | | Predicted values | | | Total |
|---------------|---------|------------------|---------|---------|-------|
| | | Class A | Class B | Class C | |
| Actual values | Class A | 8 | 3 | 1 | 12 |
| | Class B | 0 | 10 | 2 | 12 |
| | Class C | 0 | 0 | 6 | 6 |
| | Total | 8 | 13 | 9 | 30 |

Εικόνα 4.14: Αναπαράσταση ενός πίνακα σύγχυσης.

Οι μετρικές προκύπτουν συνδυάζοντας τα τέσσερα είδη πρόβλεψης κάθε κλάσης. Οι τρεις κυριότεροι τρόποι συνδυασμού των τιμών που προκύπτουν για τον υπολογισμό μιας μετρικής είναι οι εξής:

- **Μέσος όρος με προτεραιότητα στα δεδομένα (Micro average):** Πρώτα υπολογίζεται η μέση τιμή κάθε είδους πρόβλεψης λαμβάνοντας υπόψη όλες τις κλάσεις και στη συνέχεια υπολογίζεται η μετρική. Με αυτόν τον τρόπο κάθε πρόβλεψη έχει την ίδια συνεισφορά στην τελική τιμή μιας μετρικής, γεγονός που εγκυμονεί τον κίνδυνο εμφάνισης παραπλανητικών τιμών στις μετρικές όταν ένα σετ δεδομένων είναι μη ισορροπημένο.
- **Σταθμισμένος μέσος όρος (Weighted average):** Υπολογίζεται η μετρική ξεχωριστά για κάθε κλάση και στη συνέχεια κάθε μία τοπική μετρική πολλαπλασιάζεται με το αντίστοιχο ποσοστό που καταλαμβάνουν τα δεδομένα της κλάσης που ανήκει σε σχέση με όλο το σετ δεδομένων. Η τελική τιμή της μετρικής προκύπτει αθροίζοντας τις τιμές που έχουν προκύψει για κάθε κλάση. Με αυτόν τον τρόπο όσο περισσότερα δεδομένα ανήκουν σε μία κλάση, τόσο μεγαλύτερη επιρροή στην τελική τιμή της μετρικής έχει αυτή η κλάση.
- **Μέσος όρος με προτεραιότητα στις κλάσεις (Macro average):** Υπολογίζεται η μετρική ξεχωριστά για κάθε κλάση και στη συνέχεια υπολογίζεται ο μέσος όρος όλων των τοπικών μετρικών που έχουν προκύψει. Με αυτόν τον τρόπο όλες οι κλάσεις συνεισφέρουν ισάξια στην τελική τιμή της μετρικής.

Για την αξιολόγηση των μοντέλων κατηγοριοποίησης αυτού του προβλήματος εφαρμόστηκε ο μέσος όρος με προτεραιότητα στις κλάσεις. Οι τύποι που αφορούν τον υπολογισμό των μετρικών για μία κλάση είναι οι εξής:

- **Ευστοχία (Accuracy):** Η Ευστοχία αποτελεί μία από τις γνωστότερες μετρικές σε προβλήματα κατηγοριοποίησης και αναδεικνύει το ποσοστό σωστών προβλέψεων σε σχέση με όλο το σετ δεδομένων. Ο τύπος που εκφράζει την ακρίβεια είναι ο εξής:

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

Σημαντικό χαρακτηριστικό της Ευστοχίας είναι η άνιση επιρροή που έχει κάθε κλάση στο τελικό αποτέλεσμα της. Ένα σύνηθες παράδειγμα όπου το χαρακτηριστικό αυτό μπορεί να επιφέρει διαστρεβλωμένη εκτίμηση της επίδοσης ενός μοντέλου είναι η εκπαίδευσή του με χρήση μη ισορροπημένου σετ δεδομένων. Συγκεκριμένα, εάν τα περισσότερα δεδομένα ανήκουν σε μία κλάση τότε πιθανώς το μοντέλο να έχει μάθει να κατηγοριοποιεί σωστά, κατά κύριο λόγο, μόνο τα δεδομένα αυτής της κλάσης. Ωστόσο, κατά την αξιολόγησή του, εξαιτίας του γεγονότος ότι η πληθώρα των δεδομένων ανήκει σε αυτή την κλάση, η τιμή της ακρίβειας θα είναι υψηλή αποκρύπτοντας την αδυναμία γενίκευσης του μοντέλου.

- **Ακρίβεια (Precision):** Η Ακρίβεια αποτελεί ένδειξη της αξιοπιστίας που έχει μία πρόβλεψη όταν αυτή είναι Θετική. Ο τύπος που εκφράζει την Ακρίβεια είναι ο εξής:

$$Precision = \frac{TP}{TP + FP}$$

Επιστρέφοντας στο παράδειγμα του μη ισορροπημένου σετ δεδομένων, αυτή η μετρική μπορεί να αποκαλύψει την αδυναμία γενίκευσης του μοντέλου καθώς πολλές προβλέψεις που αφορούν την κλάση με τα περισσότερα δεδομένα θα είναι Ψευδώς Θετικές, μειώνοντας την τιμή της Ακρίβειας.

- **Ανάκληση (Recall):** Η Ανάκληση αποτελεί ένδειξη της ικανότητας του μοντέλου να ανιχνεύει τα Θετικά δεδομένα. Ο τύπος που εκφράζει την Ανάκληση είναι ο εξής:

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

- **F1-σκορ (F1-score):** Η μετρική F1-σκορ συνδυάζει τις μετρικές Ακρίβεια και Ανάκληση υπολογίζοντας τον αρμονικό μέσο τους και αποτελεί ένδειξη της επίδοσης του μοντέλου όσον αφορά αυτές τις δύο μετρικές. Ο τύπος που εκφράζει τη μετρική F1-σκορ είναι ο εξής:

$$F1 = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

Για παράδειγμα, έστω ένα μοντέλο με Ακρίβεια ίση με 0.9 Ανάκληση ίση με 0.2. Ο μέσος όρος αυτών των δύο τιμών ισούται με 0.55. Από την άλλη πλευρά, ο αρμονικός μέσος ισούται με 0.32, γεγονός που καθιστά ξεκάθαρο ότι για την αποφυγή χαμηλών τιμών στην μετρική F1-σκορ είναι απαραίτητο και η Ακρίβεια και η Ανάκληση να μην έχουν χαμηλές τιμές.

Συνεπώς, υπολογίζοντας τον μέσο όρο με προτεραιότητα στις κλάσεις κάθε μετρικής, η τελική μορφή τους που χρησιμοποιείται για την αξιολόγηση της επίδοσης των μοντέλων κατηγοριοποίησης είναι η εξής:

- **Μέση Ευστοχία (Average Accuracy):**

$$AverageAccuracy = \frac{\sum_{k=1}^N Accuracy_k}{N}$$

- **Μέση Ακρίβεια (Average Precision):**

$$AveragePrecision = \frac{\sum_{k=1}^N Precision_k}{N}$$

- **Μέση Ανάκληση (Average Recall):**

$$AverageRecall = \frac{\sum_{k=1}^N Recall_k}{N}$$

- **Μέσο F1-σκορ (Average F1-score):**

$$AverageF1 = \frac{\sum_{k=1}^N F1_k}{N}$$

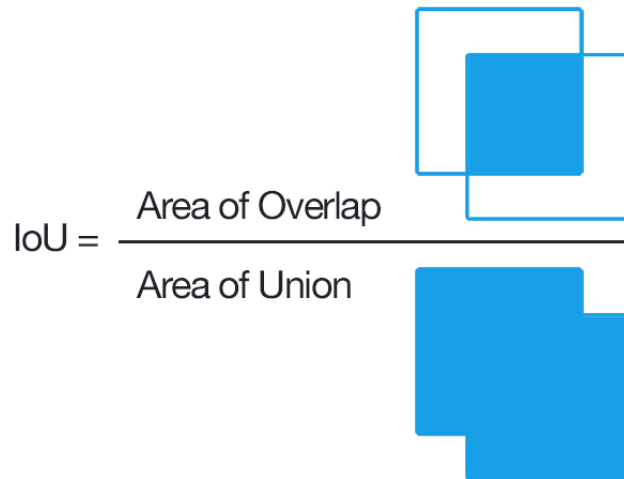
όπου N είναι ο αριθμός των κλάσεων.

4.9.2 Μετρικές μοντέλου κατάτμησης

Για το μοντέλο κατάτμησης Mask R-CNN, ως μετρική χρησιμοποιήθηκε ο συντελεστής Jaccard (Jaccard coefficient) ο οποίος ισοδύναμα ονομάζεται και Τομή ως προς την Ένωση (Intersection over Union — IoU) και εκφράζεται ως

$$J(A, B) = IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4.1)$$

Όταν ο συντελεστής Jaccard εφαρμόζεται σε εικόνες τότε η εξίσωση (4.1) μεταφράζεται ως ο λόγος του αριθμού των κοινών εικονοστοιχείων που έχει η πραγματική με την προβλεπόμενη εικόνα, με τον συνολικό αριθμό των εικονοστοιχείων που απαρτίζουν αυτές τις δύο εικόνες. Για την αρχιτεκτονική Mask R-CNN τα σύνολα της εξίσωσης (4.1) αποτελούνται από τα εικονοστοιχεία της εικόνας που βρίσκονται εντός του πραγματικού και του προβλεπόμενου οριοθετικού πλαισίου αντίστοιχα και η συγκεκριμένη μετρική αποτελεί ένδειξη του πόσο ακριβής είναι ο εντοπισμός ενός όγκου.



Εικόνα 4.15: Απεικόνιση λειτουργίας της μετρικής IoU.

Η μετρική IoU εφαρμόζεται για κάθε πρόβλεψη δεδομένου ξεχωριστά, συνεπώς για τη συνολική εκτίμηση της επίδοσης του μοντέλου υπολογίζεται ο μέσος όρος όλων των τιμών που προκύπτουν ο οποίος εκφράζεται ως

$$\text{Average IoU} = \frac{1}{N} \sum_{i=1}^N J(A_i, B_i)$$

όπου N είναι ο αριθμός των δεδομένων.

4.10 Εκπαίδευση

4.10.1 Εκπαίδευση μοντέλων κατηγοριοποίησης

Στα μοντέλα της κατηγοριοποίησης εφαρμόστηκαν και οι δύο στρατηγικές μεταφοράς μάθησης, όπως αυτές περιγράφονται στην ενότητα 4.3, και στις δύο προσεγγίσεις του προβλήματος. Και τα δύο μοντέλα έχουν ήδη εκπαιδευτεί πάνω σε ένα υποσύνολο του σετ δεδομένων ImageNet [9], το οποίο περιέχει περίπου ένα εκατομμύριο εικόνες που κατανέμονται, κατά προσέγγιση ομοιόμορφα, σε 1000 κλάσεις. Κάθε εφαρμογή μιας στρατηγικής σε ένα από τα μοντέλα, οδηγεί σε μία νέα ανεξάρτητη εκπαίδευση, συνεπώς για κάθε προσέγγιση δημιουργήθηκαν τέσσερα διαφορετικά μοντέλα κατηγοριοποίησης. Για την εκπαίδευση κάθε μοντέλου ως συνάρτηση κόστους χρησιμοποιήθηκε η κατηγορική διασταυρωμένη εντροπία. Και στις δύο στρατηγικές τα τρία τελευταία πλήρως συνδεδεμένα στρώματα που υπάρχουν εξ αρχής στις παραμετροποιήσεις VGG-16 και VGG-19 αντικαταστάθηκαν από τα εξής τρία στρώματα:

1. Πλήρως συνδεδεμένο στρώμα με 512 νευρώνες και συνάρτηση ενεργοποίησης ReLU.
2. Πλήρως συνδεδεμένο στρώμα με 256 νευρώνες και συνάρτηση ενεργοποίησης ReLU.
3. Πλήρως συνδεδεμένο στρώμα με τρεις εξόδους και συνάρτηση ενεργοποίησης Soft-max.

Πάνω στα πρώτα δύο πλήρως συνδεδεμένα στρώματα εφαρμόζεται η μέθοδος κανονικοποίησης Dropout με πιθανότητα απενεργοποίησης 50%. Αυτή η μέθοδος, σε κάθε εποχή κατά την εκπαίδευση απενεργοποιεί τυχαία ένα ποσοστό των νευρώνων του στρώματος στο οποίο εφαρμόζεται. Με αυτό τον τρόπο δίνεται η δυνατότητα στο νευρωνικό δίκτυο να μαθαίνει από διαφορετικό συνδυασμό νευρώνων σε κάθε εποχή, γεγονός που οδηγεί σε ενδεχόμενη εκμάθηση ποικίλων διαφορετικών χαρακτηριστικών. Κατά συνέπεια, αυτή η μέθοδος αποτελεί μια εξαιρετική επιλογή για την αντιμετώπιση της υπερπροσαρμογής.

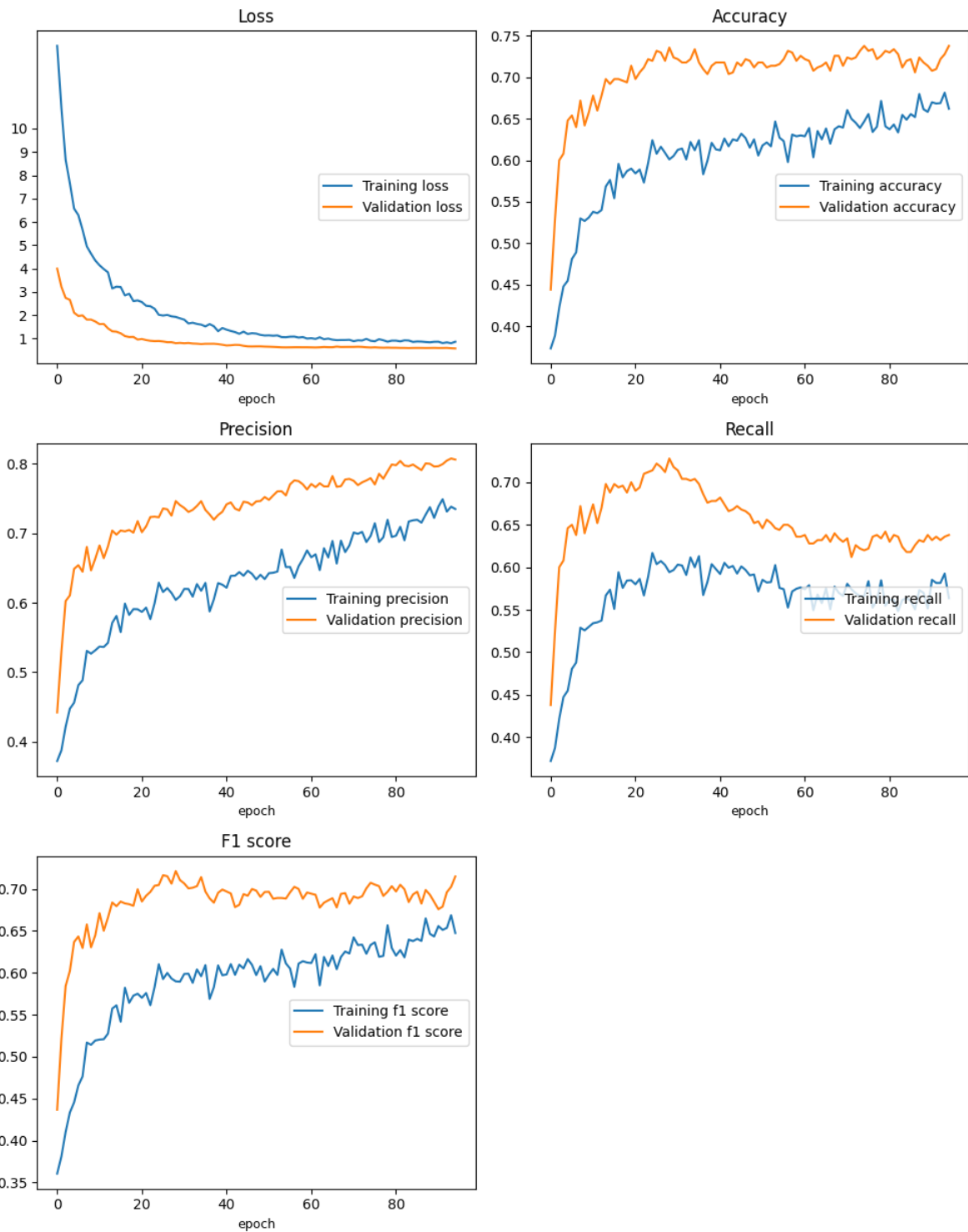
Στα πλαίσια εφαρμογής της πρώτης στρατηγικής, εκτός από την αντικατάσταση των τριών πλήρως συνδεδεμένων στρωμάτων δεν έγινε κάποια άλλη αλλαγή και όλα τα βάρη έμειναν άθικτα. Αναφορικά με την εφαρμογή της δεύτερης στρατηγικής, τα τρία νέα στρώματα μαζί με το τελευταίο σετ συνελικτικών στρωμάτων των μοντέλων VGG-16 και VGG-19 αντίστοιχα, είναι τα στρώματα τα οποία εκπαιδεύτηκαν πάνω στο σετ δεδομένων του προβλήματος, ενώ δεν τροποποιήθηκαν καθόλου τα βάρη των υπόλοιπων στρωμάτων. Ο Πίνακας 4.3 και ο Πίνακας 4.4 αποτυπώνουν την τελική μορφή των δύο αρχιτεκτονικών που χρησιμοποιήθηκαν, ενώ κατόπιν παρουσιάζονται τα γραφήματα εκπαίδευσης όλων των μοντέλων και για τις δύο προσεγγίσεις.

Πίνακας 4.3: Τελική μορφή αρχιτεκτονικής των μοντέλων που βασίζονται στην παραμετροποίηση VGG-16.

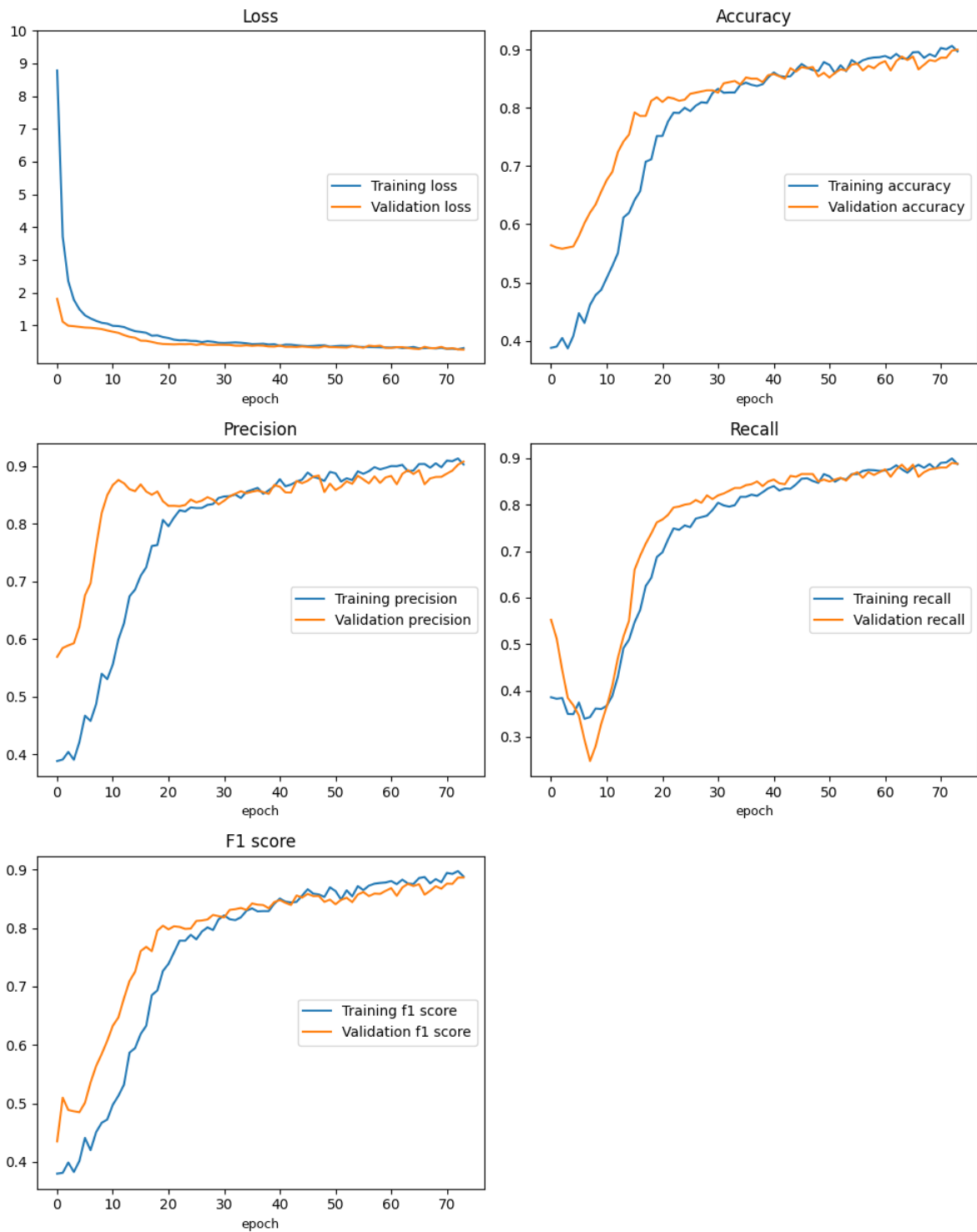
| Βάθος | Στρώμα | Χάρτες χαρακτηριστικών | Μέγεθος | Μέγεθος φίλτρου | Βήμα | Συνάρτηση ενεργοποίησης |
|-------|-----------------------|------------------------|-----------------------------|-----------------|------|-------------------------|
| - | Εισόδου | 1 | $224 \times 224 \times 3$ | - | - | - |
| 1 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Συνελικτικό | 64 | $224 \times 224 \times 64$ | 3×3 | 1 | ReLU |
| 2 | Μέγιστης συγκέντρωσης | 64 | $112 \times 112 \times 64$ | 2×2 | 2 | - |
| 3 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 4 | Συνελικτικό | 128 | $112 \times 112 \times 128$ | 3×3 | 1 | ReLU |
| 4 | Μέγιστης συγκέντρωσης | 128 | $56 \times 56 \times 128$ | 2×2 | 2 | - |
| 4 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 5 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | $56 \times 56 \times 256$ | 3×3 | 1 | ReLU |
| 6 | Μέγιστης συγκέντρωσης | 256 | $28 \times 28 \times 256$ | 2×2 | 2 | - |
| 7 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 8 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 9 | Συνελικτικό | 512 | $28 \times 28 \times 512$ | 3×3 | 1 | ReLU |
| 9 | Μέγιστης συγκέντρωσης | 512 | $14 \times 14 \times 512$ | 2×2 | 2 | - |
| 10 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 11 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 12 | Συνελικτικό | 512 | $14 \times 14 \times 512$ | 3×3 | 1 | ReLU |
| 12 | Μέγιστης συγκέντρωσης | 512 | $7 \times 7 \times 512$ | 2×2 | 2 | - |
| 14 | Πλήρως συνδεδεμένο | - | 512 | - | - | ReLU |
| 15 | Πλήρως συνδεδεμένο | - | 256 | - | - | ReLU |
| 16 | Εξόδου | - | 3 | - | - | Softmax |

Πίνακας 4.4: Τελική μορφή αρχιτεκτονικής των μοντέλων που βασίζονται στην παραμετροποίηση VGG-19.

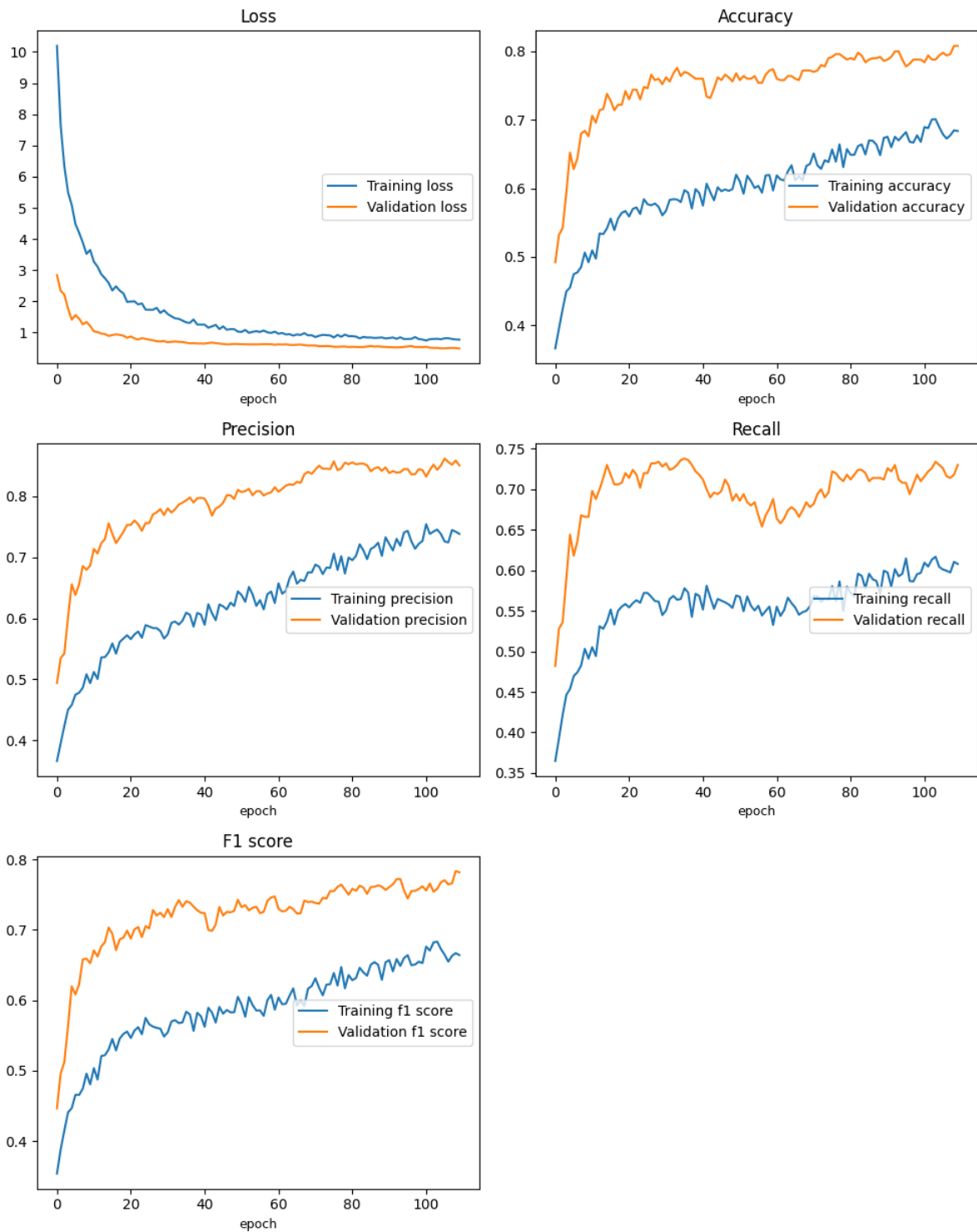
| Βάθος | Στρώμα | Χάρτες χαρακτηριστικών | Μέγεθος | Μέγεθος φίλτρου | Βήμα | Συνάρτηση ενεργοποίησης |
|-------|-----------------------|------------------------|-----------------|-----------------|------|-------------------------|
| - | Εισόδου | 1 | 224 × 224 × 3 | - | - | - |
| 1 | Συνελικτικό | 64 | 224 × 224 × 64 | 3 × 3 | 1 | ReLU |
| 2 | Συνελικτικό | 64 | 224 × 224 × 64 | 3 × 3 | 1 | ReLU |
| 2 | Μέγιστης συγκέντρωσης | 64 | 112 × 112 × 64 | 2 × 2 | 2 | - |
| 3 | Συνελικτικό | 128 | 112 × 112 × 128 | 3 × 3 | 1 | ReLU |
| 4 | Συνελικτικό | 128 | 112 × 112 × 128 | 3 × 3 | 1 | ReLU |
| 4 | Μέγιστης συγκέντρωσης | 128 | 56 × 56 × 128 | 2 × 2 | 2 | - |
| 4 | Συνελικτικό | 256 | 56 × 56 × 256 | 3 × 3 | 1 | ReLU |
| 5 | Συνελικτικό | 256 | 56 × 56 × 256 | 3 × 3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | 56 × 56 × 256 | 3 × 3 | 1 | ReLU |
| 6 | Συνελικτικό | 256 | 56 × 56 × 256 | 3 × 3 | 1 | ReLU |
| 6 | Μέγιστης συγκέντρωσης | 256 | 28 × 28 × 256 | 2 × 2 | 2 | - |
| 7 | Συνελικτικό | 512 | 28 × 28 × 512 | 3 × 3 | 1 | ReLU |
| 8 | Συνελικτικό | 512 | 28 × 28 × 512 | 3 × 3 | 1 | ReLU |
| 9 | Συνελικτικό | 512 | 28 × 28 × 512 | 3 × 3 | 1 | ReLU |
| 6 | Συνελικτικό | 512 | 28 × 28 × 256 | 3 × 3 | 1 | ReLU |
| 9 | Μέγιστης συγκέντρωσης | 512 | 14 × 14 × 512 | 2 × 2 | 2 | - |
| 10 | Συνελικτικό | 512 | 14 × 14 × 512 | 3 × 3 | 1 | ReLU |
| 11 | Συνελικτικό | 512 | 14 × 14 × 512 | 3 × 3 | 1 | ReLU |
| 12 | Συνελικτικό | 512 | 14 × 14 × 512 | 3 × 3 | 1 | ReLU |
| 6 | Συνελικτικό | 512 | 14 × 14 × 256 | 3 × 3 | 1 | ReLU |
| 12 | Μέγιστης συγκέντρωσης | 512 | 7 × 7 × 512 | 2 × 2 | 2 | - |
| 14 | Πλήρως συνδεδεμένο | - | 512 | - | - | ReLU |
| 15 | Πλήρως συνδεδεμένο | - | 256 | - | - | ReLU |
| 16 | Εξόδου | - | 3 | - | - | Softmax |



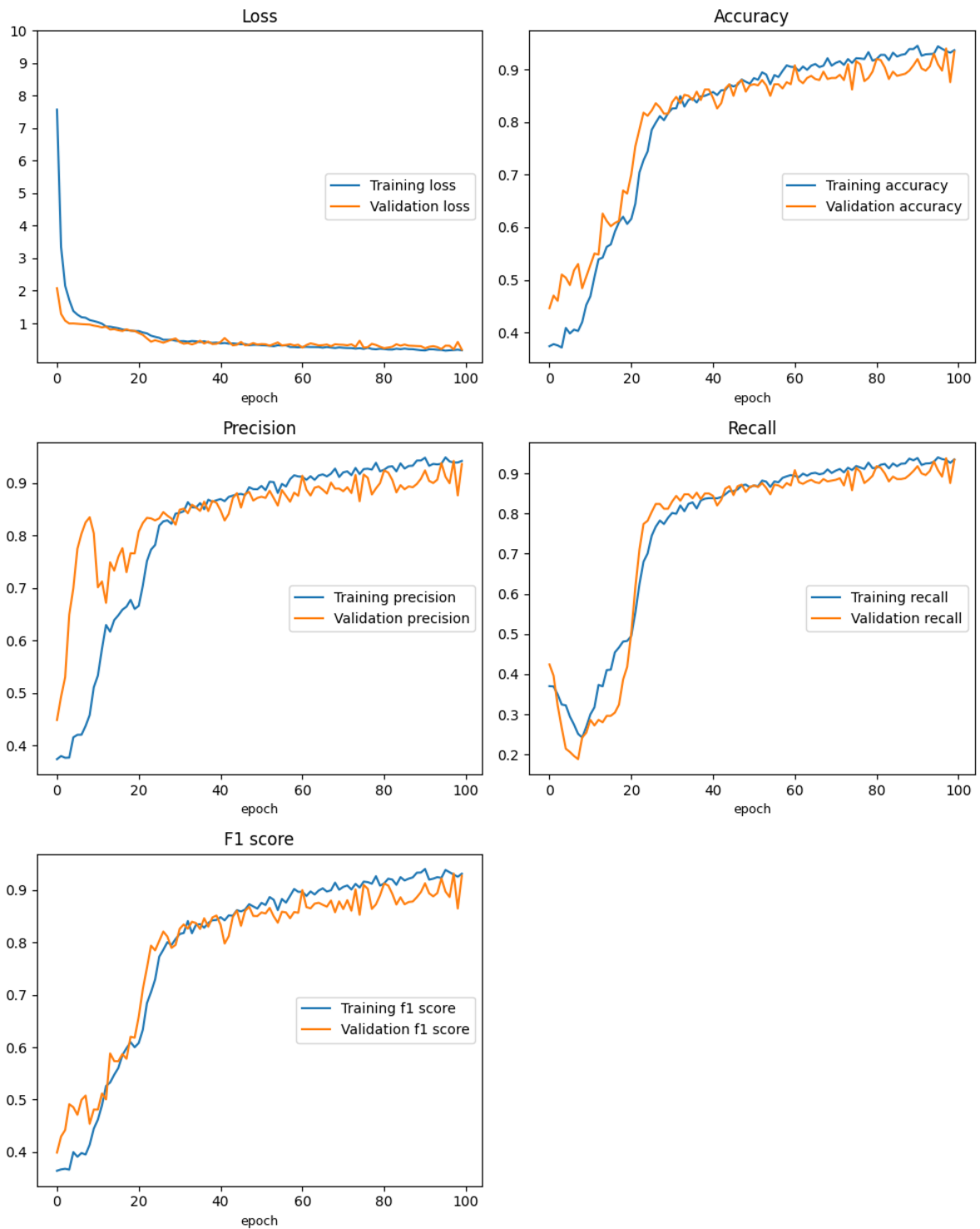
Εικόνα 4.16: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 ως εξαγωγέα χαρακτηριστικών, για την πρώτη προσέγγιση. Συνολικά 94 εποχές εκπαίδευσης.



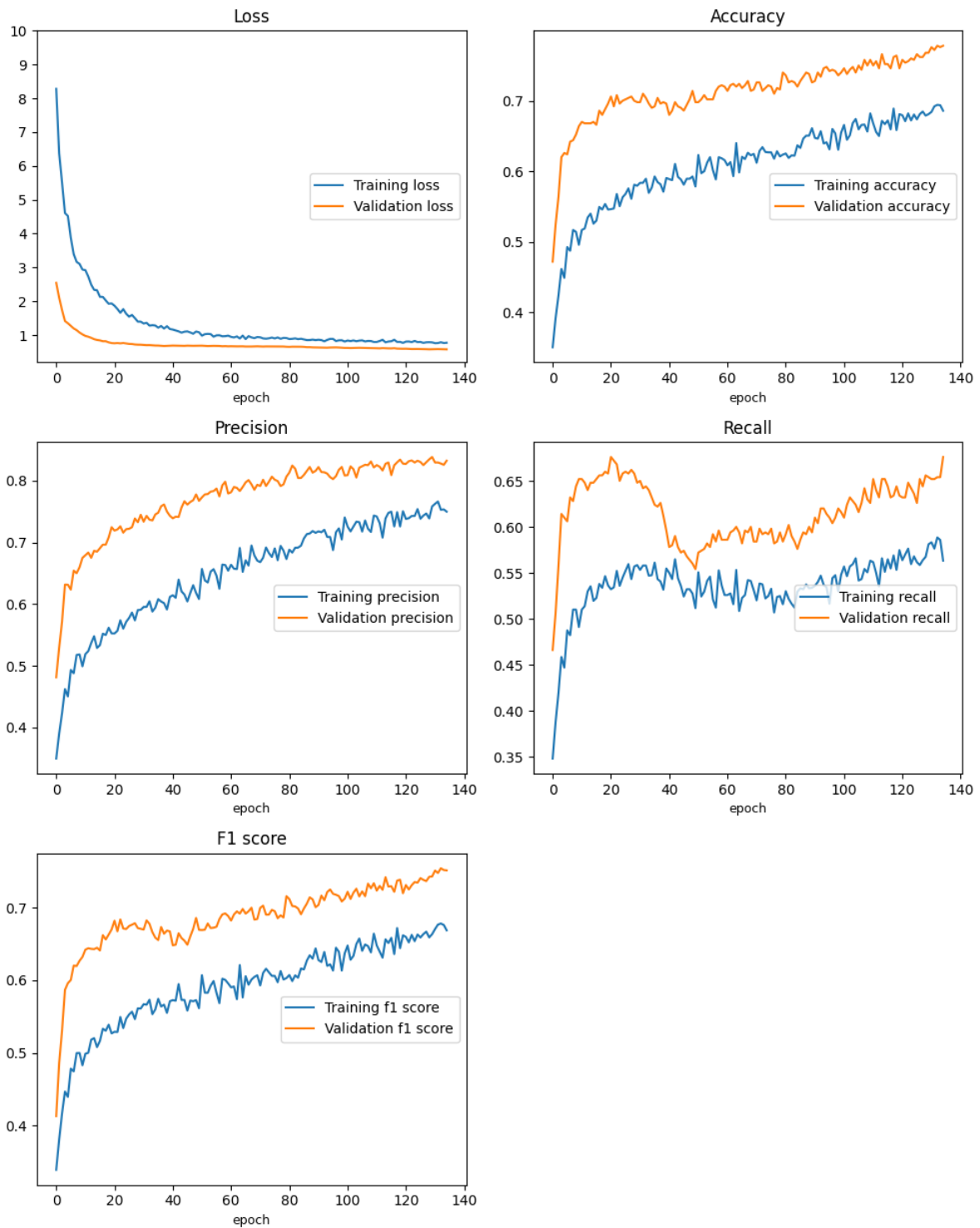
Εικόνα 4.17: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 με fine-tuning, για την πρώτη προσέγγιση. Συνολικά 73 εποχές εκπαίδευσης.



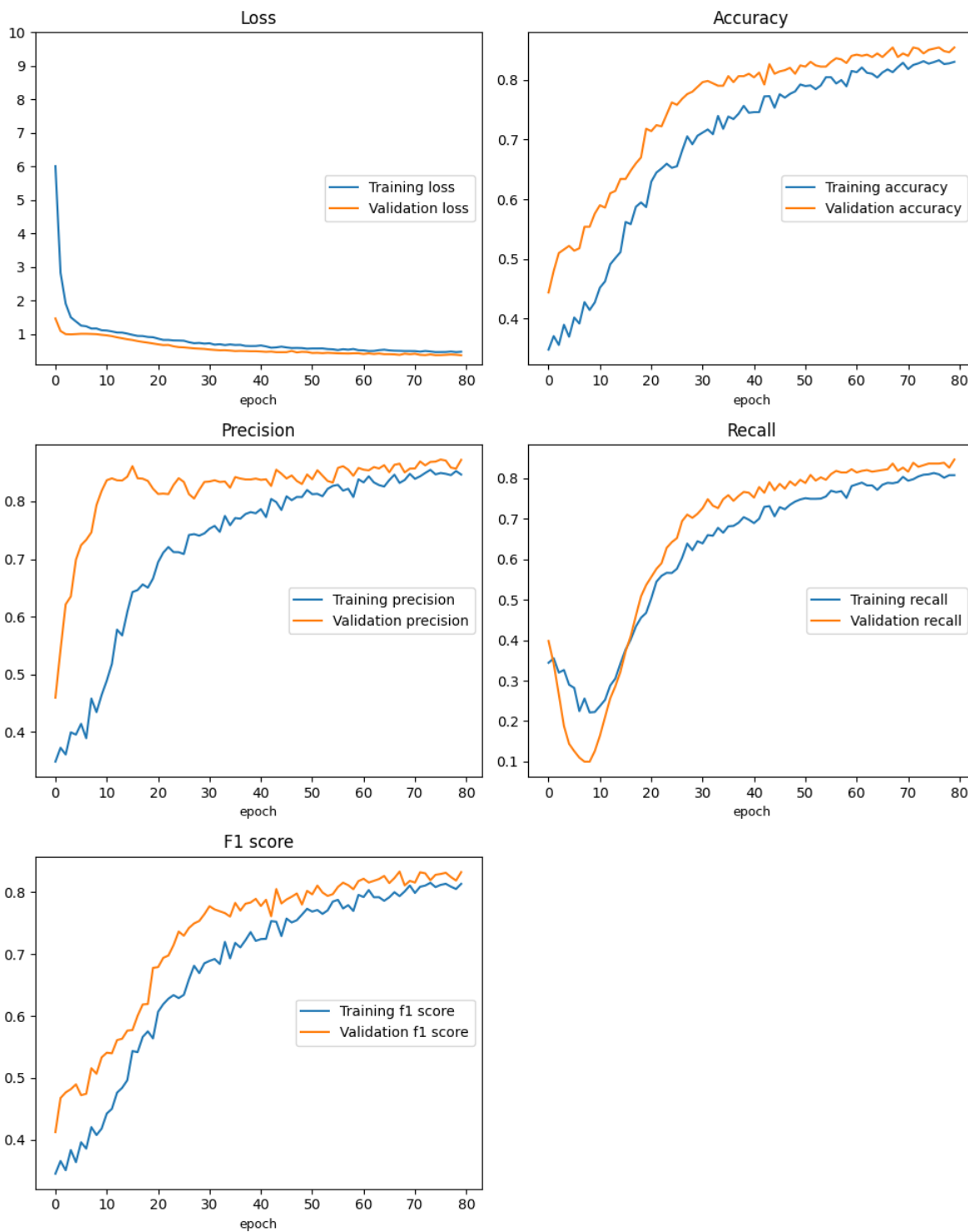
Εικόνα 4.18: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-19 ως εξαγωγέα χαρακτηριστικών, για την πρώτη προσέγγιση. Συνολικά 108 εποχές εκπαίδευσης.



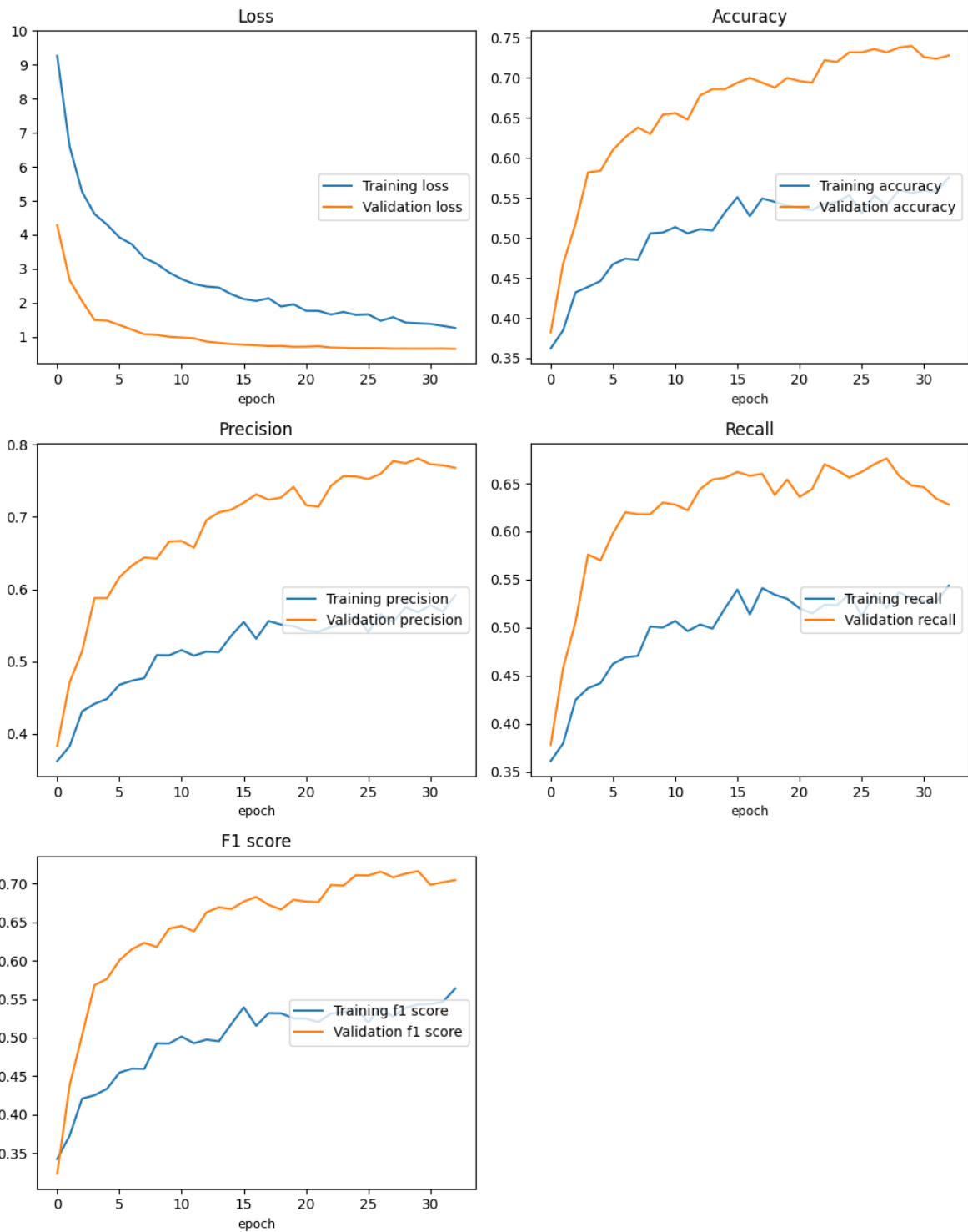
Εικόνα 4.19: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-19 με fine-tuning, για την πρώτη προσέγγιση. Συνολικά 99 εποχές εκπαίδευσης.



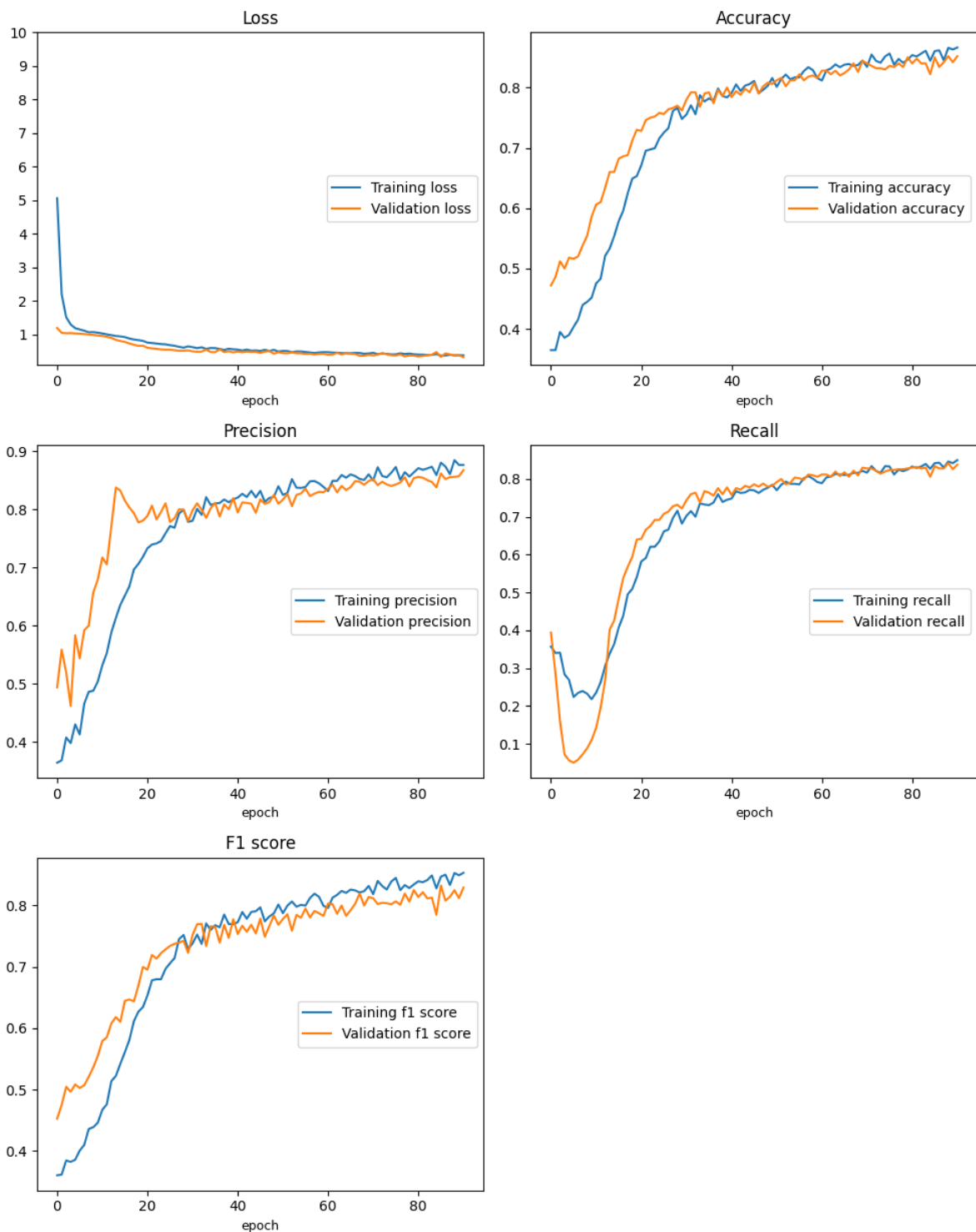
Εικόνα 4.20: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 ως εξαγωγέα χαρακτηριστικών, για τη δεύτερη προσέγγιση. Συνολικά 136 εποχές εκπαίδευσης.



Εικόνα 4.21: Γραφήματα της εκπαίδευσης του μοντέλου που χρησιμοποιεί την παραμετροποίηση VGG-16 με fine-tuning, για τη δεύτερη προσέγγιση. Συνολικά 78 εποχές εκπαίδευσης.



Εικόνα 4.22: Γραφήματα της εκπαίδευσης του μοντέλου που είναι βασισμένο στην παραμετροποίηση VGG-19 ως εξαγωγή χαρακτηριστικών, για τη δεύτερη προσέγγιση. Συνολικά 32 εποχές εκπαίδευσης.



Εικόνα 4.23: Γραφήματα της εκπαίδευσης του μοντέλου που είναι βασισμένο στην παραμετροποίηση VGG-19 με fine-tuning, για τη δεύτερη προσέγγιση. Συνολικά 87 εποχές εκπαίδευσης.

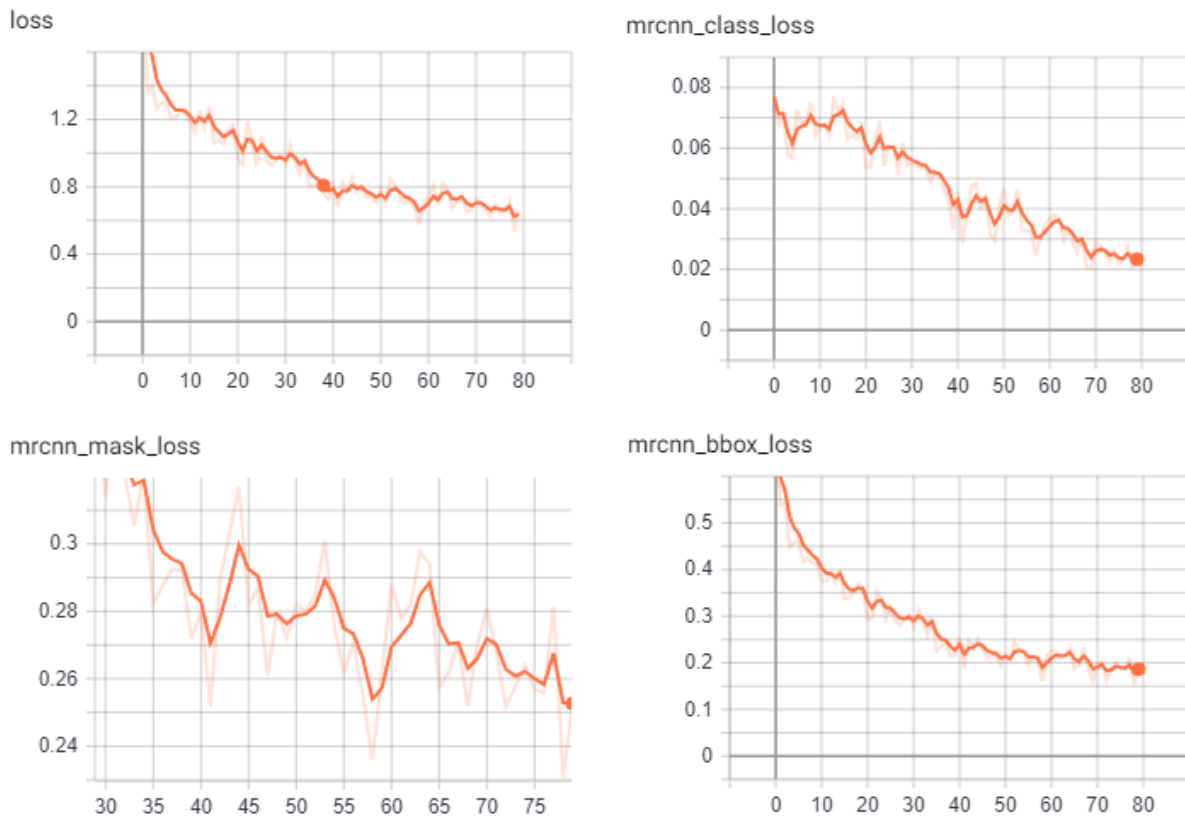
4.10.2 Εκπαίδευση μοντέλων κατάτμησης

Στο μοντέλο κατάτμησης Mask R-CNN εφαρμόστηκε μόνο η στρατηγική fine-tuning κρατώντας όμως όλα τα στρώματα ακέραια, ακόμα και τα στρώματα εξόδου. Το μοντέλο Mask R-CNN είναι προ-εκπαιδευμένο πάνω στο σετ δεδομένων COCO [23] το οποίο περιέχει 330 χιλιάδες λέξεις που κατανέμονται σε 91 κλάσεις. Η περαιτέρω εκπαίδευσή

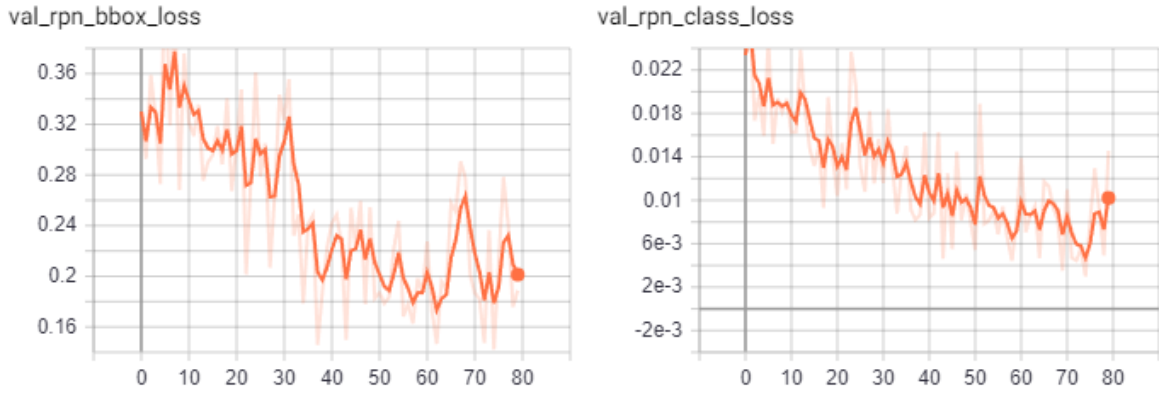
του πάνω στο σετ δεδομένων του συγκεκριμένου προβλήματος έγινε εκπαιδύοντας το μοντέλο εξ ολοκλήρου για 80 εποχές. Ωστόσο οι καλύτερες επιδόσεις καταγράφηκαν την εποχή 72, συνεπώς για την τελική αξιολόγηση διατηρήθηκαν τα βάρη εκείνης της εποχής. Τέλος, οι παρακάτω εικόνες αποτυπώνουν τη συμπεριφορά του μοντέλου κατά την εκπαίδευση, μέσω των συναρτήσεων απώλειας που περιγράφονται στο άρθρο [14].



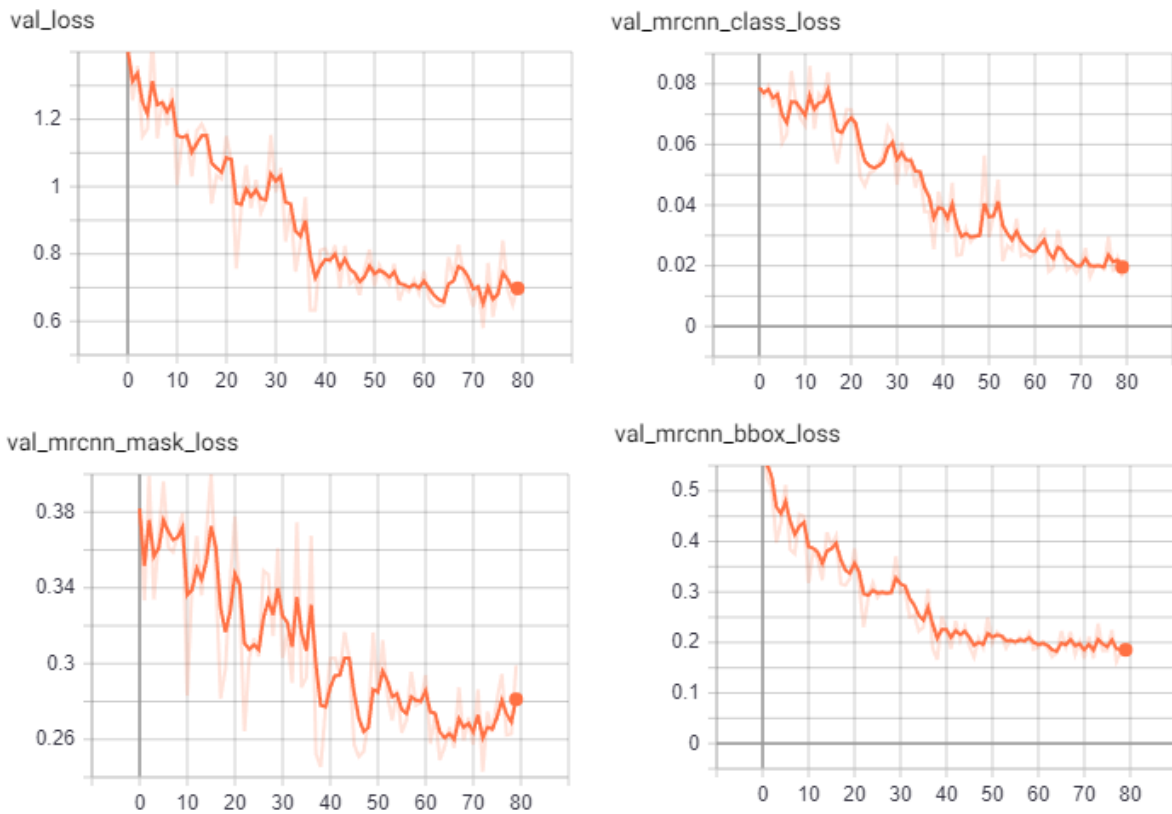
Εικόνα 4.24: Γράφημα των συναρτήσεων απώλειας του Δικτύου Πρότασης Περιοχών για το σετ εκπαίδευσης.



Εικόνα 4.25: Γράφημα των συναρτήσεων απώλειας του συνολικού μοντέλου Mask R-CNN για το σετ εκπαίδευσης.



Εικόνα 4.26: Γράφημα των συναρτήσεων απώλειας του Δικτύου Πρότασης Περιοχών για το σετ επικύρωσης.

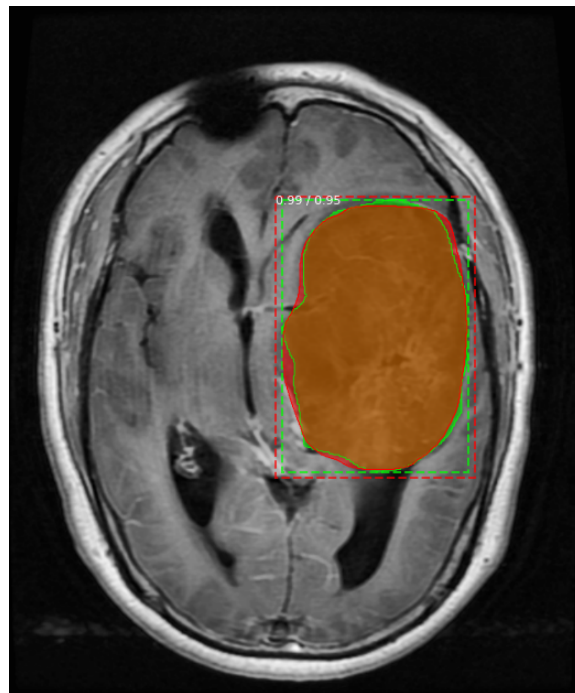


Εικόνα 4.27: Γραφήμα της εκπαίδευσης συνολικά του μοντέλου Mask R-CNN για το σετ επικύρωσης.

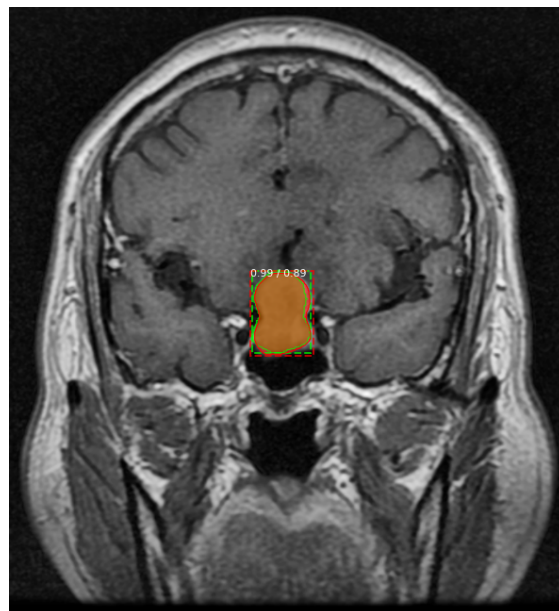
4.11 Αποτελέσματα

Ενδεικτικά, η Εικόνα 4.28, η Εικόνα 4.29 και η Εικόνα 4.30 απεικονίζουν κάποια από τα αποτελέσματα του μοντέλου κατάτμησης στο σετ αξιολόγησης. Με κόκκινο χρώμα παρουσιάζεται η μάσκα και το οριοθετικό πλαίσιο που παρήχθησαν από το μοντέλο, ενώ με πράσινο χρώμα απεικονίζονται τα πραγματικά χωρικά όρια του όγκου. Παράλληλα, σε κάθε Εικόνα παρουσιάζονται και δύο πραγματικοί αριθμοί που αντιπροσωπεύουν την

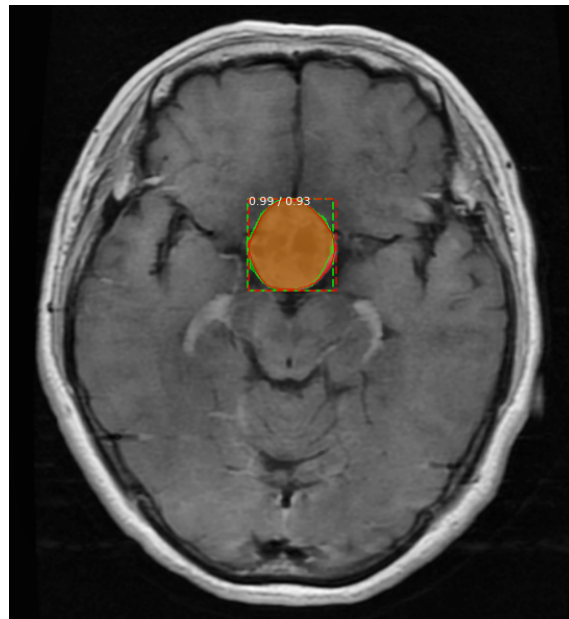
αυτοπεποίθηση που είχε το μοντέλο για τη συγκεκριμένη πρόβλεψη και την μετρική IoU αντίστοιχα.



Εικόνα 4.28: Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.95.



Εικόνα 4.29: Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.89.



Εικόνα 4.30: Απεικόνιση αποτελέσματος του μοντέλου κατάτμησης με αυτοπεποίθηση πρόβλεψης 0.99 και IoU 0.93.

Τέλος, ο Πίνακας 4.5 και ο Πίνακας 4.6 παρουσιάζουν τα αποτελέσματα της κατηγοριοποίησης της πρώτης και δεύτερης προσέγγισης αντίστοιχα.

Πίνακας 4.5: Αποτελέσματα πρώτης προσέγγισης.

| Πρώτη Προσέγγιση | Ευστοχία | Ακρίβεια | Ανάκληση | F1 |
|--------------------------|----------|----------|----------|-------|
| VGG-16 χωρίς fine-tuning | 0.800 | 0.786 | 0.801 | 0.785 |
| VGG-16 με fine-tuning | 0.9300 | 0.921 | 0.925 | 0.923 |
| VGG-19 χωρίς fine-tuning | 0.800 | 0.778 | 0.787 | 0.779 |
| VGG-19 με fine-tuning | 0.953 | 0.946 | 0.948 | 0.947 |

Πίνακας 4.6: Αποτελέσματα δεύτερης προσέγγισης.

| Δεύτερη Προσέγγιση | IoU | Ευστοχία | Ακρίβεια | Ανάκληση | F1 |
|--------------------------|-------|----------|----------|----------|-------|
| Mask R-CNN | 0.746 | - | - | - | - |
| VGG-16 χωρίς fine-tuning | - | 0.731 | 0.713 | 0.713 | 0.705 |
| VGG-16 με fine-tuning | - | 0.796 | 0.778 | 0.796 | 0.783 |
| VGG-19 χωρίς fine-tuning | - | 0.783 | 0.770 | 0.773 | 0.768 |
| VGG-19 με fine-tuning | - | 0.805 | 0.804 | 0.805 | 0.796 |

5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

5.1 Συμπεράσματα

Η παρούσα εργασία αναδεικνύει την ικανότητα της βαθιάς μάθησης να αντεπεξέρχεται σε απαιτητικά και ουσιώδη για την ανθρωπότητα προβλήματα, όπως είναι αυτό της έγκαιρης και σωστής διάγνωσης εγκεφαλικών όγκων. Για την επίτευξη αυτού του σκοπού χρησιμοποιήθηκαν τρεις διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων και δύο στρατηγικές μεταφοράς μάθησης, οι οποίες αποτέλεσαν τη βάση για την υλοποίηση πέντε διαφορετικών μοντέλων, ενώ ακολουθήθηκαν δύο διαφορετικές προσεγγίσεις. Αναφορικά με τη χρήση μεταφοράς μάθησης, τα αποτελέσματα δείχνουν πως η χρήση της στρατηγικής fine-tuning βελτιώνει σημαντικά την επίδοση όλων των μοντέλων κατηγοριοποίησης και στις δύο προσεγγίσεις. Η ύπαρξη μεγαλύτερου βάθους και κατά συνέπεια περισσότερων εκπαιδευσιμων βαρών, έδωσε τη δυνατότητα στο μοντέλο που βασίστηκε στην παραμετροποίηση VGG-19 να έχει υψηλότερες επιδόσεις, και στις δύο προσεγγίσεις, από εκείνο που βασίστηκε στην VGG-16. Βέβαια οι επιδόσεις και των δύο είναι άκρως ικανοποιητικές καθιστώντας και τις δύο παραμετροποιήσεις κατάλληλες να εφαρμοστούν στο συγκεκριμένο πρόβλημα. Το μοντέλο κατάτμησης που βασίστηκε στην αρχιτεκτονική Mask R-CNN παρουσιάζει ικανοποιητικά αποτελέσματα, ωστόσο η προσέγγιση στην οποία η κατηγοριοποίηση εφαρμόζεται στην αρχική εικόνα και όχι στον κατατμημένο όγκο παρουσιάζει καλύτερα αποτελέσματα. Παρ' όλα αυτά, τόσο το μοντέλο κατάτμησης αυτούσιο όσο και η δεύτερη προσέγγιση που προτάθηκε για το πρόβλημα, παρουσιάζουν εξαιρετικό ενδιαφέρον. Συγκεκριμένα, με τη χρήση του μοντέλου κατάτμησης προσδίδεται διαδραστικότητα μεταξύ του υπολογιστή και του ανθρώπου η οποία μπορεί να λειτουργήσει ως μία επιπλέον δικλείδα ασφαλείας σε ένα αυτοματοποιημένο σύστημα κατηγοριοποίησης εγκεφαλικών όγκων, καθώς θα δίνεται η δυνατότητα στο γιατρό να ελέγξει τον τρόπο με τον οποίο παράχθηκε το αποτέλεσμα, εάν αυτό κριθεί σκόπιμο. Επίσης, μέσω της κατάτμησης αφαιρούνται μέρη της εικόνας, όπως είναι το κρανίο και τα μάτια, τα οποία μπορούν να λειτουργήσουν μόνο ως θόρυβος για έναν κατηγοριοποιητή βλάπτοντας την απόδοση του, γεγονός που επισημαίνει τις δυναμικά υψηλές δυνατότητες της δεύτερης προσέγγισης. Βέβαια, για ορισμένους εγκεφαλικούς όγκους σημαντική παράμετρο αποτελεί ο περιβάλλον χώρος τους. Για παράδειγμα, πιθανώς εάν είχε ακολουθηθεί μια λιγότερο αυστηρή κατάτμηση, όσον αφορά τα χωρικά όρια, και είχε συμπεριληφθεί ο περιβάλλον χώρος στις εικόνες ασθενών με όγκο υπόφυσης ο οποίος εμφανίζεται σε σταθερή θέση στον εγκέφαλο, να είχαν επιτευχθεί καλύτερα αποτελέσματα στην δεύτερη προσέγγιση. Εν κατακλείδι, αυτή τη στιγμή υπάρχουν ποικίλοι τρόποι με τους οποίους η βαθιά μάθηση μπορεί να φανεί χρήσιμη για τον εντοπισμό και την κατηγοριοποίηση εγκεφαλικών όγκων, ορισμένοι από τους οποίους παρουσιάστηκαν σε αυτή την εργασία. Επίσης, είναι δεδομένο ότι η ίδια δεν έχει φτάσει καν κοντά στο μέγιστο των δυνατοτήτων της, τόσο σε γενικό πλαίσιο όσο και στον τομέα της ιατρικής, μιας και αποτελεί έναν ταχύτατα αναπτυσσόμενο κλάδο της τεχνητής νοημοσύνης. Ο συνδυασμός των δύο παραπάνω γεγονότων καθιστά το ενδεχόμενο της πλήρους ένταξης της βαθιάς μάθησης στην ιατροφαρμακευτική περίθαλψη με επιδόσεις καλύτερες ακόμα και από το ιατρικό προσωπικό, εξαιρετικά πιθανό στο μέλλον.

5.2 Μελλοντικές εργασίες

Υπάρχουν πολλές τροποποιήσεις, επεκτάσεις και πειραματισμοί που αξίζουν να λάβουν χώρα για το συγκεκριμένο πρόβλημα, αλλά είναι αδύνατο να συμπεριληφθούν όλα σε μία εργασία. Ειδικότερα, ενδιαφέρον παρουσιάζει η ιδέα του γίνει η εκπαίδευση και η αξιολόγηση των μοντέλων σε ένα μεγαλύτερο σετ δεδομένων με περισσότερα είδη εγκεφαλικών όγκων. Με αυτόν τον τρόπο, θα πηγάζει μεγαλύτερη αυτοπεποίθηση από τα όποια συμπεράσματα ληφθούν. Φυσικά, σκόπιμη κρίνεται η εφαρμογή ποικίλων συνδυασμών αρχιτεκτονικών και υπερπαρμέτρων για το ίδιο πρόβλημα. Μία αξιοσημείωτη επέκταση θα ήταν η προσθήκη της δυνατότητας αξιολόγησης του βαθμού κακοήθειας (tumor grade) ενός εγκεφαλικού όγκου. Ακολουθώντας το ίδιο μοτίβο, καθοριστική στον τρόπο αντιμετώπισης ενός όγκου θα ήταν η εμφάνιση ενός ποσοστού θνησιμότητας η οποία θα ήταν βασισμένη στο ιστορικό ανθρώπων με ίδιο είδος όγκου και παρόμοια βιολογικά χαρακτηριστικά. Επίσης, εξαιρετικά ενδιαφέρουσα θα ήταν η προσθήκη φυσικής επεξεργασίας γλώσσας (natural language processing – NLP) μέσω της οποίας το είδος του όγκου μαζί με άλλα χαρακτηριστικά του θα συντελούσαν ένα κείμενο φυσικής γλώσσας, προσομοιώνοντας τη διαδικασία συγγραφής ιατρικής γνωμάτευσης.

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

| | |
|-------|---|
| MRI | magnetic resonance imaging |
| AI | artificial intelligence |
| CAD | computer assisted diagnosis |
| RNN | Recurrent neural network |
| LSTM | Long short-term memory |
| CNN | Convolutional neural network |
| RPN | Region Proposal Network |
| R-CNN | Region-Based Convolutional Neural Network |
| FPN | Feature Pyramid Network |
| TP | True Positive |
| TN | True Negative |
| FP | False Positive |
| FN | False Negative |
| IoU | Intersection over Union |
| VGG | Visual Geometry Group |
| NLP | Natural Language Processing |

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Waleed Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017.
- [2] Parnian Afshar, Konstantinos N. Plataniotis, and Arash Mohammadi. Capsule networks' interpretability for brain tumor classification via radiomics analyses. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3816–3820, 2019.
- [3] Zeynettin Akkus, Issa Ali, Jiri Sedlar, Timothy L. Kline, Jay P. Agrawal, Ian F. Parney, Caterina Giannini, and Bradley J. Erickson. Predicting 1p19q chromosomal deletion of low-grade gliomas from mr images using deep learning, 2016.
- [4] Zeynettin Akkus, Jiri Sedlar, Lucie Coufalova, Panagiotis Korfiatis, Timothy L Kline, Joshua D Warner, Jay Agrawal, and Bradley J Erickson. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. *Cancer Imaging*, 15(1):1–10, 2015.
- [5] Javeria Amin, Muhammad Sharif, Mussarat Yasmin, Tanzila Saba, Muhammad Almas Anjum, and Steven Lawrence Fernandes. A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *Journal of medical systems*, 43(11):1–16, 2019.
- [6] Sheenam Bansal and Raman Maini. Performance analysis of color based region split and merge and otsu's thresholding techniques for brain tumor extraction.
- [7] Ahmad Chaddad. Automated feature extraction in brain tumor by magnetic resonance imaging using gaussian mixture models. *International Journal of Biomedical Imaging*, 2015:1–11, 07 2015.
- [8] Jun Cheng. brain tumor dataset, Apr 2017.
- [9] 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. Ieee, 2009.
- [10] RB Dubey, M Hanmandlu, SK Gupta, and SK Gupta. Region growing for mri brain tumor volume analysis rb dubey1, m. hanmandlu2, sk gupta3 and sk gupta4. *Indian Journal of Science and Technology*, 2(9), 2009.
- [11] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
- [12] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [13] Nidhi Gupta and Pritee Khanna. A non-invasive and adaptive cad system to detect brain tumor from t2-weighted mris using customized otsu's thresholding with prominent features and supervised learning. *Signal Processing: Image Communication*, 59:18–26, 2017.
- [14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [15] Jeff Heaton. *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.
- [16] Karsten Held, E Rota Kops, Bernd J Krause, William M Wells, Ron Kikinis, and H-W Muller-Gartner. Markov random field segmentation of brain mr images. *IEEE transactions on medical imaging*, 16(6):878–886, 1997.
- [17] Umit Ilhan and Ahmet Ilhan. Brain tumor segmentation based on a new threshold approach. *Procedia Computer Science*, 120:580–587, 2017. 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary.
- [18] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge, 2018.

- [19] T Kalaiselvi and P Nagaraja. Brain tumor segmentation of mri brain images through fcm clustering and seeded region growing technique. *International Journal of Applied Engineering Research*, 10(76):427–432, 2015.
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [21] Yang Li, Ye Mingquan, and Zhang Hao. Tumor diagnosis based on the gmm feature decision classification of brain mr images. *International journal of Multimedia and Ubiquitous Engineering*, 11(3):37–44, 2016.
- [22] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection, 2017.
- [23] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.
- [24] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [25] Raghav Mehta and Jayanthi Sivaswamy. M-net: A convolutional neural network for deep brain structure segmentation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 437–440. IEEE, 2017.
- [26] Norhashimah Mohd Saad, Syed Ab Rahman Abu Bakar, Sobri Muda, and Musa Mokji. Brain lesion segmentation of diffusion-weighted mri using thresholding technique. volume 35, pages 249–254, 11 2011.
- [27] Vinay Rao, M Shari Sarabi, and Ayush Jaiswal. Brain tumor segmentation with deep learning. *MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS)*, 59, 2015.
- [28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.
- [29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [30] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [31] Yasser Salman. Modified technique for volumetric brain tumor measurements. *Journal of Biomedical Science and Engineering*, 02:16–19, 01 2009.
- [32] Mie Sato, S. Lakare, Ming Wan, A. Kaufman, and M. Nakajima. A gradient magnitude based region growing algorithm for accurate segmentation. In *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, volume 3, pages 448–451 vol.3, 2000.
- [33] Roshan G Selkar and MN Thakare. Brain tumor detection and segmentation by using thresholding and watershed algorithm. *International Journal of Advanced Information and Communication Technology*, 1(3):321–4, 2014.
- [34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [35] N Subbanna and T Arbel. Probabilistic gabor and markov random fields segmentation of brain tumours in mri volumes. *Proc MICCAI Brain Tumor Segmentation Challenge (BRATS)*, pages 28–31, 2012.
- [36] Nagesh Subbanna, Doina Precup, and Tal Arbel. Iterative multilevel mrf leveraging context and voxel information for brain tumour segmentation in mri. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [37] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.
- [38] Ken CL Wong, Tanveer Syeda-Mahmood, and Mehdi Moradi. Building medical image classifiers with very limited data using segmentation networks. *Medical image analysis*, 49:105–116, 2018.
- [39] Liang Zhao, Wei Wu, and Jason J. Corso. Brain tumor segmentation based on gmm and active contour method with a model-aware edge map. 2012.