



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ  
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

## Χρήση Τεχνικών Βαθιάς Μάθησης για την Κατάτμηση Εγκεφαλικών Όγκων για τη Διάγνωση της Νόσου Alzheimer

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΛΚΙΒΙΑΔΗΣ – ΜΑΡΙΟΣ Ε. ΚΟΥΚΟΣ

ΙΩΑΝΝΗΣ Θ. ΦΙΛΙΠΠΑΣ

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

**Συνεπιβλέποντες :** Θάνος Τάγαρης

Υποψήφιος Διδάκτορας Ε.Μ.Π.

Μάρα Σδράκα

Υποψήφια Διδάκτορας Ε.Μ.Π.

Αθήνα, Νοέμβριος 2018





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ  
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

## Χρήση Τεχνικών Βαθιάς Μάθησης για την Κατάτμηση Εγκεφαλικών Όγκων για την Διάγνωση της Νόσου Alzheimer

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΛΚΙΒΙΑΔΗΣ – ΜΑΡΙΟΣ Ε. ΚΟΥΚΟΣ

ΙΩΑΝΝΗΣ Θ. ΦΙΛΙΠΠΑΣ

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

**Συνεπιβλέποντες :** Θάνος Τάγαρης Μάρα Σδράκα  
Υποψήφιος Διδάκτορας Ε.Μ.Π. Υποψήφια Διδάκτορας Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13<sup>η</sup> Νοεμβρίου 2018.

.....

Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

.....

Γεώργιος Στάμου  
Αναπληρωτής Καθηγητής Ε.Μ.Π.

.....

Κωνσταντίνα Νικήτα  
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Νοέμβριος 2018

.....

**Αλκιβιάδης – Μάριος Ε. Κούκος**

**Ιωάννης Θ. Φίλιππας**

Διπλωματούχοι Ηλεκτρολόγοι Μηχανικοί και Μηχανικοί Υπολογιστών Ε.Μ.Π.

Copyright © Αλκιβιάδης – Μάριος Ε. Κούκος, 2018.

Copyright © Ιωάννης Θ. Φίλιππας, 2018.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Περίληψη

Τις τελευταίες δεκαετίες η ερευνητική δραστηριότητα στον τομέα της Ανάλυσης Εικόνας έχει σημειώσει ραγδαία ανάπτυξη μέσα από πληθώρα μελετών και εφαρμογών με βασικό στόχο να υποστηριχθούν και να διευκολυνθούν βασικές λειτουργίες στην καθημερινότητα των ανθρώπων ανά την υφήλιο. Προς επίτευξη του συγκεκριμένου σκοπού, πολλές από αυτές τις εφαρμογές αξιοποιούν τεχνικές από τους τομείς της Τεχνητής Νοημοσύνης και συγκεκριμένα της Μηχανικής Μάθησης και της Βαθιάς Μάθησης. Ειδικότερα στον τομέα της Ιατρικής, οι σύγχρονες επιστημονικές τάσεις διατυπώνουν ότι οι διαδικασίες της διάγνωσης αλλά και της αντιμετώπισης διαφόρων παθήσεων μπορούν πλέον να γίνονται όχι μόνο με στατιστικές μεθόδους, αλλά και με γνώμονα τόσο το ιστορικό του όσο και το περιβάλλον στο οποίο ζει ο ασθενής.

Στόχος της παρούσας διπλωματικής εργασίας είναι η αξιοποίηση τεχνικών Μηχανικής Μάθησης για την πρόβλεψη της νόσου Alzheimer, χρησιμοποιώντας τη μαγνητική τομογραφία του εγκεφάλου του εξεταζόμενου. Πρώτο βήμα προς επίτευξη του συγκεκριμένου στόχου αποτέλεσε η κατάτμηση των εικόνων των μαγνητικών τομογραφιών εγκεφάλων. Πιο συγκεκριμένα, για την κατάτμηση μελετήθηκε και υλοποιήθηκε το πλήρως συνελκτικό νευρωνικό δίκτυο FC - DENSENET. Η εκπαίδευση του δικτύου υλοποιήθηκε σε επισημασμένα δεδομένα μαγνητικών τομογραφιών εγκεφάλων που προήλθαν από το διαγωνισμό MICCAI 2012 Multi-Atlas Labelling. Επιχειρήθηκαν τρεις διαφορετικές προσεγγίσεις για την κατάτμηση και πιο συγκεκριμένα, μια απλή εφαρμογή του μοντέλου FC-DenseNet103-Tiramisu, μία που αγνοούσε την κλάση του υποβάθρου και μια που εφάρμοζε βάρη σε κάθε μία από τις κλάσεις ενδιαφέροντος. Η πρώτη μέθοδος απέτυχε να δώσει αξιόπιστα αποτελέσματα καθώς ταξινομούσε όλα τα εικονοστοιχεία στην κλάση του υποβάθρου. Η μέθοδος με τη χρήση βαρών είχε καλύτερη προσαρμογή στα δεδομένα εκπαίδευσης από αυτήν που αγνοούσε την κλάση του υποβάθρου και έτσι επιλέχθηκε ως η κύρια μέθοδος κατάτμησης για το επόμενο στάδιο της διάγνωσης μέσω ταξινόμησης.

Στη συνέχεια, επιχειρήθηκε η χρήση των κατατετμημένων από το FC-DenseNet εικόνων μαγνητικής τομογραφίας εγκεφάλου, για το διαχωρισμό ατόμων σε υγιή και πάσχοντα από τη νευροεκφυλιστική νόσο του Alzheimer. Για το σκοπό αυτό αξιοποιήθηκαν δεδομένα από το ADNI, μια βάση δεδομένων με ιατρικά στοιχεία ασθενών και μη της συγκεκριμένης νόσου. Για την ταξινόμηση των ατόμων επιλέχθηκαν προς μελέτη οι περιοχές της αμυγδαλής, του ενδορινικού φλοιού, του υπόκαμπου και της παραιπποκάμπειας έλικας του εγκεφάλου των ασθενών. Συγκεκριμένα, υπολογίστηκε η αναλογία του μεγέθους (σε εικονοστοιχεία) των περιοχών αυτών προς το μέγεθος του συνολικού εγκεφάλου για κάθε εξεταζόμενο. Για την ταξινόμηση δοκιμάστηκαν δύο μέθοδοι μάθησης για την εξαγωγή αποτελεσμάτων. Η πρώτη μέθοδος βασίστηκε σε διαγράμματα διασποράς των ποσοστών αναλογίας και η δεύτερη υλοποιήθηκε με χρήση ενός ταξινομητή Random Forest, ο οποίος λαμβάνει ως είσοδο τα ποσοστά αναλογίας των παραπάνω περιοχών. Τα αποτελέσματα ανέδειξαν ότι στους ασθενείς με τη νόσο του Alzheimer επηρεάζονται κυρίως τα αριστερά μέρη των περιοχών της αμυγδαλής, του υπόκαμπου και της παραιπποκάμπειας έλικας. Λαμβάνοντας υπόψη τα παραπάνω, συμπεραίνεται ότι το μοντέλο FC-DenseNet απέδωσε αποτελέσματα ακριβούς κατάτμησης εγκεφαλικών εικόνων και σε συνδυασμό με έναν απλό ταξινομητή μηχανικής μάθησης επιτεύχθηκε ο διαχωρισμός πασχόντων και μη από τη νόσο Alzheimer με ποσοστά ακρίβειας πάνω από 95%. Η επιτυχία της παρούσας μεθοδολογίας δημιουργεί προσδοκίες για ανάπτυξη παρόμοιων εφαρμογών για την υποστήριξη της ιατρικής διάγνωσης και για άλλες νευρο-εκφυλιστικές νόσους.

**Λέξεις Κλειδιά:** Τεχνητή Νοημοσύνη, Βαθιά Μάθηση, Κατάτμηση Εγκεφαλικών Εικόνων, FC-DenseNet, Εκπαίδευση, Alzheimer, Ταξινόμηση



# Abstract

The last decades research efforts on the field of Image Analysis have undergone unprecedented growth and development through numerous studies and applications targeting to support and facilitate basic aspects of human every-day life. Towards this goal, most of these applications utilize tools and techniques from the domains of Artificial Intelligence such as Machine Learning and Deep Learning for a variety of fields. Concerning the applications on the Medicine field, current scientific trends suggest that processes of diagnosis and even treatment for various diseases should be achieved not only through statistical methods but also with regard to both patient's medical history and the environment in which he lives.

To this end, this Diploma Thesis presents an application of Machine Learning techniques for the prediction of Alzheimer's disease by using the Magnetic Resonance Imaging (MRI) of the subject's brain. The first step of the methodology was the segmentation of the MRI images. More specifically, the fully convolutional neural network FC - DENSENET was studied and implemented for the segmentation. Network training was implemented on labeled MRI data from the MICCAI 2012 Multi-Atlas Labelling contest. Three different approaches to segmentation were tested, and in particular a simple application of the 103-tiramisu model, one that ignored the class of the background and one that applied weights to each one of the segmentation classes. The first method failed to produce reliable results as it classified all pixels in the background class. The class weights method presented a better-fitting model on the training data than the one ignoring the background class, and thus it was chosen as the basic segmentation method before getting to the next stage of diagnostic classification.

Next, the segmented, via the FC – DENSENET, MRI images were used towards the binary classification of patients and non-patients of the Alzheimer's disease. To this end, data from a medical database – ADNI- were used. For training and classification, the brain areas of the amygdala, the entorhinal cortex, the hippocampus, and the parahippocampal gyrus were selected for study. Specifically, the ratio of the size (in pixels) of these regions to the total brain size for each subject was calculated. Two learning methods were tested for classification. The first method was based on scatter plots and the second was implemented using a Random Forest classifier, which was trained using the proportional size ratios mentioned above. The results indicated that the left regions of amygdala, hippocampus and parahippocampal gyrus were mainly degenerated for the patients with Alzheimer's disease. The combined analysis of the results highlighted that the use of the FC-DenseNet produced accurate brain segmentation results and that in combination with a simple machine learning classifier, the binary classification of Alzheimer's and non-Alzheimer's patients was achieved with accuracy rates above 95%. To sum up, the success of the proposed methodology raises expectations for developing similar applications to support medical diagnosis for other neuro-degenerative diseases as well.

**Key words:** Artificial Intelligence, Deep Learning, Brain MRI Segmentation, FC-DenseNet, Training, Alzheimer, Classify





# Ευχαριστίες

Η παρούσα διπλωματική εκπονήθηκε στα πλαίσια του Προπτυχιακού Προγράμματος Σπουδών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου και συγκεκριμένα, στο εργαστήριο Ευφών Συστημάτων. Πριν όμως από οποιαδήποτε αναφορά στη διαδικασία που ακολουθήθηκε και στα αποτελέσματα που προέκυψαν, θα θέλαμε να ευχαριστήσουμε θερμά τους ανθρώπους οι οποίοι συνέβαλαν στην ολοκλήρωση της εργασίας αυτής.

Θα θέλαμε καταρχάς να ευχαριστήσουμε εκ βαθέων τον επιβλέποντα Καθηγητή κ. Ανδρέα-Γεώργιο Σταφυλοπάτη για την πολύτιμη βοήθεια και καθοδήγησή του, για την ολοκλήρωση της παρούσας εργασίας καθώς και για την δυνατότητα που μας προσέφερε να ασχοληθούμε με ένα τόσο ενδιαφέρον και ελκυστικό ερευνητικό αντικείμενο. Επιπροσθέτως, ευχαριστούμε θερμά τον Ερευνητικό Συνεργάτη κ. Γεώργιο Σιόλα για την καθοδήγησή του ιδιαίτερα στις αρχές αυτής της προσπάθειας. Ακόμη, θα θέλαμε να απευθύνουμε τις ευχαριστίες μας στην κ. Κωνσταντίνα Νικήτα, Καθηγήτρια Ε.Μ.Π. μέλος της τριμελούς επιτροπής εξέτασης, αλλά και στον Αναπληρωτή Καθηγητή κ. Γεώργιο Στάμου ο οποίο συνέβαλε τα μέγιστα στην επιλογή αυτού του αντικειμένου.

Ιδιαίτερη αναφορά αξίζει να σημειωθεί στους Υποψήφιους Διδάκτορες Θάνο Τάγαρη και Μάρα Σδράκα χωρίς την καθοριστική συνεισφορά των οποίων η εργασία δε θα μπορούσε υλοποιηθεί. Τόσο η επιστημονική όσο και η πνευματική τους στήριξη ήταν ιδιαίτερα σημαντικές για εμάς.

Τέλος, σε προσωπικό επίπεδο θα θέλαμε, ο καθένας ξεχωριστά, να ευχαριστήσουμε την οικογένεια και τους φίλους μας, οι οποίοι στάθηκαν δίπλα μας καθ' όλη τη διάρκεια των προπτυχιακών μας σπουδών, και ιδιαίτερα να σημειώσουμε τη συνεισφορά της Χριστίνας Καρακίζη αλλά και του Τάσου Φαντάκη

Αλκιβιάδης-Μάριος Ε. Κούκος

Ιωάννης Θ. Φίλιππας

Αθήνα, 13/11/2018

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>1</b>	<b>Εισαγωγή.....</b>	<b>13</b>
1.1	Τεχνητή Νοημοσύνη.....	13
1.2	Μηχανική μάθηση.....	14
1.2.1	Εισαγωγή.....	14
1.2.2	Μοντέλα Μηχανικής Μάθησης.....	15
1.3	Νευρωνικά δίκτυα.....	16
1.4	Το πρόβλημα της κατάτμησης εικόνας.....	16
1.5	Εφαρμογές της Μηχανικής Μάθησης στην Ιατρική.....	17
1.6	Βαθιά Μάθηση.....	18
<b>2</b>	<b>Θεωρητικό Υπόβαθρο.....</b>	<b>19</b>
2.1	Νευρωνικά Δίκτυα.....	19
2.1.1	Perceptron.....	20
2.1.2	Το Πολυστρωματικό Perceptron (Multi Layer Perceptron-MLP).....	20
2.1.3	Συναρτήσεις Ενεργοποίησης.....	21
2.1.4	Backpropagation (Ανάστροφη Διάδοση).....	23
2.1.5	Υπερπροσαρμογή (Overfitting).....	24
2.1.6	Παράδειγμα Εκπαίδευσης MLP.....	25
2.2	Συνελικτικά Δίκτυα.....	26
2.2.1	Συνελικτικό Επίπεδο (Convolutional Layer).....	27
2.2.2	Επίπεδο Συγκέντρωσης (Pooling Layer).....	29
2.2.3	Επίπεδα Ενεργοποίησης (Activation Layers).....	30
2.2.4	Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer).....	30
2.2.5	Επίπεδο Κανονικοποίησης (Normalization Layer).....	30
2.2.6	Απόσυρση (Dropout).....	31
2.3	Μέτρηση Απόδοσης.....	31
2.3.1	Συναρτήσεις Σφάλματος.....	31
2.3.2	Μετρικές.....	32
2.3.3	Συναρτήσεις Βελτιστοποίησης (Optimization function).....	35
<b>3</b>	<b>Βαθιά Συνελικτικά Μοντέλα.....</b>	<b>41</b>
3.1	Εισαγωγή.....	41
3.2	DenseNet.....	42
3.3	Το Πλήρως Συνδεδεμένο Πυκνό Δίκτυο (FC-DenseNet).....	44
<b>4</b>	<b>Μεθοδολογία και Αποτελέσματα Κατάτμησης.....</b>	<b>47</b>
4.1	Δεδομένα και Προεπεξεργασία.....	47

4.2	Εκπαίδευση .....	47
4.2.1	Αρχιτεκτονική του Μοντέλου .....	47
4.2.2	Εκπαίδευση .....	49
4.3	Αποτελέσματα .....	50
4.3.1	Στατιστικά Αποτελέσματα .....	50
4.3.2	Εμπειρικά Αποτελέσματα.....	54
<b>5</b>	<b>Εφαρμογή της Κατάτμησης σε Ιατρικές Εικόνες και Αποτελέσματα.....</b>	<b>57</b>
5.1	Συλλογή και Προεπεξεργασία Δεδομένων .....	57
5.2	Μεθοδολογία .....	58
5.3	Αποτελέσματα .....	58
5.3.1	Τεστ Συσχέτισης Περιοχών με τη Μεταβλητή Εξόδου .....	59
5.3.2	Εκπαίδευση Ταξινομητών .....	59
5.3.3	Εμπειρική Αξιολόγηση.....	61
<b>6</b>	<b>Συμπεράσματα και Μελλοντικοί Στόχοι .....</b>	<b>71</b>
6.1	Συμπεράσματα .....	71
6.2	Μελλοντικοί Στόχοι .....	72
<b>7</b>	<b>Βιβλιογραφία.....</b>	<b>74</b>



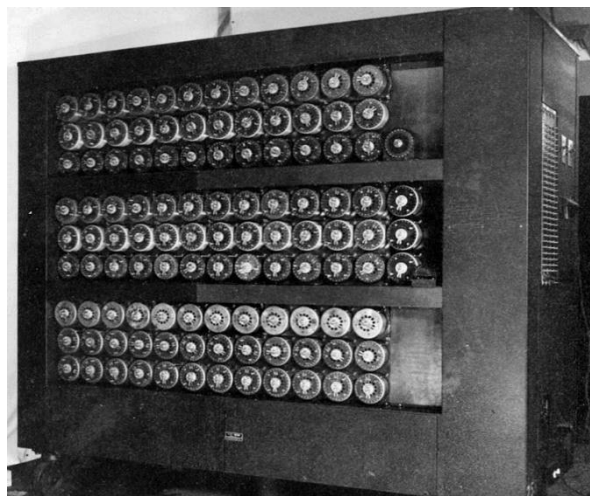
# Κεφάλαιο 1<sup>ο</sup>

## 1 ΕΙΣΑΓΩΓΗ

---

### 1.1 ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

Ανέκαθεν σκοπός της επιστήμης της πληροφορικής ήταν να μοντελοποιεί συστήματα ώστε να τα μελετά πιο εύκολα. Η βιομηχανική «επανάσταση» και οι παράπλευρες συνέπειες της στον τομέα της ηλεκτρονικής είχαν ως αποτέλεσμα μια έξαρση στην οικονομία των μέσων που καλούμαστε να χρησιμοποιούμε, για να αντιμετωπίσουμε τις καθημερινές μας ανάγκες. Από την εφεύρεση του τηλεγράφου από τον Samuel Morse το 1838 [1], έως τη μηχανικό σύστημα Enigma που σχεδίασε ο Alan Turing [2] για να αποκωδικοποιήσει τον κώδικα επικοινωνίας του γερμανικού στρατού, μεσολάβησαν περίπου εκατό χρόνια, στα οποία η ανθρωπότητα είχε εκπαιδευτεί να κατασκευάζει όλο και πιο πολύπλοκα ηλεκτρομηχανικά συστήματα.



*Σχήμα 1.1: Το μηχανικό σύστημα Enigma του Alan Turing*

Παράλληλα, η ραγδαία ανάπτυξη της επιστήμης της θεωρητικής πληροφορικής κατά το ίδιο διάστημα, προσέφερε στην επιστημονική κοινότητα όλο και περισσότερα μοντέλα τα οποία μπορούσαν να εφαρμοστούν για τη μελέτη πολύπλοκων συστημάτων. Η κατασκευή όλο και απλούστερων και γενικευμένων μοντέλων έδινε μέρα τη μέρα την ευκαιρία στους επιστήμονες να σχηματοποιούν καλύτερα το πρόβλημά τους, όσο σύνθετο κι αν ήταν το σύστημα από το οποίο προερχόταν το συγκεκριμένο πρόβλημα.

Σήμερα, η κορύφωση αυτής της παράλληλης προόδου σε όλους τους επιστημονικούς κλάδους, μας δίνει τη δυνατότητα να αντιμετωπίζουμε προβλήματα από όλους τους πιθανούς χώρους, με τρόπο σχεδόν αυτόματο. Η εκπαίδευση των σύνθετων υπολογιστικών συστημάτων

επάνω σε κατάλληλα δεδομένα για τη επίλυση ενός προβλήματος καλείται σήμερα μηχανική μάθηση.

## 1.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

### 1.2.1 Εισαγωγή

Με βάση τον ορισμό που δόθηκε από τον Tom M. Mitchell [3] η μηχανική μάθηση περιγράφεται ως εξής: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία  $E$  ως προς μια κλάση εργασιών  $T$  και ένα μέτρο επίδοσης  $P$ , αν η επίδοσή του σε εργασίες της κλάσης  $T$ , όπως αποτιμάται από το μέτρο  $P$ , βελτιώνεται με την εμπειρία  $E$ ».

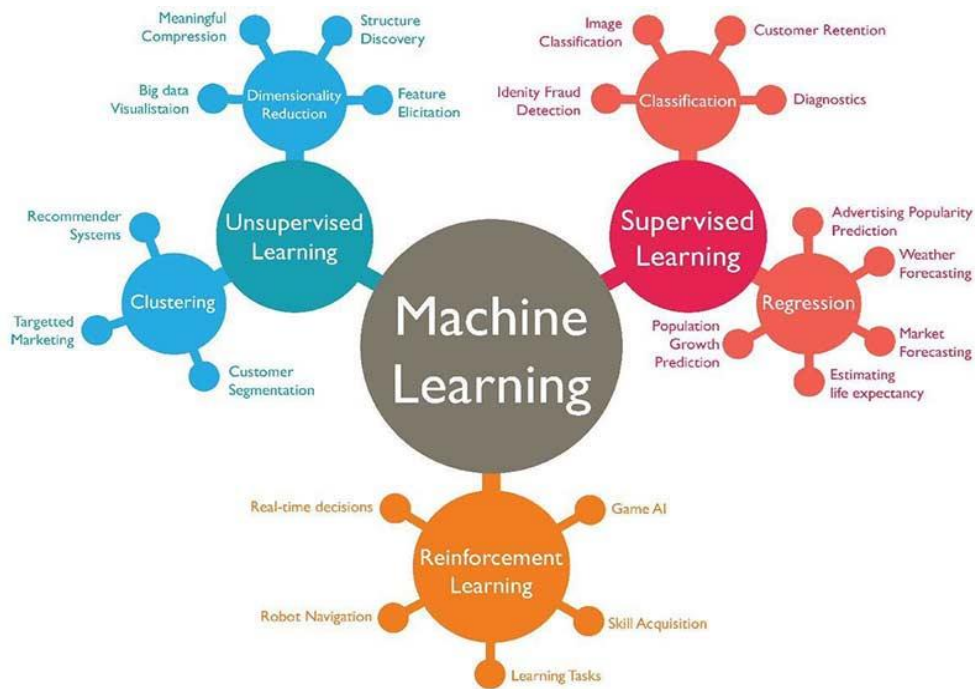
Για παράδειγμα, εάν σκοπός ενός προγράμματος είναι να μάθει να παίζει σκάκι (κλάση εργασιών  $T$ ), μπορεί χρησιμοποιώντας έναν αλγόριθμο μηχανικής μάθησης να παίζει παιχνίδια εξάσκησης ενάντια στον εαυτό του (εμπειρία  $E$ ), και όσο μεγαλύτερο είναι το ποσοστό νικών του εναντίων πραγματικών αντιπάλων (μέτρο επίδοσης  $P$ ), τόσο καλύτερα θα έχει «μάθει» να παίζει.

Σκοπός της μηχανικής μάθησης ως επιστήμης, είναι να επεκτείνει τον παραδοσιακό προγραμματισμό, υπό την έννοια ότι ένα πρόγραμμα μηχανικής μάθησης δεν τροφοδοτεί έναν υπολογιστικό πυρήνα με αλγοριθμικά οργανωμένες εντολές, με άμεσο σκοπό την επίλυση ενός προκαθορισμένου προβλήματος. Αντίθετα, χρησιμοποιώντας κατάλληλα σχεδιασμένους αλγορίθμους, ένα πρόγραμμα μηχανικής μάθησης αλληλεπιδρά με έναν υπολογιστικό πυρήνα ώστε και αυτό να προσαρμοστεί ανάλογα με το πρόβλημα για το οποίο αναζητούμε κάποια βέλτιστη προσέγγιση.

Οι εφαρμογές που ενσωματώνουν τεχνικές μηχανικής μάθησης γράφονται, κυρίως, σε γλώσσες σεναρίου (script languages), όπως είναι η Python και η R. Βέβαια σε εφαρμογές πολύ ειδικού σκοπού, όπως είναι ο σχεδιασμός ηλεκτρονικών παιχνιδιών ή η τηλεκίνηση ρομποτικών συστημάτων, οι γλώσσες χαμηλού επιπέδου, όπως η C/C++ παραμένουν κυρίαρχες.

Η μηχανική μάθηση ως πεδίο εμπεριέχει μεγάλο πλήθος *αλγοριθμικών τεχνικών* και *εκπαιδευτικών στρατηγικών*. Οι τρεις βασικές κατηγορίες μάθησης είναι η **επιβλεπόμενη**, η **μη επιβλεπόμενη** και η ενισχυτική μάθηση [4]. Όσον αφορά τις δύο πρώτες, η βασική τους διαφορά είναι ότι, στην πρώτη περίπτωση θεωρούμε δεδομένη τη γνώση που θέλουμε το σύστημά μας να διδαχθεί, οπότε κατά την εκπαίδευση το τροφοδοτούμε με δεδομένα αλλά και τις επισημάνσεις (labels). Αντίθετα, στη μη επιβλεπόμενη μάθηση, τροφοδοτούμε το σύστημα με δεδομένα εκπαίδευσης, χωρίς όμως να τα χαρακτηρίζουμε ως προς κάποιο γνώρισμά τους. Στη μη επιβλεπόμενη μάθηση, λοιπόν, περιμένουμε από το σύστημά μας να εκμαιεύσει κάποιες πληροφορίες από τα δεδομένα εκπαίδευσης, ενώ στην επιβλεπόμενη φροντίζουμε ώστε το σύστημά μας να «μάθει» να αναγνωρίζει συγκεκριμένα χαρακτηριστικά από τα δεδομένα είσοδο.

Η τρίτη κατηγορία μάθησης, η **ενισχυτική μάθηση** αποτελεί μια σύνθεση των δύο προηγούμενων, μόνο που εισάγει και την έννοια της ανταμοιβής (reward). Τελείως συνοπτικά, το σύστημα προσπαθεί να εξάγει χαρακτηριστικά από μη επισημασμένα δεδομένα (unlabeled data) και, όταν με βάση κάποιο κριτήριο, τα καταφέρνει καλά, τότε επιβραβεύεται με κάποια ανταμοιβή. Αντίστοιχα, όταν δεν τα καταφέρνει, τότε «τιμωρείται».



Σχήμα 1.2: Κατηγορίες ταξινόμησης προβλημάτων μηχανικής μάθησης

### 1.2.2 Μοντέλα Μηχανικής Μάθηση

Πολλά θεωρητικά μοντέλα που σχεδιάστηκαν από τη δεκαετία του 1950 κι έπειτα έχουν υλοποιηθεί τα τελευταία χρόνια κι έχουν προσφέρει χρήσιμα εργαλεία σε ερευνητές κι επαγγελματίες από όλα τα επιστημονικά πεδία. Ενδεικτικά αναφέρουμε τις εξής σημαντικές κατηγορίες συστημάτων:

- Δένδρα Αποφάσεων (Decision Trees)
- Τεχνητά Νευρωνικά Δίκτυα
- Μηχανές Διανυσμάτων Υποστήριξης
- Δίκτυα Bayes
- Γενετικοί Αλγόριθμοι
- Επαγωγικός Λογικός Προγραμματισμός
- Συστήματα Συσταδοποίησης (Clustering)
- Αναπαραστατική Μάθηση

Στο πλαίσιο της παρούσας εργασίας, υλοποιήθηκε ένα πλήρως Συνελικτικό Τεχνητό Νευρωνικό Δίκτυο βαθιάς Μάθησης, γι' αυτό θα δοθεί έμφαση στο θεωρητικό υπόβαθρο μοντέλων αυτού του τύπου.

### 1.3 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

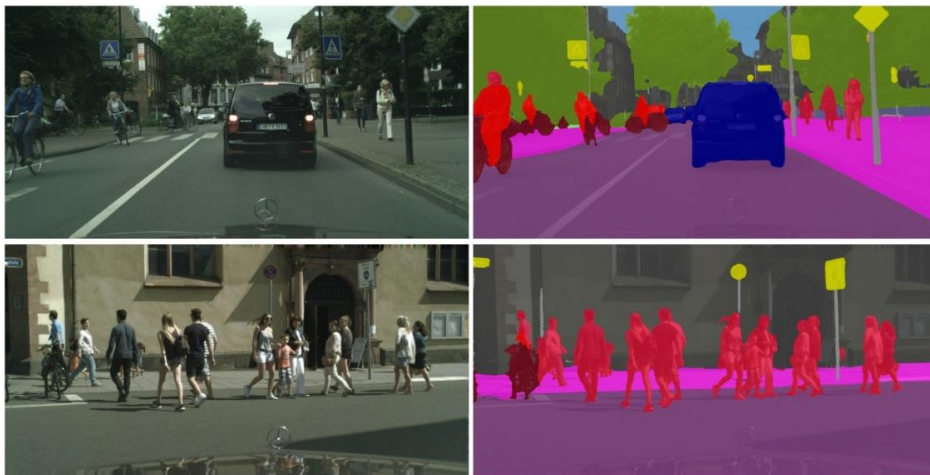
Τα νευρωνικά δίκτυα είναι σύνθετες υπολογιστικές δομές, που μπορούν να μοντελοποιούν εξαιρετικά πολλές από τις αλγοριθμικές τεχνικές της μηχανικής μάθησης. Ένα μεγάλο πλεονέκτημά τους είναι ότι δίνουν τεράστια ευχέρεια κινήσεων στον προγραμματιστή, ώστε να μπορεί να προσθέτει ή να αφαιρεί από αυτά *στρώματα* ή *παραμέτρους* μέχρι να πετύχει την *υπολογιστική ακρίβεια* που επιθυμεί. Η προσθήκη νέων επιπέδων ή ο εμπλουτισμός των ήδη υπάρχοντων σε ένα τεχνητό νευρωνικό δίκτυο απαιτεί πειραματισμό, καθώς η υπολογιστική πολυπλοκότητα της δομής μπορεί να αυξηθεί εκθετικά.

Παρά την πολυπλοκότητά τους όμως, η αρχή λειτουργίας των νευρωνικών δικτύων είναι εξαιρετικά κομψή καθώς δανείζεται στοιχεία από τη λειτουργία των εγκεφαλικών νευρώνων, όπως άλλωστε μαρτυρά και η ονομασία τους. Για να καταστεί ευκολότερα αντιληπτή η λειτουργία τους, θα αναλυθεί εκτενέστερα στο επόμενο κεφάλαιο της παρούσας διπλωματικής εργασίας, η λειτουργία του νευρώνα *perceptron*, ο οποίος είναι και ένα από τα αρχαιότερα μοντέλα μηχανικής μάθησης.

### 1.4 ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ

Το πρόβλημα της κατάτμησης εικόνας [5] είναι ένα από τα σημαντικότερα στην περιοχή της υπολογιστικής όρασης, αφού ο διαχωρισμός των *περιοχών ενδιαφέροντος* (regions of interest) μιας εικόνας είναι από τα βασικά βήματα για την εξόρυξη των πληροφοριών που μπορεί να αναπαριστά. Ακριβώς επειδή το πρόβλημα είναι δύσκολο να καθοριστεί επακριβώς για οποιοδήποτε τύπο εικόνας, προτείνονται πολλοί τρόποι προσέγγισής του στη βιβλιογραφία [6]. Γενικά, μπορεί να οριστεί ως η διαδικασία χωρισμού της εικόνας σε ανεξάρτητες περιοχές, οι οποίες εμφανίζουν μια ομοιογένεια ως προς κάποιο χαρακτηριστικό (για παράδειγμα, τη φωτεινότητα). Η ένωση όλων των περιοχών πρέπει να έχει ως αποτέλεσμα την αρχική εικόνα.

Το πρόβλημα της κατάτμησης δεν είναι μονοσήμαντο. Ανάλογα την εφαρμογή, μια εικόνα μπορεί να κατατμηθεί με περισσότερους από έναν τρόπους. Παραδείγματος χάριν, συχνά ζητείται ο διαχωρισμός του κυρίως περιεχομένου της εικόνας από το παρασκήνιο, οπότε οι περιοχές ενδιαφέροντος είναι δύο (background detection). Σε άλλες εφαρμογές αυτό δεν είναι αρκετό και απαιτείται η κατάτμηση της εικόνας σε ακόμα περισσότερα τμήματα.



Σχήμα 1.3: Παράδειγμα κατάτμησης εικόνων



Ανεξάρτητα, όμως, από το περιεχόμενο της εφαρμογής, για να θεωρείται επιτυχημένη η κατάτμηση θα πρέπει το αποτέλεσμά της να πληροί τα ακόλουθα κριτήρια:

- Οι περιοχές στις οποίες έχει κατατμηθεί η εικόνα να παρουσιάζουν μια ομοιογένεια ως προς κάποιο μετρήσιμο χαρακτηριστικό
- Γειτονικές περιοχές πρέπει να παρουσιάζουν αισθητές διαφορές
- Τα όρια των περιοχών πρέπει να είναι σαφώς ορισμένα και ευδιάκριτα
- Το εσωτερικό των περιοχών να μην περιέχει οπές, η άλλου είδους ανωμαλίες

Ανάλογα με το περιεχόμενο που θέλουμε να ξεχωρίσει το σύστημά μας, πρέπει να αποφασίσουμε σε ποια από τα παραπάνω κριτήρια θα επικεντρωθούμε αλγοριθμικά.

## 1.5 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ

Ήδη εδώ και μια δεκαετία η μηχανική μάθηση έχει αρχίσει να βρίσκει εφαρμογές σε όλες τις βαριές βιομηχανίες της παγκόσμιας οικονομίας. Από τα αυτόματα λογισμικά συστήματα (softbots) που καταφέρνουν να πραγματοποιούν εμπορικές συναλλαγές στο Χρηματιστήριο Αξιών [7], έως τα νευρωνικά δίκτυα που χρησιμοποιούνται για τη βελτιστοποίηση των παραμέτρων μιας πολεοδομικής εργασίας, είναι ολοφάνερο ότι η μηχανική μάθηση διαθέτει εργαλεία που μπορούν να λύσουν τα χέρια επαγγελματιών που αναζητούν αυτοματοποιημένες λύσεις στα προβλήματά τους.

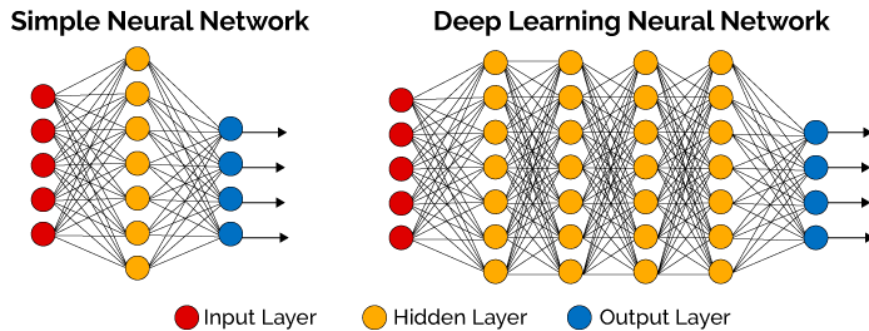
Στον τομέα της Ιατρικής Έρευνας, η μηχανική μάθηση έχει ήδη διαπιστεύσει την αξιοπιστία της όσον αφορά την *έγκαιρη* και *έγκυρη* γνωμάτευση της κατάστασης της υγείας ενός ασθενούς, η οποία δίνει την ευκαιρία στον επικεφαλής ιατρό να παρέχει μίαν όσο το δυνατό πλήρη διάγνωση [8]. Παράλληλα, καθημερινά αναπτύσσονται εφαρμογές για τη συνεχή παρακολούθηση ενός ασθενή, εφαρμογές οι οποίες δε χρειάζονται ούτε συνεχή επίβλεψη αλλά ούτε και ιδιαίτερη διάδραση με τον υπό παρακολούθηση ασθενή. Αυτό προσφέρει τρομερές δυνατότητες στον ιατρό, καθώς η διαδικασία της *εξέτασης* απλοποιείται σημαντικά, αν και επιβαρύνεται με παραπάνω *θόρυβο*, καθώς ο ασθενής συνήθως χρησιμοποιεί αυτές τις εφαρμογές στον προσωπικό του χώρο και όχι σε κάποιο ιατρικό εργαστήριο [9]

Τα τελευταία χρόνια διαπιστώθηκε ότι με χρήση συνελκτικών νευρωνικών δικτύων, μπορούσε να εξαχθεί σημαντικός όγκος πληροφορίας για νευροεκφυλιστικές ασθένειες όπως είναι η νόσος του Alzheimer ή του Parkinson. Πρόκειται για δύο νόσους οι οποίες καταλήγουν σε τελείως διαφορετικού τύπου συμπτώματα, καθώς η πρώτη, στα τελικά της στάδια, επιφέρει πλήρη άνοια, ενώ η δεύτερη αποκλείει την άμεση επικοινωνία των μυών με το νευρικό σύστημα, έτσι ώστε ο ασθενής να μην μπορεί να ελέγξει την κίνησή του [10].

Παρά ταύτα, οι ερευνητές γνωρίζουν ότι οι δύο νόσοι έχουν πιθανώς κοινή καταγωγή, καθώς καταστρέφουν σταδιακά ολόκληρες νευροεγκεφαλικές δομές, έως ότου καταστούν μη λειτουργικές. Η έγκαιρη διάγνωση τέτοιων παθήσεων κρίνεται απολύτως απαραίτητη, αφού στα πρώτα στάδια τους είναι πολύ πιο εύκολα αντιμετωπίσιμες.

Το γεγονός αυτό μας έδωσε το κίνητρο να εφαρμόσουμε τα συμπεράσματά μας για το κεντρικό μοντέλο αυτής της εργασίας, στην αναγνώριση προτύπων που καταδεικνύουν την κατάσταση ενός πιθανού ασθενή με Alzheimer, χρησιμοποιώντας ως δεδομένο αποκλειστικά την *απεικόνιση* της μαγνητικής του τομογραφικής εξέτασης (brain MRi) [11].

## 1.6 ΒΑΘΙΑ ΜΑΘΗΣΗ



Σχήμα 1.4. Παράδειγμα απλού νευρωνικού δικτύου(αριστερά) και βαθιού νευρωνικού δικτύου(δεξιά).

Μια σημαντική καινοτομία που εισήγαγε η Μηχανική Μάθηση στην Επιστήμη των Υπολογιστών είναι η **Βαθιά Μάθηση** (Deep Learning) [12]. Ο όρος εισήχθη το 1986 από τη Rina Dechter και χρησιμοποιείται για να περιγράψει νευρωνικά μοντέλα τα οποία είναι μεγάλα σε βάθος, δηλαδή αποτελούνται από έναν μεγάλο αριθμό στρωμάτων στοιβαγμένων το ένα μετά το άλλο. Πολύ συχνά τα μοντέλα αυτά είναι **Συνελικτικά** δηλαδή εκμεταλλεύονται τις ιδιότητες της συνέλιξης συναρτήσεων για να επεξεργαστούν δεδομένα που εκτείνονται σε δύο ή περισσότερες διαστάσεις και, άρα, η απευθείας επεξεργασία τους είναι υπολογιστικά χρονοβόρα.

Ακόμη, τις τελευταίες δεκαετίες του περασμένου αιώνα, έγινε σημαντική πρόοδος στην Επιστήμη της Πληροφορικής, και έδωσε τη δυνατότητα στους ερευνητές να υλοποιήσουν τις τεχνικές της βαθιάς μάθησης. Τα διαθέσιμα δεδομένα τα οποία αυξάνονται συνεχώς αλλά και η ανάπτυξη συστημάτων **παράλληλης επεξεργασίας δεδομένων**, που οδήγησαν στις σύγχρονες πανίσχυρες μονάδες γραφικής επεξεργασίας (*GPUs*), έχουν παίξει καθοριστικό ρόλο στην εξέλιξη της.

# Κεφάλαιο 2<sup>ο</sup>

## 2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

---

### 2.1 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Η δημιουργία μηχανών που θα μπορούν να σκέφτονται, να εξάγουν λογικά συμπεράσματα και να επιτελούν, περισσότερο ή λιγότερο, απλά καθήκοντα αποτελούσε διακαή πόθο των ερευνητών τις τελευταίες δεκαετίες, αλλά και διάφορων επιστημόνων ανά τους αιώνες. Οι προσπάθειες που είχαν ξεκινήσει να γίνονται συστηματικά από τα τέλη του περασμένου αιώνα κατεδείκνυαν ότι ο παραδοσιακός προγραμματισμός αντιμετώπιζε πολύ σοβαρά προβλήματα απέναντι σε αυτή την πρόκληση. Η σχεδίαση *συστημάτων νοημοσύνης* που στηρίζονταν αποκλειστικά σε κανόνες, φαινόταν αρκετή όταν τα συστήματα προορίζονταν για την επίλυση σαφώς ορισμένων προβλημάτων (για παράδειγμα να βρεις τη βέλτιστη κίνηση στο σκάκι δεδομένης μιας σκακιέρας, DEEPBLUE [13]), όταν όμως ο χώρος κατάστασης γινόταν σημαντικά μεγαλύτερος, όπως είναι το παραδοσιακό κινεζικό παιχνίδι Go, ήταν προφανές ότι χρειαζόταν διαφορετική προσέγγιση του ζητήματος.

Για να ξεπεραστεί αυτό, δοκιμάστηκαν νέες προσεγγίσεις, οι οποίες δεν αποσκοπούσαν στην απομνημόνευση κανόνων, τους οποίους το σύστημα θα εφάρμοζε επάνω στα δεδομένα εισόδου. Αντιθέτως, στόχευαν στο σχεδιασμό συστημάτων που θα μπορούσαν να εξάγουν αυτόματα κανόνες, μετά την εκπαίδευσή τους πάνω σε δεδομένα που είχαν υποστεί σημαντική προεπεξεργασία. Η διαδικασία αυτή ονομάστηκε μηχανική μάθηση (*machine learning*) και τα αποτελέσματά της οδήγησαν στην έκρηξη που γνωρίζεται το πεδίο της τεχνητής νοημοσύνης στις μέρες μας.

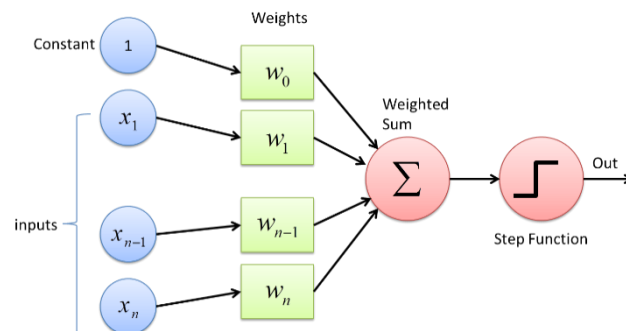
Τα νευρωνικά δίκτυα αποτελούν συστήματα που υλοποιούν τη μηχανική μάθηση. Ονομάστηκαν έτσι γιατί η δομή τους προσομοιάζει τον τρόπο με τον οποίο επικοινωνούν οι νευρώνες στον ανθρώπινο, και όχι μόνο, εγκέφαλο. Έχουν σημειώσει εξαιρετικά αποτελέσματα και για αυτό έχουν κυριαρχήσει σε πεδία όπως η επεξεργασία και αναγνώριση εικόνας, η επεξεργασία και αναγνώριση φωνής και στην αναγνώριση προτύπων γενικότερα. Για να γίνει κατανοητή η λειτουργία ενός νευρωνικού δικτύου θα δοθεί ένα παράδειγμα από την αναγνώριση εικόνων. Ένα νευρωνικό δίκτυο χρησιμοποιώντας εικόνες που απεικονίζουν μία γάτα ή έναν σκύλο και έχουν χαρακτηριστεί ανάλογα, μπορεί να «μάθει» να αναγνωρίζει ποιο από τα δύο ζώα αναπαρίσταται. Αυτό συμβαίνει χωρίς αρχικά να έχει καμία γνώση για τα χαρακτηριστικά των ζώων, όπως μέγεθος, τρίχωμα, μύτη. Κατά τη διάρκεια όμως της εκπαίδευσης δημιουργεί δικούς του κανόνες οι οποίοι προκύπτουν από τα δεδομένα στα οποία εκπαιδεύεται.

Σήμερα υπάρχει πληθώρα επιλογών όσον αφορά τη σχεδίαση ενός νευρωνικού μοντέλου. Ιδιαίτερα για τη γλώσσα Python, υπάρχουν ολόκληρες βιβλιοθήκες τις οποίες μπορεί κανείς να εκμεταλλευτεί προκειμένου να υλοποιήσει ένα δίκτυο. Στο πλαίσιο αυτής της εργασίας χρησιμοποιήθηκε το Keras [14], μια **διεπαφή προγραμματισμού εφαρμογών** (*application programming interface*), η οποία λειτουργεί ως βιβλιοθήκη υψηλού επιπέδου, αφού ενσωματώνει το **tensorflow**, μια από τις σημαντικότερες δομές ανάπτυξης λογισμικού για εφαρμογές μηχανικής μάθησης, το CNTK, μια εργαλειοθήκη για βαθιά μάθηση σχεδιασμένο από τη Microsoft, καθώς και το Theano, μια βιβλιοθήκη σχεδιασμένη για το χειρισμό διανυσμάτων μεγάλων διαστάσεων και το σχεδιασμό σύνθετων μαθηματικών δομών. Για τις δικές μας ανάγκες χρησιμοποιήθηκε το tensorflow.

Σε αυτό το κεφάλαιο θα αναλύσουμε μερικά πολύ σημαντικά εργαλεία που έχει προσφέρει η επιστήμη της Μηχανικής Μάθησης στην ανάλυση και επεξεργασία εικόνας, τα **Συνελικτικά Νευρωνικά Δίκτυα** (Convolutional Neural Networks). Θα μας βοηθήσει σε αυτό η ανάλυση πιο βασικών μοντέλων που χρησιμοποιούν την εμπρόσθια τροφοδότηση (feed forward) και την οπισθοδρόμηση (backpropagation), όπως είναι τα ακολουθιακά μοντέλα.

### 2.1.1 Perceptron

Ο **Perceptron**, όπως δηλώνει και το όνομά του (*perceiving and recognizing automaton*), είναι ένας νευρώνας που αντιλαμβάνεται και αναγνωρίζει με αυτόματο τρόπο. Εφευρέθηκε το 1957 στο Αεροναυτικό Εργαστήριο του Cornell από τον Frank Rosenblatt [15] και είναι από τους αρχαιότερους αλγορίθμους μηχανικής μάθησης. Πρόκειται ουσιαστικά για έναν δυαδικό ταξινομητή που αποτελείται από έναν νευρώνα, με είσοδο ένα διάνυσμα και έξοδο έναν δυαδικό αριθμό. Η είσοδος  $\mathbf{x}$  του νευρώνα πολλαπλασιάζεται με έναν πίνακα βαρών  $\mathbf{W}$  και προστίθεται με έναν αριθμό που ονομάζεται πόλωση  $\mathbf{b}$  (bias). Αν το αποτέλεσμα της πράξης είναι μεγαλύτερο από κάποιο κατώφλι, τότε η έξοδος του συστήματος είναι 1 διαφορετικά είναι 0.

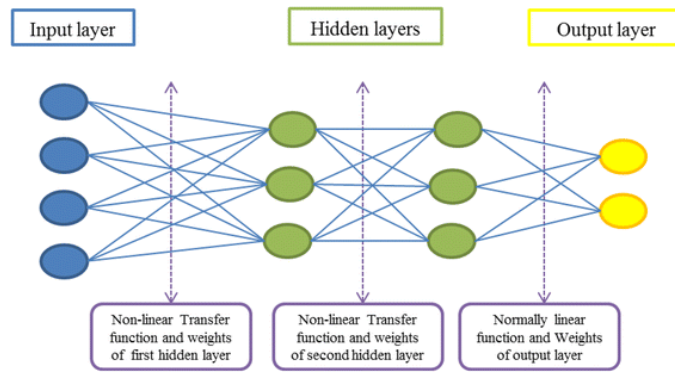


Σχήμα 2.1 Το μοντέλο perceptron

### 2.1.2 Το Πολυστρωματικό Perceptron (Multi Layer Perceptron-MLP)

Έχοντας στο μυαλό μας το μοντέλο perceptron είμαστε έτοιμοι να περιγράψουμε ένα πλήρες *ακολουθιακό* μοντέλο νευρωνικού δικτύου, το πολυστρωματικό perceptron (MultiLayer Perceptron). Είναι ένα δίκτυο το οποίο μπορεί να είναι αρκετά *βαθύ* με την έννοια ότι δεν υπάρχει περιορισμός στο πόσα *κρυφά* στρώματα θα διαθέτει πέρα από τα σίγουρα στρώματα της εισόδου και της εξόδου. Τα κρυφά αυτά στρώματα στο εσωτερικό του είναι αυτά που προσφέρουν το βάθος του. Οι νευρώνες μεταξύ δυο διαδοχικών κρυφών επιπέδων συνδέονται πλήρως μεταξύ τους. Συγκεκριμένα, η έξοδος κάθε νευρώνα ενός επιπέδου δίνεται ως είσοδος σε κάθε νευρώνα του αμέσως επόμενου επιπέδου. Συνήθως δεν επιτρέπονται συνδέσεις μεταξύ νευρώνων που δεν ανήκουν σε διαδοχικά στρώματα. Στη γενική περίπτωση τα πολλά κρυφά επίπεδα αυξάνουν το υπολογιστικό κόστος, συνεπώς ένα δίκτυο πρέπει να είναι τόσο βαθύ, όσο πολύπλοκο είναι το σύστημα που θέλουμε να μοντελοποιήσουμε. Η έξοδος του  $i$ -οστού επιπέδου προκύπτει από το εσωτερικό γινόμενο του διανύσματος εξόδου του προηγούμενου επιπέδου  $\mathbf{y}_{i-1}$  με τον πίνακα των βαρών  $\mathbf{W}_i$  προσθέτοντας το διάνυσμα πόλωσης  $\mathbf{b}_i$  και δίνοντας το αποτέλεσμα της πράξης αυτής ως είσοδος σε μια μη γραμμική συνάρτηση ενεργοποίησης  $f$ .

$$y_i = f(W_i y_{i-1} + b_i)$$



Σχήμα 2.2 Ένα MLP με 3 κρυφά επίπεδα. Αυτό το δίκτυο είναι Πλήρως Συνδεδεμένο Δίκτυο (Fully Connected Network (FCN)), μιας και κάθε νευρώνας είναι συνδεδεμένος με όλους τους νευρώνες από το προηγούμενο επίπεδο.

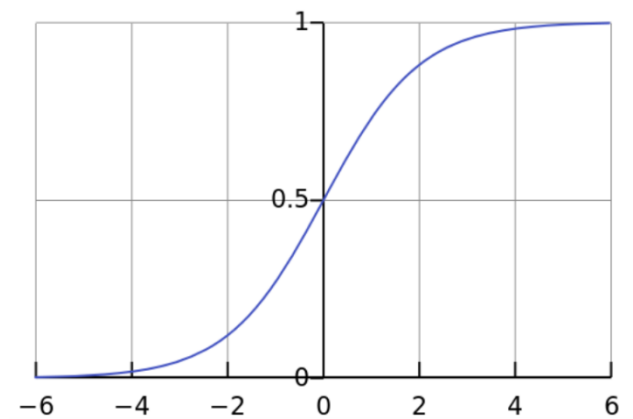
Η διαδικασία που περιγράφηκε ονομάζεται εμπρόσθια διάδοση (Forward Propagation) [16]. Η επιλογή της συνάρτησης ενεργοποίησης είναι πολύ σημαντική για την αποδοτικότητα του ενός νευρωνικού δικτύου. Ανάλογα με τη φύση του προβλήματος, επιλέγεται και κατάλληλη συνάρτηση ενεργοποίησης.

### 2.1.3 Συναρτήσεις Ενεργοποίησης

#### Σιγμοειδής συνάρτηση

Παρακάτω φαίνεται η μαθηματική εξίσωση που περιγράφει τη σιγμοειδή συνάρτηση, καθώς και η γραφική της αναπαράσταση:

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



Σχήμα 2.3 Γραφική αναπαράσταση της σιγμοειδούς συνάρτησης.

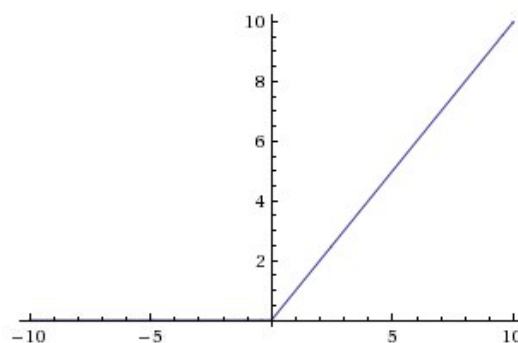
Πρόκειται για μία συνεχή και μη γραμμική συνάρτηση. Όταν εφαρμοστεί το σταθμισμένο άθροισμα των βαρών ενός μοντέλου στη θέση του  $x$ , οι τιμές τους κανονικοποιούνται στο διάστημα  $(0,1)$ . Πιο συγκεκριμένα, οι υψηλές σε απόλυτη τιμή αρνητικές τιμές θα πάρουν την τιμή 0, ενώ οι αντίστοιχες θετικές την τιμή 1. Ωστόσο, όπως είναι εμφανές, αλλαγές στην είσοδο  $x$ , τείνουν να έχουν πολύ μικρό αντίκτυπο στις αλλαγές του  $y$ . Η παράγωγος σε αυτήν την περιοχή θα είναι μικρή, με αποτέλεσμα να δημιουργεί ένα πρόβλημα που ονομάζεται “*vanishing gradient*” [17]. Στις ακραίες περιοχές της σιγμοειδούς συνάρτησης, η μερική παράγωγος έχει χαμηλή τιμή ή αλλιώς έχει εξαφανιστεί (*vanished*), κι έτσι το δίκτυο αρνείται να μάθει περαιτέρω ή βελτιώνεται με δραματικά αργούς ρυθμούς. Παρά το παραπάνω πρόβλημα, παραμένει μία από τις πιο δημοφιλείς συναρτήσεις σε προβλήματα ταξινόμησης.

### Ανορθωμένη Γραμμική Μονάδα (ReLU)

Η Ανορθωμένη Γραμμική Μονάδα (Rectified Linear Unit – ReLU), είναι μία από τις πλέον δημοφιλέστερες συναρτήσεις ενεργοποίησης, της οποίας ο μαθηματικός ορισμός φαίνεται παρακάτω:

$$A(x) = \max(0, x)$$

Είναι εντυπωσιακό πως μια τόσο απλή συνάρτηση καταφέρνει να εισάγει σε ένα μοντέλο την ιδιότητα της μη γραμμικότητας. Είναι φθηνή υπολογιστικά και κατ’ επέκταση, το μοντέλο χρειάζεται λιγότερο χρόνο για να εκπαιδευτεί. Σε αντίθεση με τη σιγμοειδή συνάρτηση, αποφεύγεται εδώ το φαινόμενο του *vanishing gradient*, καθώς η παράγωγος είναι πάντα σταθερή και ίση με 1, για θετική είσοδο, με αποτέλεσμα να συγκλίνει γρηγορότερα. Ακόμη, η ReLU ενεργοποιείται αραιά, δηλαδή από τη στιγμή που παίρνει την τιμή μηδέν για κάθε αρνητική είσοδο, είναι αρκετά πιθανό κάποια συγκεκριμένη μονάδα του δικτύου να μην ενεργοποιηθεί. Η αραιότητα (*sparsity*) οδηγεί συνήθως σε συνοπτικά μοντέλα, τα οποία έχουν μεγαλύτερη προβλεπτική ισχύ και λιγότερη υπερπροσαρμογή. Σε ένα αραιό δίκτυο, είναι αρκετά πιο πιθανό οι νευρώνες του να είναι πραγματικά επεξεργαστικά σημαντικές πτυχές του προβλήματος. Το μειονέκτημα όμως του γεγονότος ότι όλες οι αρνητικές εισοδοί δίνουν μηδενική έξοδο είναι ένα πρόβλημα που ονομάζεται «ετοιμοθάνατη» (“*dying*”) ReLU [18]. Ένας νευρώνας θεωρείται νεκρός όταν παίρνει συνεχώς αρνητικές τιμές και κατά συνέπεια έχει πάντοτε μηδενική έξοδο. Επειδή η κλίση της ReLU είναι επίσης μηδενική στον αρνητικό ημίξονα, εφόσον ένας νευρώνας πάρει αρνητική τιμή, δε θα μπορέσει ποτέ να αλλάξει πρόσημο στην τιμή του. Έτσι, είναι πιθανό ένα μέρος του δικτύου εν τέλει να μη χρησιμεύει σε τίποτα. Το φαινόμενο αυτό αντιμετωπίζεται, σε κάποιο βαθμό, χρησιμοποιώντας χαμηλές τιμές για το ρυθμό απόδοσης (*learning rate*), ο οποίος θα εξηγηθεί αναλυτικά στην παράγραφο 2.4.



Σχήμα 2.4 Η γραφική παράσταση της συνάρτησης ReLU.

## Softmax

Σε προβλήματα ταξινόμησης με περισσότερες από δύο κλάσεις, δεν είναι χρήσιμη καμία από τις παραπάνω μεθόδους. Για αυτό το σκοπό χρησιμοποιείται μια άλλη συνάρτηση, η κανονικοποιημένη εκθετική συνάρτηση (Softmax), η οποία συμπιέζει την έξοδο κάθε μονάδας στο διάστημα μεταξύ 0 και 1, όπως ακριβώς και η σιγμοειδής. Η διαφορά τους όμως είναι ότι κάθε έξοδος διαιρείται με το άθροισμα όλων των εξόδων, ούτως ώστε το συνολικό άθροισμα τους να είναι ίσο με 1. Πρακτικά, η έξοδος της Softmax είναι ισοδύναμη με μια κατανομή πιθανοτήτων στις κλάσεις, δείχνει δηλαδή την πιθανότητα κάθε κλάσης να είναι αληθής. Ο μαθηματικός της τύπος φαίνεται παρακάτω, όπου το  $z$  αντιπροσωπεύει το διάνυσμα εισόδου και  $K$  οι συνολικές κλάσεις του μοντέλου:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

Για να γίνει πιο κατανοητή η λειτουργία της Softmax δίνεται ένα απλό παράδειγμα. Εάν η είσοδος ενός επιπέδου του μοντέλου είναι [1.2, 0.9, 0.4], η έξοδός του θα είναι [0.46, 0.34, 0.20] όπως φαίνεται και στο παρακάτω σχήμα.



Σχήμα 2.5 Παράδειγμα της συνάρτησης Softmax

### 2.1.4 Backpropagation (Ανάστροφη Διάδοση)

Ο αλγόριθμος της ανάστροφης διάδοσης παρουσιάστηκε αρχικά τη δεκαετία του 70', αλλά η σπουδαιότητά του δεν είχε εκτιμηθεί μέχρι το 1986 [19]. Πειράματα τότε έδειξαν ότι ο αλγόριθμος αυτός λειτουργούσε πολύ πιο γρήγορα από τις προηγούμενες προσεγγίσεις, ανοίγοντας έτσι νέους ορίζοντες στον κλάδο της μηχανικής μάθησης, και των νευρωνικών δικτύων ειδικότερα.

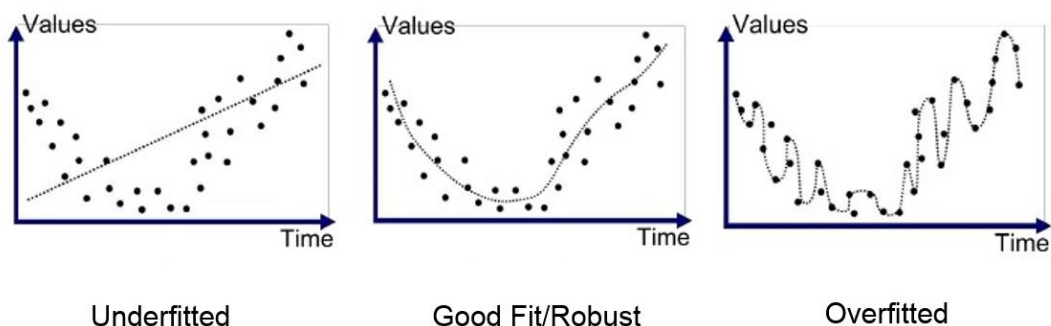
Η ανάστροφη οπισθοδρόμηση είναι μια μέθοδος η οποία υπολογίζει τις μερικές παραγώγους της συνάρτησης απώλειας ως προς τα βάρη και τις πολώσεις τους. Σκοπός της είναι να βελτιστοποιήσει τις παραμέτρους του δικτύου, ώστε το νευρωνικό δίκτυο να μπορεί να μάθει σωστά να αντιστοιχίζει τυχαίες εισόδους με τις αντίστοιχες εξόδους τους, που θα αναλυθεί περισσότερο στην παράγραφο 2.4. Για να μπορέσει να λειτουργήσει θα πρέπει να ικανοποιούνται δύο συνθήκες σχετικά με τη μορφή της συνάρτησης απώλειας. Η πρώτη είναι να μπορεί να εκφραστεί ως μέσος όρος των συναρτήσεων απώλειας κάθε επιμέρους δεδομένου εισόδου. Ο λόγος που απαιτείται αυτό είναι ότι ο back propagation υπολογίζει τις μερικές παραγώγους ως προς μόνο ένα παράδειγμα εισόδου, οπότε οι τελικές τιμές πρέπει να ανακατασκευαστούν παίρνοντας τον μέσο όρο στα παραδείγματα εισόδου. Η δεύτερη συνθήκη είναι ότι η συνάρτηση απώλειας μπορεί να γραφτεί ως συνάρτηση των εξόδων του δικτύου.

Πιο αναλυτικά τώρα η λειτουργία του αλγορίθμου είναι η εξής. Αρχικά μέσω της εμπρόσθιας διάδοσης υπολογίζεται η έξοδος του δικτύου. Αμέσως μετά υπολογίζεται το σφάλμα των νευρώνων στην έξοδο του δικτύου, ως η μερική παράγωγος της συνάρτησης απώλειας ως προς την είσοδό της. Στη συνέχεια, με δεδομένο το σφάλμα της εξόδου, υπολογίζονται τα σφάλματα σε όλα τα προηγούμενα επίπεδα μέχρι το επίπεδο εισόδου, χρησιμοποιώντας τον κανόνα της αλυσίδας.

## 2.1.5 Υπερπροσαρμογή (Overfitting)

Για να είναι δυνατή η εκπαίδευση ενός νευρωνικού δικτύου, είναι απαραίτητη η χρήση ενός συνόλου δεδομένων. Σε ένα πρόβλημα ταξινόμησης τα δεδομένα αυτά αποτελούνται από κάποια διανύσματα εισόδου και ένα διάνυσμα απάντησης, ή απλούστερα στόχο. Για να υπάρχει μια αντικειμενική εικόνα της απόδοσής του, καθώς και για να μπορούν να αντιμετωπιστούν πιθανά προβλήματα, όπως η υπερπροσαρμογή που παρουσιάζεται παρακάτω, το διαθέσιμο σύνολο δεδομένων χωρίζεται σε τρία υποσύνολα. Το πρώτο είναι το σύνολο εκπαίδευσης (**train set**), το οποίο αφορά στο σύνολο που χρησιμοποιείται για την εκπαίδευση του μοντέλου. Με βάση τα δεδομένα του, το μοντέλο παράγει μια έξοδο η οποία συγκρίνεται με τον στόχο, ώστε να γίνει η ενημέρωση των παραμέτρων του δικτύου χρησιμοποιώντας κάποιον αλγόριθμο μάθησης. Αφού ολοκληρωθεί αυτή η διαδικασία, η οποία ονομάζεται **εποχή**, συνεχίζει να επαναλαμβάνεται για έναν καθορισμένο αριθμό επαναλήψεων (δηλαδή εποχών). Το δεύτερο υποσύνολο ονομάζεται σύνολο δοκιμής (**test set**). Πρόκειται για ένα σύνολο ανεξάρτητο του συνόλου εκπαίδευσης, το οποίο όμως ακολουθεί ακριβώς την ίδια κατατομή πιθανότητας με αυτό. Σκοπός του είναι κατά κύριο λόγο η αξιολόγηση του μοντέλου μετά το πέρας της εκπαίδευσης, ώστε να διαπιστωθεί αν το μοντέλο έχει αποκτήσει την ικανότητα της γενίκευσης. Τέλος, υπάρχει και το σύνολο επαλήθευσης (**validation set**) το οποίο όμως δε χρησιμοποιείται πάντοτε. Το υποσύνολο αυτό χρησιμοποιείται για να παρέχει μια αμερόληπτη αξιολόγηση του μοντέλου που προσαρμόζει το σύνολο εκπαίδευσης, ενώ ρυθμίζει τις υπερπαραμέτρους του μοντέλου.

Το φαινόμενο της υπερ-προσαρμογής (overfitting) αποτελεί ένα από τα πιο σημαντικά προβλήματα στην εκπαίδευση ενός νευρωνικού δικτύου. Υπερπροσαρμογή παρατηρείται όταν ένα μοντέλο μαθαίνει τις λεπτομέρειες και το θόρυβο του συνόλου δεδομένων σε τέτοιο βαθμό, ώστε να επηρεάζεται αρνητικά η απόδοσή του σε νέα δεδομένα. Αυτό σημαίνει ότι ο θόρυβος και οι τυχαίες διακυμάνσεις του συνόλου δεδομένων συλλέγονται και μαθαίνονται ως έννοιες για το μοντέλο. Το πρόβλημα είναι ότι αυτές οι έννοιες δεν εφαρμόζονται σε νέα δεδομένα και επηρεάζεται αρνητικά η ικανότητα του μοντέλου να γενικευτεί. Η υπερπροσαρμογή είναι πιθανότερη σε «πιο πολύπλοκα» μοντέλα τα οποία έχουν περισσότερη ευελιξία στην εκπαίδευση. Ένας τρόπος να εντοπιστεί το φαινόμενο αυτό είναι χρησιμοποιώντας ένα σύνολο επαλήθευσης (**validation set**) ή χρησιμοποιώντας τεχνικές για να εκτιμηθεί η ακρίβεια του μοντέλου. Μια από τις πιο δημοφιλείς τεχνικές είναι η k-fold cross validation. Επιτρέπει στο μοντέλο να εκπαιδευτεί και να αξιολογηθεί k φορές, σε διαφορετικά υποσύνολα του συνόλου εκπαίδευσης και κατασκευάζει μια εκτίμηση της απόδοσης του μοντέλου σε άγνωστα δεδομένα. Η χρήση ενός συνόλου επαλήθευσης, δηλαδή ενός συνόλου ανεξάρτητου του συνόλου εκπαίδευσης, το οποίο χρησιμοποιείται στο τέλος κάθε εποχής για την αξιολόγηση των αποτελεσμάτων, εφόσον υπάρχουν επαρκή δεδομένα, είναι επίσης μια εξαιρετική και αποδοτική τεχνική.



Σχήμα 2.6 Παράδειγμα υπερπροσαρμογής.

### Υποπροσαρμογή (Underfitting)



Η υποπροσαρμογή συμβαίνει όταν ένα μοντέλο μηχανική μάθησης δεν είναι αρκετά πολύπλοκο για να συλλάβει επακριβώς τις σχέσεις μεταξύ των χαρακτηριστικών των δεδομένων, με αποτέλεσμα να έχει αρκετά χαμηλή απόδοση και στο σύνολο δεδομένων.

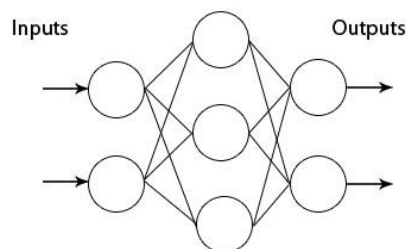
### 2.1.6 Παράδειγμα Εκπαίδευσης MLP

Με σκοπό να κατανοηθεί καλύτερα η λειτουργία ενός πολυστρωματικού perceptron παρουσιάζεται παρακάτω ένα παράδειγμα. Έστω ότι έχουμε σαν δεδομένο τον παρακάτω πίνακα με βαθμολογίες φοιτητών.

Ώρες Μελέτης	Βαθμός Προόδου	Τελικός Βαθμός
35	6,7	1 (Επιτυχία)
17	8,9	0 (Αποτυχία)
8	9,0	1 (Επιτυχία)
42	5,6	1 (Επιτυχία)
10	7,5	0 (Αποτυχία)

Πίνακας 2.1 Παραδείγματος MLP

Αρχικά θα περιγραφεί η διαδικασία εκπαίδευσης ενός mlp ώστε να μπορεί να μάθει να προβλέπει αν ένας φοιτητής θα πετύχει ή όχι στις τελικές εξετάσεις, δεδομένων των ωρών μελέτης και του βαθμού προόδου του. Για την εκπαίδευση του μοντέλου χρησιμοποιείται ο παραπάνω πίνακας, θέτοντας τις ώρες μελέτης και το βαθμό προόδου ως εισόδους και τον τελικό βαθμό ως έξοδο. Τα δεδομένα αυτά αποτελούν το σύνολο εκπαίδευσης (**train set**). Λόγω της απλότητας του προβλήματος, αρκεί ένα νευρωνικό δίκτυο με ένα επίπεδο τριών νευρώνων, που θα έχει 2 εισόδους (τις ώρες μελέτης και το βαθμό προόδου) και 2 νευρώνες εξόδου, από τους οποίους ο πρώτος δείχνει την πιθανότητα της αποτυχίας και ο δεύτερος αυτήν της επιτυχίας. Ένα τέτοιο δίκτυο φαίνεται στο σχήμα 2.7. Επιπλέον, ως συνάρτηση ενεργοποίησης θα μπορούσε να χρησιμοποιηθεί η σιγμοειδής, μιας και όπως ειπώθηκε μπορεί να λειτουργήσει αποδοτικά σε προβλήματα ταξινόμησης δύο κλάσεων.



Σχήμα 2.7 Πολυστρωματικό perceptron με ένα κρυφό επίπεδο και ένα επίπεδο εξόδου.

Οι εισόδοι πολλαπλασιάζονται με τα αντίστοιχα βάρη του πρώτου επιπέδου, τα αποτελέσματά τους δίνονται ως είσοδος στην συνάρτηση ενεργοποίησης και η διαδικασία αυτή επαναλαμβάνεται και για τις εξόδους του κρυφού επιπέδου, που δίνονται ως εισόδοι στους νευρώνες εξόδου του δικτύου. Έστω ότι από την παραπάνω διαδικασία προκύπτει ως έξοδος [0.2, 0.8] για τον πρώτο φοιτητή και [0.6, 0.4] για τον δεύτερο. Οι επιθυμητές έξοδοι είναι [0, 1] και

[1,0] όπως φαίνεται στον πίνακα 2.1. Παρατηρείται εδώ ότι για τον έναν φοιτητή η έξοδος είναι αρκετά κοντά ενώ για τον δεύτερο δεν πετυχαίνει τόσο καλά αποτελέσματα. Σε αυτό το σημείο υπολογίζονται τα σφάλματα στις εξόδους και μέσω του αλγορίθμου ανάστροφης οπισθοδρόμησης υπολογίζονται εκ νέου τα βάρη και επαναλαμβάνεται η διαδικασία ξανά και ξανά μέχρι το σφάλμα να φτάσει στο επιθυμητό επίπεδο.

Ας υποθέσουμε τώρα ότι θέλουμε να προβλέψουμε αν θα περάσει τις τελικές εξετάσεις ένας φοιτητής που διάβασε 25 ώρες και έγραψε 7,0 στην πρόοδο. Τα δεδομένα αυτά, που χρησιμοποιούνται δηλαδή χωρίς να γνωρίζουμε εκ των προτέρων το αποτέλεσμα, και στα οποία επιθυμούμε να το προβλέψουμε ονομάζονται σύνολο δοκιμής (**test set**). Με βάση τα δεδομένα εκπαίδευσης το μοντέλο θα κάνει μια πρόβλεψη, που θα αφορά τη νέα είσοδο. Η πρόβλεψη του μοντέλου εξαρτάται από τα βάρη που διαμορφώθηκαν στο στάδιο της εκπαίδευσης. Επίσης, στο στάδιο της δοκιμής τα βάρη του δικτύου δεν προσαρμόζονται και τα δεδομένα βοηθούν μόνο για την αξιολόγηση του μοντέλου.



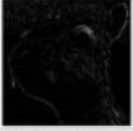




## 2.2 ΣΥΝΕΛΙΚΤΙΚΑ ΔΙΚΤΥΑ

Τα συνελικτικά νευρωνικά δίκτυα, προκειμένου να αναλύσουν δεδομένα μεγαλύτερων διαστάσεων από απλές χρονοσειρές, εκμεταλλεύονται την ιδιότητα της *συνέλιξης σημάτων*. Όπως στη θεωρία *σημάτων μιας διάστασης* (της χρονικής), η **ανάλυση Φουριέ**, προσφέρει τη δυνατότητα *αποσύνθεσης* ενός πολύπλοκου σήματος σε απλά σήματα, αλλά και τη δυνατότητα σύνθεσης πολύπλοκων σημάτων από απλές αρμονικές συναρτήσεις. Έτσι και στην *επεξεργασία εικόνας*, μπορούμε να αναλύσουμε μια πλούσια σε περιεχόμενο εικόνα χωρίς να την επεξεργαστούμε «άμεσα». Η πράξη της συνέλιξης μας προσφέρει τη δυνατότητα να αποσυνθέτουμε εικόνες και να επεξεργαζόμαστε τα απλούστερα χαρακτηριστικά τους, γεγονός που μας εξοικονομεί σημαντικούς υπολογιστικούς πόρους. Στη συνέχεια, πάλι αξιοποιώντας τις ιδιότητες της συνέλιξης, μπορούμε να ανασυνθέσουμε την αρχική εικόνα από τα δομικά της στοιχεία, τα οποία έχουμε ήδη αναλύσει, και να εξάγουμε συμπεράσματα για αυτήν.

Η διαδικασία είναι ανάλογη με τον τρόπο με τον οποίο αντιλαμβάνεται τα οπτικά ερεθίσματα ο ανθρώπινος εγκέφαλος. Στην αρχή του οπτικού δικτύου μέσα στον εγκέφαλο, υπάρχουν νευρώνες εξειδικευμένοι στην αναγνώριση πολύ απλών χαρακτηριστικών φωτεινότητας, όπως είναι οι ευθείες γραμμές ή οι γωνίες. Όσο το σήμα ταξιδεύει προς το κέντρο του εγκεφαλικού φλοιού, οι νευρώνες ή ομάδες νευρώνων οργανώνονται ώστε να αναγνωρίζουν χαρακτηριστικά υψηλότερου επιπέδου, όπως γεωμετρικά σχήματα και αντικείμενα. Έχει παρατηρηθεί πειραματικά πως υπάρχουν συστοιχίες νευρώνων σε έναν ενήλικο ανθρώπινο εγκέφαλο που ενεργοποιούνται κάθε φορά που παρατηρείται κάποια γνωστή φιγούρα, ένα σύμβολο αθλητικής ομάδας ή κάποιο αγαπημένο πρόσωπο.

### Συνέλιξη

Σε δυσδιάστατα σήματα όπως οι εικόνες, η συνέλιξη χρησιμοποιείται, είτε για εφαρμογή ενός φίλτρου σε μια εικόνα και εξαγωγή χαρακτηριστικών κατά τα στάδια δειγματοληψίας, είτε για ανακατασκευή μιας εικόνας από τα χαρακτηριστικά της, κατά την από-δειγματοληψία (*upsampling*). Οι εικόνες που προκύπτουν από τη συνέλιξη μιας εικόνας με ένα φίλτρο ονομάζονται *χάρτες χαρακτηριστικών* της εικόνας. Οι διαστάσεις τους εξαρτώνται άμεσα από τις διαστάσεις του συνελικτικού *φίλτρου*. Όσον αφορά τη διαδικασία ανακατασκευής της εικόνας έχει επικρατήσει ο όρος της από-συνέλιξης (*deconvolution*) ή ανάστροφης συνέλιξης (**transposed convolution**). Ο ρόλος της λειτουργίας αυτής είναι να δημιουργεί στην έξοδο, εικόνες μεγαλύτερης διάστασης, έτσι ώστε να είναι δυνατόν να ανακατασκευαστεί μια εικόνα από τους χάρτες χαρακτηριστικών της.

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Σχήμα 2.8 Εφαρμογές της συνέλιξης διαφόρων φίλτρων με μια εικόνα.

### Βασικά Επίπεδα

Μερικά από τα πλέον χρησιμοποιούμενα επίπεδα ενός συνελκτικού δικτύου είναι τα εξής :

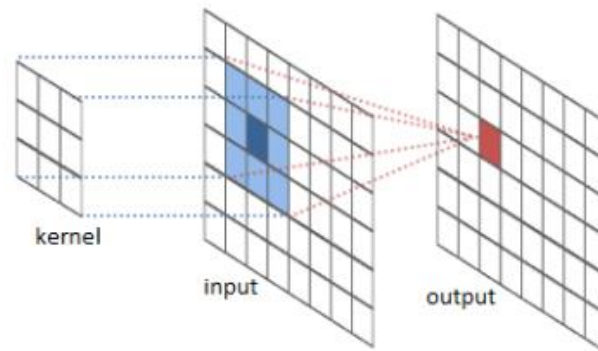
- Συνελκτικό επίπεδο (Convolution layer)
- Επίπεδο Συγκέντρωσης (Pooling Layer)
- Επίπεδο Ενεργοποίησης (Activation Layer)
- Επίπεδο Κανονικοποίησης (Normalization layer)

#### 2.2.1 Συνελκτικό Επίπεδο (Convolutional Layer)

##### Πυρήνας (Kernel)

Μια πολύ σημαντική έννοια που πρέπει να εξηγηθεί ώστε να κατανοήσουμε πλήρως την επιτυχία των συνελκτικών δικτύων στην ανάλυση εικόνας είναι αυτή του συνελκτικού *φίλτρου* (filter) ή *πυρήνα* (kernel). Πρόκειται για έναν μικρό, σε σχέση με το μέγεθος της εικόνας, τετραγωνικό πίνακα ο οποίος επιτελεί αντίστοιχη λειτουργία με αυτή του διανύσματος  $\mathbf{W}$  σε ένα νευρώνα perceptron. Όπως λοιπόν το διάνυσμα  $\mathbf{W}$  σταθμίζει την είσοδο ενός στρώματος perceptron

σε ένα MLP, έτσι και τα φίλτρα ενός συνελκτικού στρώματος εφαρμόζονται στην είσοδό του, προκειμένου να εξάγουν τα *χαρακτηριστικά* (features) αυτής της εισόδου. Το αποτέλεσμα στην έξοδο κάθε συνελκτικού νευρώνα συνδυάζεται με αυτά του υπόλοιπου στρώματος, και έτσι, στην έξοδο του στρώματος σχηματίζεται ο *χάρτης χαρακτηριστικών* (feature map) της εικόνας εισόδου.



Σχήμα 2.9 Εικονική αναπαράσταση της πράξης της συνέλιξης. Ένα φίλτρο εφαρμόζεται σε μια περιοχή ενός χάρτη χαρακτηριστικών και δίνει τιμή στην αντίστοιχη θέση του νέου χάρτη χαρακτηριστικών.

### Τοπική Συνδεσιμότητα (Local Connectivity)

Σε εισόδους υψηλής ανάλυσης, όπως είναι οι εικόνες, είναι πρακτικά αδύνατο να συνδεθούν όλοι οι νευρώνες του ενός επιπέδου με αυτούς του προηγούμενου. Γι' αυτό το λόγο κάθε νευρώνας συνδέεται μόνο με μία τοπική περιοχή της τρισδιάστατης εισόδου. Το χωρικό εύρος αυτής της συνδεσιμότητας είναι μια υπερπαράμετρος που ονομάζεται πεδίο υποδοχής (**receptive field**) του νευρώνα, που ισοδυναμεί με το μέγεθος του φίλτρου. Είναι σημαντικό να αναφερθεί ότι το βάθος, δηλαδή η τρίτη διάσταση, είναι ίδια στους νευρώνες και στα φίλτρα.

Εάν, για παράδειγμα, δώσουμε ως είσοδο μια εικόνα 32x32x3 (μέγεθος εικόνας 32x32 σε RGB μορφοποίηση) και το επίπεδο έχει πεδίο υποδοχής 5x5, τότε κάθε νευρώνας του συνελκτικού επιπέδου θα έχει συνολικά  $5 * 5 * 3 = 75$  συνδέσεις στον όγκο εισόδου. Τονίζεται ότι, η μόνη κοινή διάσταση των δύο εισόδων, είναι η τρίτη, αυτή που μετράει το βάθος τους και αυτό δεν είναι τυχαίο, αλλιώς δε θα ήταν δυνατή η συνέλιξη των σημάτων.

### Χωρική συνδεσιμότητα (Spatial Connectivity)

Οι κύριες υπερπαράμετροι του συνελκτικού επιπέδου είναι ο *αριθμός των φίλτρων* (**depth**), το μέγεθος του φίλτρου, το *βήμα* (**stride**) και το *γέμισμα* (**padding**). Χρησιμοποιώντας αυτές τις παραμέτρους γίνεται η διαδικασία της συνέλιξης. Η εικόνα εισόδου περνάει από τα φίλτρα χαρακτηριστικών του επιπέδου κι εξάγεται ένας πρώτος χάρτης χαρακτηριστικών της. Το μέγεθος του χάρτη εξαρτάται από το πλήθος και το είδος των φίλτρων του επιπέδου. Συγκεκριμένα, για κάθε φίλτρο του επιπέδου, υπολογίζονται τα εσωτερικά γινόμενα της εισόδου και του φίλτρου, και κάθε φίλτρο χρησιμεύει για ανιχνεύει κάποιο διαφορετικό γνώρισμα της εισόδου. Το βήμα προσδιορίζει τον αριθμό των εικονοστοιχείων που θα ολισθαίνει το φίλτρο κατά τη διαδικασία της συνέλιξης. Εάν το βήμα είναι 1 το φίλτρο μετακινείται στην εικόνα κατά ένα εικονοστοιχείο τη φορά, αν είναι 2 κατά δύο κ.ο.κ. Ταυτόχρονα, είναι συχνά χρήσιμο να προσθέτουμε μηδενικές τιμές στο περίγραμμα της εικόνας εισόδου (**zero-padding**). Με αυτό τον τρόπο, μπορούμε να ελέγξουμε

τις διαστάσεις της εξόδου, χωρίς παράλληλα να αλλοιώσουμε επί της ουσίας το περιεχόμενο της εικόνας εισόδου.

### Διαμοιρασμός παραμέτρων (Parameter sharing)

Ακόμη, είναι σημαντικό να αναφερθεί ότι τα συνελκτικά δίκτυα χρησιμοποιούν ένα σχήμα διαμοιρασμού παραμέτρων (parameter sharing) έτσι ώστε να μειώσουν το πλήθος των παραμέτρων. Με αυτό τον τρόπο η εκπαίδευση του νευρωνικού δικτύου γίνεται πιο γρήγορα, καθώς χρησιμοποιούνται λιγότερες πηγές δεδομένων εκπαίδευσης, χωρίς αυτό να επηρεάζει σημαντικά την τελική επίδοση του μοντέλου.

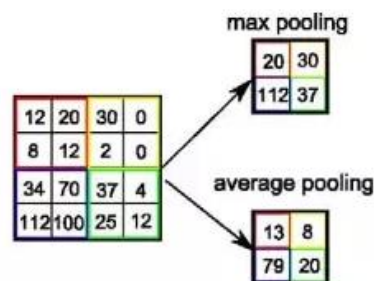
Συνοψίζοντας τα παραπάνω, ένα συνελκτικό επίπεδο λαμβάνει μια είσοδο με διαστάσεις  $W_1 \times H_1 \times D_1$ , έχει  $K$  σε αριθμό φίλτρα διαστάσεων  $F \times F \times D_1$ , βήμα  $S$  και μέγεθος zero padding  $P$ , θα έχει ως έξοδο  $W_2 \times H_2 \times D_2$ , όπου:

- $W_2 = (W_1 - F + 2P) / S + 1$ ,
- $H_2 = (H_1 - F + 2P) / S + 1$
- $D_2 = K$

Μια συνηθισμένη επιλογή παραμέτρων είναι  $F=3, S=1, P=1$ .

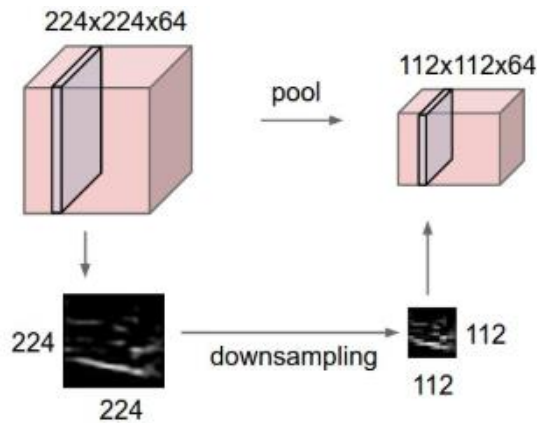
### 2.2.2 Επίπεδο Συγκέντρωσης (Pooling Layer)

Γενικά τα επίπεδα συγκέντρωσης χρησιμοποιούνται ανάμεσα σε διαδοχικά επίπεδα συνέλιξης, επειδή το κομμάτι της συνέλιξης προκαλεί τη δημιουργία αρκετών παραμέτρων. Η βασική τους λειτουργία είναι να μειώνουν το συνολικό αριθμό παραμέτρων του δικτύου, το οποίο βοηθάει στον έλεγχο της υπερπροσαρμογής. Επιπλέον, με τη χρήση ενός τέτοιου επιπέδου το δίκτυο δεν επηρεάζεται ιδιαίτερα από μικρές αλλοιώσεις ή μετασχηματισμούς της εισόδου.



Σχήμα 2.10 Η λειτουργίες της μέγιστης συγκέντρωσης και της συγκέντρωσης μέσου όρου.

Οι πιο συνηθισμένες μορφές σε συνελκτικά δίκτυα είναι του Μέγιστου, του Μέσου ή του Αθροίσματος. Στην πρώτη περίπτωση, από μια τετραγωνική «γειτονιά» ενός τετραγωνικού πίνακα κρατείται μόνο ο μεγαλύτερος όρος, στη δεύτερη κρατείται ο μέσος όρος των στοιχείων ενώ στην τελευταία το άθροισμά τους. Η πιο συνηθισμένη από αυτές είναι η Μέγιστη Συγκέντρωση (**Max Pooling**).



Σχήμα 2.11 Η λειτουργία της συγκέντρωσης.

### 2.2.3 Επίπεδα Ενεργοποίησης (Activation Layers)

Τα επίπεδα ενεργοποίησης είναι εξαιρετικά σημαντικά για ένα τεχνητό νευρωνικό δίκτυο. Είναι τα επίπεδα τα οποία εφαρμόζουν πάνω στους χάρτες ενεργοποίησης μια μη γραμμική απεικόνιση στα στοιχεία τους. Οι πιο διαδομένες συναρτήσεις ενεργοποίησης έχουν αναφερθεί στο κεφάλαιο 2.2

### 2.2.4 Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

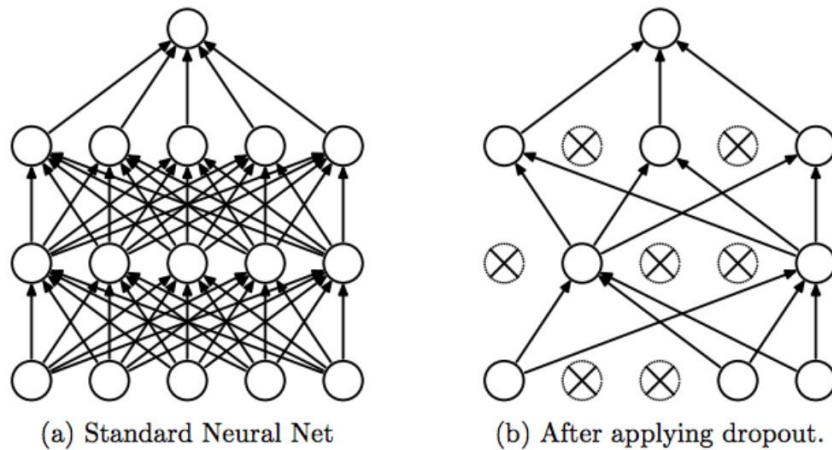
Στο πλήρως συνδεδεμένο επίπεδο κάθε νευρώνας του επιπέδου έχει πλήρεις συνδέσεις με όλους του προηγούμενου επιπέδου. Η έξοδος των συνελκτικών επιπέδων και των επιπέδων συγκέντρωσης αναπαριστά υψηλής ποιότητας χαρακτηριστικά. Η πρόσθεση ενός πλήρως συνδεδεμένου επιπέδου είναι ένας εύκολος τρόπος εκμάθησης μη γραμμικών συνδυασμών αυτών των χαρακτηριστικών. Ακόμη, σε προβλήματα ταξινόμησης, μετά την εξαγωγή των χαρακτηριστικών είναι απαραίτητη ταξινόμηση των δεδομένων στις διάφορες κατηγορίες. Αυτό μπορεί να συμβεί πολύ εύκολα χρησιμοποιώντας ένα πλήρως συνδεδεμένο επίπεδο.

### 2.2.5 Επίπεδο Κανονικοποίησης (Normalization Layer)

Η κανονικοποίηση ενός επιπέδου επιτυγχάνεται με προσαρμογή και κλιμάκωση των ενεργοποιήσεων. Για παράδειγμα, όταν έχουμε χαρακτηριστικά με τιμές από 0 μέχρι 1 και κάποια άλλα από 1 μέχρι 1000, θα έπρεπε να κανονικοποιηθούν ώστε να επιταχυνθεί η μάθηση. Πιο συγκεκριμένα, η κανονικοποίηση παρτίδων (batch Normalization) κανονικοποιεί την έξοδο του προηγούμενου επιπέδου ενεργοποίησης, αφαιρώντας το μέσον όρο της παρτίδας και διαιρώντας με την τυπική της απόκλιση. Ωστόσο, μετά από αυτή τη μετατροπή των εξόδων, τα βάρη στο επόμενο επίπεδο δεν είναι πλέον βέλτιστα. Η συνάρτηση βελτιστοποίησης αναιρεί αυτήν την κανονικοποίηση, εφόσον υπάρχει κάποιος τρόπος για αυτήν ώστε να ελαχιστοποιήσει την συνάρτηση απώλειας. Για να αποφευχθεί αυτό, η κανονικοποίηση παρτίδων προσθέτει σε κάθε επίπεδο 2 επιπλέον παραμέτρους προς βελτιστοποίηση, την παράμετρο  $\gamma$  με την οποία πολλαπλασιάζεται το επίπεδο ενεργοποίησης και την παράμετρο  $\beta$  η οποία προστίθεται σε αυτό. Με άλλα λόγια, επιτρέπει στην συνάρτηση βελτιστοποίησης να εφαρμόζει κανονικοποίηση αλλάζοντας μόνο αυτές τις 2 παραμέτρους για κάθε επίπεδο ενεργοποίησης, αντί να χάνεται η σταθερότητα του δικτύου αλλάζοντας όλα τα βάρη.

### 2.2.6 Απόσυρση (Dropout)

Ο όρος αυτός αναφέρεται στην «αγνόηση» νευρώνων κατά τη διάρκεια της εκπαίδευσης, οι οποίοι επιλέγονται τυχαία. Λέγοντας ότι οι νευρώνες αγνοούνται, εννοείται ότι δε λαμβάνονται υπόψιν κατά τη διάρκεια μια εμπρόσθιας ή οπίσθιας διάδοσης. Πιο τεχνικά, σε κάθε στάδιο εκπαίδευσης, κάθε νευρώνας με τα αντίστοιχα βάρη του, είτε απορρίπτεται με μια πιθανότητα  $1-p$  είτε όχι με πιθανότητα  $p$ , ώστε να δημιουργηθεί ένα δίκτυο με λιγότερους νευρώνες. Ο βασικός λόγος που χρησιμοποιείται το dropout είναι η αποφυγή της υπερπροσαρμογής. Ένα πλήρως συνδεδεμένο επίπεδο απασχολεί της πλειοψηφίας των νευρώνων, και κατ' επέκταση, αυτοί αναπτύσσουν σχέσεις αλληλεξάρτησης μεταξύ τους, το οποίο περιορίζει την ατομική ισχύ κάθε ξεχωριστού νευρώνα, οδηγώντας έτσι σε υπερπροσαρμογή στα δεδομένα εκπαίδευσης.



Σχήμα 2.12 Η λειτουργία της απόσυρσης. Αριστερά φαίνεται ένα πλήρως συνδεδεμένο δίκτυο και δεξιά το ίδιο μετά τη λειτουργία της απόσυρσης.

## 2.3 ΜΕΤΡΗΣΗ ΑΠΟΔΟΣΗΣ

### 2.3.1 Συναρτήσεις Σφάλματος

Για την εκπαίδευση ενός μοντέλου είναι **απαραίτητη** η χρησιμοποίηση μιας *συνάρτησης σφάλματος* (loss function). Μια συνάρτηση σφάλματος υπολογίζει πόσο κοντά στα αναμενόμενα αποτελέσματα βρίσκονται οι προβλέψεις το δίκτυο που εκπαιδεύεται. Διαφορετικές συναρτήσεις σφάλματος θα δώσουν διαφορετικά αποτελέσματα, οπότε γίνεται αντιληπτό ότι η επιλογή της κατάλληλης συνάρτησης παίζει καθοριστικό ρόλο στην επίδοση ενός μοντέλου. Παρακάτω παρουσιάζονται μερικές από τις πιο συνηθισμένες συναρτήσεις σφάλματος για προβλήματα ταξινόμησης.

- Εντροπία

Έστω ότι έχουμε  $n$  σύμβολα που ακολουθούν μία κατανομή πιθανοτήτων  $y$ , και κάθε ένα από τα σύμβολα αυτά θέλουμε να αντιστοιχηθεί σε μια δυαδική ακολουθία από bits. Αποδεικνύεται ότι για να χρησιμοποιηθεί ο μικρότερος αριθμός bits κατά μέσο όρο, θα πρέπει να ανατεθούν  $\log_2 \frac{1}{y_i}$  bits στο  $i$ -οστό σύμβολο. Εκμεταλλευόμενοι πλήρως την γνωστή κατανομή  $y$ , με τον παραπάνω τρόπο, επιτυγχάνεται βέλτιστος αριθμός bits για κάθε ανάθεση. Ο βέλτιστος συνολικός αριθμός bits ονομάζεται *εντροπία* και ο μαθηματικός της τύπος είναι:

$$H(y) = \sum_i y_i * \log \frac{1}{y_i} = - \sum_i y_i * \log y_i$$

- Σταυροειδής Εντροπία

Σε συνέχεια του ορισμού της εντροπίας, αν θεωρηθεί μια κατανομή ως ένα εργαλείο που κωδικοποιεί σύμβολα, η εντροπία μετράει τον αριθμό των bits που θα χρειαστούν εάν χρησιμοποιηθεί το σωστό εργαλείο  $y$ . Αντιθέτως, η *σταυροειδής εντροπία* (cross entropy) αναπαριστά τον αριθμό των bits που θα χρειαστούν αν κωδικοποιήσουμε σύμβολα από το  $y$ , χρησιμοποιώντας το λάθος εργαλείο  $\hat{y}$  [20]. Έτσι, στο  $i$ -οστό σύμβολο θα ανατίθενται  $\log \frac{1}{\hat{y}_i}$  bits αντί για  $\log \frac{1}{y_i}$ . Προφανώς χρησιμοποιείται η αναμενόμενη τιμή της πραγματικής κατανομής  $y$ , οπότε ο μαθηματικός ορισμός της σταυρωτής εντροπίας είναι:

$$H(y) = \sum_i y_i * \log \frac{1}{\hat{y}_i} = - \sum_i y_i * \log \hat{y}_i$$

### 2.3.2 Μετρικές

Μία μετρική συνάρτηση είναι μία συνάρτηση που χρησιμοποιείται για να κριθεί η απόδοση ενός μοντέλου. Η αποτίμησή της γίνεται με δεδομένες τις προβλέψεις καθώς και τις πραγματικές τιμές των δεδομένων του προβλήματος. Σε αντίθεση με τις συναρτήσεις σφάλματος, οι μετρικές χρησιμεύουν μόνο για την αξιολόγηση των αποτελεσμάτων, αλλά όχι κατά την εκπαίδευση του μοντέλου. Πριν παρουσιαστούν κάποιες από τις συνηθέστερες μετρικές συναρτήσεις, θα αναφερθούν κάποιες βασικές ορολογίες.

Σε προβλήματα ταξινόμησης ορίζονται οι εξής ποσότητες:

**Αληθώς Θετική-True Positive (TP):** Αναφέρεται στις προβλέψεις που εκτιμήθηκαν ότι ανήκουν σε μία κλάση και αυτό είναι αληθές.

**Ψευδώς Θετική-False Positive (FP):** Αναφέρεται στις προβλέψεις που εκτιμήθηκαν σε μία κλάση, αλλά αυτό είναι ψευδές.

**Αληθώς Αρνητική-True Negative (TN):** Αναφέρεται στις προβλέψεις που εκτιμήθηκαν ότι δεν ανήκουν σε μία κλάση και αυτό είναι αληθές.

**Ψευδώς Αρνητική-False Negative (FN):** Αναφέρεται στις προβλέψεις που εκτιμήθηκαν ότι δεν ανήκουν σε μία κλάση και αυτό είναι ψευδές.

Για παράδειγμα στο παρακάτω σχήμα παρουσιάζεται ένας πίνακας προβλέψεων σε ένα πρόβλημα ταξινόμησης αλληλογραφιών που κατηγοριοποιούνται ως spam ή no-spam. Το μοντέλο έχει κατηγοριοποιήσει 15 e-mail ως spam, από τα οποία τα 12 είναι σωστές και τα 4 λανθασμένες προβλέψεις, και 84 ως no-spam 3 από τα οποία έχουν προβλεφθεί λάθος. Όσον αφορά την κλάση spam λοιπόν, τα TP είναι 12, τα FP 4, τα TN 81 και τα FN 3.



		Predicted Class	
		Spam	No-spam
Actual Class	Spam	12	3
	No-spam	4	81

Σχήμα 2.14 Πίνακας σύγχυσης

Παρακάτω παρουσιάζονται μερικές από τις πιο συνηθισμένες μετρικές συναρτήσεις.

### Ορθότητα (Accuracy)

Η ακρίβεια είναι η πιο συνηθισμένη διαισθητική μέτρηση απόδοσης και είναι απλώς η αναλογία των σωστά προβλεπόμενων παρατηρήσεων σε σχέση με τις συνολικές παρατηρήσεις. Η μετρική αυτή είναι κατάλληλη μόνο όταν πρόκειται για συμμετρικά δεδομένα όπου τα FP και τα FN είναι περίπου ίσα. Για παράδειγμα, στα δικά μας πειράματα όπως θα παρατηρηθεί και στο κεφάλαιο 4, η ακρίβεια δεν αποτελεί κατάλληλη μέτρηση της απόδοσης.

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

### Ακρίβεια (Precision)

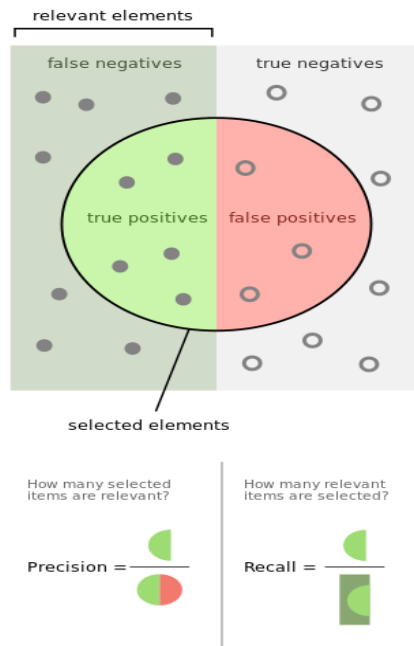
Δείχνει την αναλογία των επιτυχιών προς τις συνολικές προβλέψεις μιας κλάσης. Είναι αρκετά καλή μετρική ώστε να καθορίζει αν το ποσοστό των FP είναι υψηλό.

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

### Ανάκλαση (Recall)

Η ανάκληση είναι το ποσοστό των επιτυχιών προβλέψεων μιας κλάσης προς τα συνολικά πραγματικά στιγμιότυπά της. Αν και φαίνεται αρκετά αντιπροσωπευτική, η χρήση μόνο αυτής σπάνια είναι αρκετή και δε συνιστάται. Γι' αυτό το λόγο χρησιμοποιείται συνήθως μαζί με την precision.

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



Σχήμα 2.15 Σχηματική επεξήγηση του precision και του recall.

### Μετρική F1 (F1-score ή Dice coefficient)

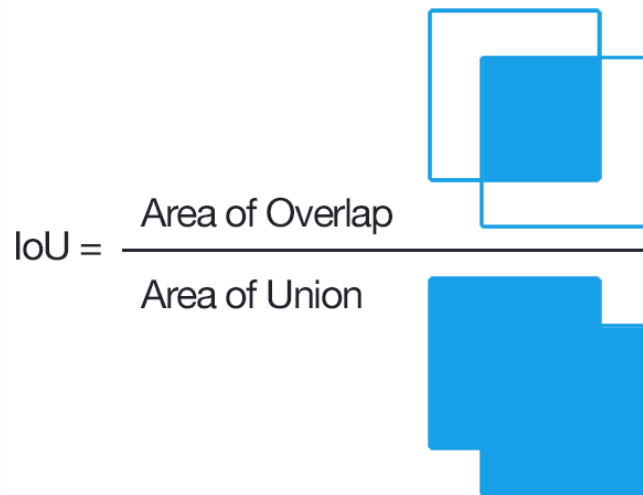
Πρόκειται για τον αρμονικό μέσο όρο των δύο παραπάνω μετρικών. Συνεπώς, λαμβάνει υπόψιν και τα FP και τα FN. Αν και εποπτικά είναι δύσκολη η αντίληψή της ως ακρίβεια, συνήθως είναι πιο χρήσιμη από την accuracy, ιδιαίτερα σε προβλήματα ανισόρροπων κλάσεων, όπως είναι αυτό που αντιμετωπίστηκε στην παρούσα διπλωματική.

$$F_1 = 2 * \frac{Recall * Precision}{Recall + Precision}$$

### Τομή προς ένωση (Jaccard coefficient ή Intersection Over Union)

Είναι μία αυστηρή μετρική που χρησιμοποιείται για τη σύγκριση της ομοιότητας και της ποικιλομορφίας του συνόλου των δειγμάτων. Ορίζεται ως το κλάσμα της τομής προς την ένωση του δείγματος δεδομένων. Η χρήση της σε προβλήματα αναγνώρισης αντικειμένων και κατάτμησης εικόνων είναι πολύ συνηθισμένη.

$$Jaccard = \frac{TP}{Tp + FP + FN}$$



Σχήμα 2.16 Σχηματική επεξήγησης της jaccard. Η τομή των δύο περιοχών διαιρείται με την ένωσή τους.



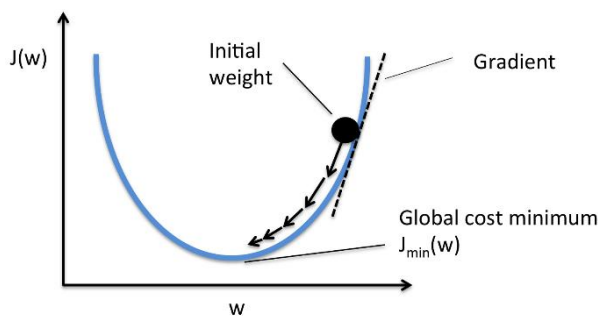
Σχήμα 2.17 Η απόδοση της jaccard για διαφορετικές περιπτώσεις. Όσο πιο μεγάλη είναι η κοινή επιφάνεια τόσο πιο καλά αποτελέσματα δίνει.

### 2.3.3 Συναρτήσεις Βελτιστοποίησης (Optimization function)

Η συνάρτηση βελτιστοποίησης είναι μια συνάρτηση με μεταβλητές τις εσωτερικές παραμέτρους του δικτύου, δηλαδή τα βάρη και την προκατάληψη τις πολώσεις (bias). Για να έχουμε ακριβείς προβλέψεις σε ένα, είναι απαραίτητο να ελαχιστοποιηθεί το υπολογιζόμενο της συνάρτησης βελτιστοποίησής του. Έχουν προταθεί πολλοί αλγόριθμοι για τη βελτιστοποίηση ενός δικτύου, και παρακάτω παρουσιάζονται οι πιο σημαντικοί.

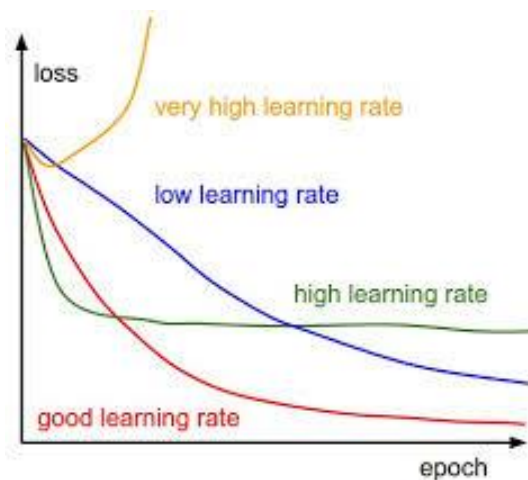
#### Gradient Descent

Η gradient decent αποτελεί θεμέλιο του τρόπου εκπαίδευσης και βελτιστοποίησης των ευφών συστημάτων στις μέρες μας και είναι ο δημοφιλέστερος τρόπος βελτιστοποίησης των νευρωνικών δικτύων. Πρόκειται για έναν επαναληπτικό αλγόριθμο ελαχιστοποίησης μια συνάρτησης απώλειας  $J(\theta)$ , όπου  $\theta$  οι παράμετροι του μοντέλου. Οι προσαρμογές που γίνονται στις παραμέτρους  $\theta$  είναι στην αντίθετη κατεύθυνση από το διάνυσμα της κλίσης  $\nabla_{\theta} J(\theta)$ . Αυτό συμβαίνει διότι το διάνυσμα των πρώτων μερικών παραγώγων της συνάρτησης δείχνει προς την κατεύθυνση όπου θα περισσότερο η τιμή της. Έχοντας βρει την κατεύθυνση προς την οποία ενημερώνονται οι παράμετροι, αυτό που μένει είναι να βρεθεί και το πόσο θα μεταβληθούν.



Σχήμα 2.18 Το μονοπάτι που ακολουθεί ο αλγόριθμος καθώς προσεγγίζει την ελάχιστη τιμή.

Για αυτόν τον σκοπό χρησιμοποιείται μια υπέρ-παράμετρος που ονομάζεται βαθμός εκμάθησης (**learning rate**). Ο βαθμός εκμάθησης καθορίζει το μέγεθος του βήματος με το οποίο μεταβάλλονται τα βάρη με σκοπό να προσεγγιστεί ένα ελάχιστο, σε σχέση με την κλίση της συνάρτησης  $J$ . Η επιλογή του βαθμού απόδοσης είναι αρκετά σημαντική. Αν το μέγεθος ενός βήματος είναι πολύ υψηλό, τότε το ελάχιστο θα ξεπερνιέται, με αποτέλεσμα ο αλγόριθμος να μη συγκλίνει σε κάποια ελάχιστη τιμή. Αντιθέτως, αν η τιμή του βαθμού απόδοσης είναι αρκετά μικρή, θα χρειαστεί πολλές επαναλήψεις για να προσεγγιστεί το ελάχιστο. Επομένως, υπάρχει ένα εύρος τιμών για το οποίο αυξάνεται η πιθανότητα για σύγκλιση του αλγορίθμου σε λογικά χρονικά όρια.



Σχήμα 2.19 Φαίνεται η επίδραση του βαθμού εκμάθησης στη συνάρτηση απώλειας.

Υπάρχουν τρεις ακόμα παραλλαγές του gradient descent, οι οποίες διαφέρουν στην ποσότητα των δεδομένων που χρειάζεται για να υπολογιστούν οι μερικές παράγωγοι. Θα αναλυθούν και οι τρεις παρακάτω.

### Batch Gradient Descent

Ο αλγόριθμος Batch Gradient Descent, γνωστός και ως vanilla gradient descent, είναι η αρχική υλοποίηση του gradient descent, στον οποίο οι μερικές παράγωγοι της συνάρτησης

απώλειας σε σχέση με τις παραμέτρους  $\theta$  του δικτύου, υπολογίζονται σε ολόκληρο το σύνολο των δεδομένων.

$$\theta = \theta - a \cdot \nabla_{\theta} J(\theta)$$

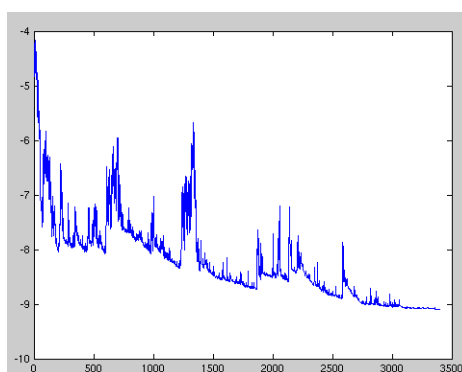
Τα πλεονεκτήματα της μεθόδου αυτής είναι παράγει σταθερό σφάλμα κλίσης και εξασφαλίζει τη σύγκλιση σε ολικό ελάχιστο για κυρτές επιφάνειες και σε τοπικό ελάχιστο για μη κυρτές. Ωστόσο, το σταθερό σφάλμα στην κλίση μπορεί μερικές φορές να οδηγήσει σε κάποιο ελάχιστον που δεν είναι το καλύτερο δυνατό. Επιπλέον, ο υπολογισμός των παραμέτρων για όλα τα δεδομένα σε ένα βήμα, καθιστά τον αλγόριθμο άχρηστο για μεγάλα σύνολα δεδομένων που δε χωρούν στη μνήμη.

### Stochastic Gradient Descent (SGD)

Το Stochastic Gradient Descent, αντιθέτως, ενημερώνει τις για κάθε ένα από τα παραδείγματα εκπαίδευσης του συνόλου δεδομένων. Για ένα παράδειγμα εκπαίδευσης  $x_i$  με την αντίστοιχη ετικέτα  $y_i$  η ενημέρωση των παραμέτρων γίνεται ως εξής:

$$\theta = \theta - a \cdot \nabla_{\theta} J(\theta, x^{(i)}, y^{(i)})$$

Η προηγούμενη μέθοδος σε μεγάλα σύνολα δεδομένων εκτελεί περισσότερους υπολογισμούς από ότι χρειάζεται, υπολογίζοντας συχνά παρόμοιες παραγώγους πριν από κάθε ενημέρωση. Το πλεονέκτημα εδώ είναι ότι εξαλείφεται αυτός ο πλεονασμός υπολογισμών, πραγματοποιώντας μία ενημέρωση σε κάθε βήμα. Η SGD είναι λοιπόν αρκετά ταχύτερη και μπορεί να χρησιμοποιηθεί για online μάθηση. Μπορεί να οδηγήσει σε μεγάλες μεταβολές της αντικειμενικής συνάρτησης, αφού εκτελεί συνεχώς ενημερώσεις με μεγάλη διακύμανση, αλλά το γεγονός αυτό μπορεί να οδηγήσει σε θετικά αποτελέσματα, όπου ο αλγόριθμος οδηγεί σε καλύτερα ελάχιστα τα οποία το batch gradient descent δε θα έβρισκε ποτέ. Επιπλέον, πειράματα έχουν αποδείξει ότι μειώνοντας τον βαθμό απόδοσης σταδιακά, ο αλγόριθμος θα συγκλίνει με υψηλό βαθμό βεβαιότητας σε κάποιο τοπικό, ή ακόμα και ολικό, ελάχιστο, σε μια μη κυρτή, η κυρτή βελτιστοποίηση αντίστοιχα [21].



Σχήμα 2.20 Διακυμάνσεις της SGD.

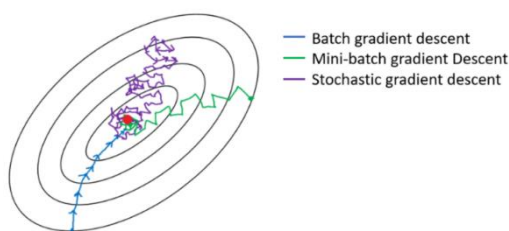
### Mini-batch Gradient Descent

Ο Mini-batch Gradient Descent είναι ένας συνδυασμός των δυο προηγούμενων αλγορίθμων. Το σύνολο δεδομένων χωρίζεται σε μικρές παρτίδες (batches) από  $b$  παραδείγματα εκπαίδευσης και πραγματοποιείται μια ενημέρωση για κάθε μία από αυτές τις παρτίδες:

$$\theta = \theta - a \cdot \nabla_{\theta} J(\theta, x^{(i:i+b)}, y^{(i:i+b)})$$

Με τον τρόπο αυτό, μειώνεται την διακύμανση των ενημερώσεων των παραμέτρων του δικτύου, το οποίο μπορεί να οδηγήσει σε σταθερή σύγκλιση. Τα τελευταία χρόνια είναι η πιο συνηθισμένη μέθοδος που χρησιμοποιείται στο πεδίο της βαθιάς μάθησης και για συντομία παρουσιάζεται ως SGD.

Ωστόσο, ο SGD δεν επιτυγχάνει πάντα βέλτιστη λύση. Αρχικά, όπως προαναφέρθηκε, η τιμή του βαθμού απόδοσης αποτελεί πολύ σημαντική επιλογή, καθώς μια μικρή τιμή προκαλεί πολύ αργή σύγκλιση, ενώ αντίθετα μια μεγάλη τιμή μπορεί να αποτρέψει τον αλγόριθμο από εύρεση λύσης και να δημιουργεί διακύμανση της συνάρτησης σφάλματος.



Σχήμα 2.21 Απεικόνιση της σύγκλισης των αλγορίθμων *batch gradient descent*, *mini-batch gradient descent* και *stochastic gradient descent*.

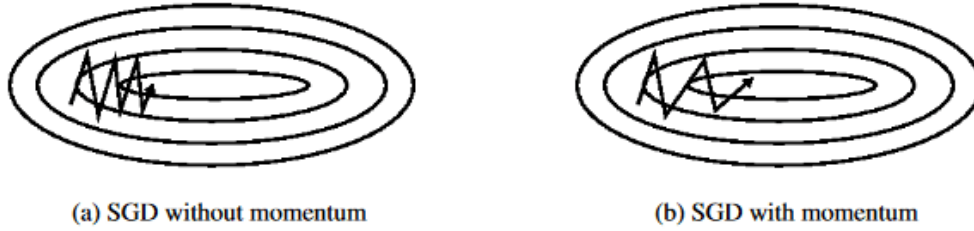
## Ορμή (Momentum)

Η SGD δεν αποδίδει τόσο καλά όταν έχει να διασχίζει μια κοιλάδα, δηλαδή μια περιοχή που η καμπυλότητα στη μία διάστασή είναι αρκετά πιο απότομη απ' ότι σε μια άλλη, οι οποίες βρίσκονται συνήθως κοντά σε τοπικά ελάχιστα. Σε τέτοιες περιπτώσεις, η SGD ταλαντώνεται μεταξύ των «πλαγιών» της κοιλάδας, ενώ κάνει πολύ μικρά βήματα προς το ζητούμενο ελάχιστο. Η ορμή [22] είναι μία μέθοδος που βοηθάει την SGD να επιταχύνει προς τη σωστή κατεύθυνση και μειώνει σημαντικά τις ταλαντώσεις (βλ. Σχήμα 2). Αυτό επιτυγχάνεται προσθέτοντας έναν όρο ορμής στο διάλυμα ενημερώσεων, ο οποίος είναι ίσος με το προηγούμενο διάλυμα ενημερώσεων πολλαπλασιασμένο με έναν παράγοντα  $\gamma$ , του οποίου οι τιμές :

$$v_t = \gamma v_{t-1} + \alpha \cdot \nabla_{\theta} J(\theta)$$

$$\theta = \theta - v_t$$

Διαισθητικά, μπορούμε να σκεφτούμε την επίδραση της ορμής σαν την ώθηση μιας μπάλας στην πλαγιά ενός λόφου. Η μπάλα αποκτά συνεχώς περισσότερη ορμή και αυξάνει την ταχύτητά της. Με το ίδιο σκεπτικό, ο όρος της ορμής αυξάνεται για παραμέτρους των οποίων οι κλίσεις έχουν την ίδια κατεύθυνση και μειώνεται γι' αυτές που οι κλίσεις τους αλλάζουν συνεχώς κατεύθυνση. Ως αποτέλεσμα, αποκομίζουμε ταχύτερη σύγκλιση και μειωμένη ταλάντωση.



Σχήμα 2.22 Επιρροή του όρου ορμής στο SGD.

## Adagrad

Ο Adagrad [23] είναι ένας αλγόριθμος ο οποίος προσαρμόζει το βαθμό εκμάθησης των παραμέτρων, εκτελώντας μικρότερες ενημερώσεις (δηλαδή χαμηλούς βαθμούς εκμάθησης) για τις παραμέτρους που σχετίζονται με τα χαρακτηριστικά που εμφανίζονται συχνά και μεγαλύτερες ενημερώσεις (δηλαδή υψηλούς βαθμούς εκμάθησης) για αυτές που αφορούν λιγότερο συχνά εμφανιζόμενα χαρακτηριστικά. Γι' αυτό το λόγο είναι ιδανικό για αραιά δεδομένα. Αντί, λοιπόν, να εκτελείται μια ενημέρωση για όλες τις παραμέτρους  $\theta$ , με κοινό βαθμό εκμάθησης, κάθε παράμετρος  $\theta_i$  διαθέτει τον δικό της βαθμό εκμάθησης τη χρονική στιγμή  $t$ . Αυτό επιτυγχάνεται διαιρώντας τον βαθμό εκμάθησης με την τετραγωνική ρίζα του αθροίσματος των τετραγώνων των παραγώγων ως προς την  $i$ -οστή παράμετρο μέχρι τη χρονική στιγμή  $t$ :

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\alpha}{\sqrt{G_{t,ii} + \epsilon}} \cdot \nabla_{\theta} J(\theta_{t,i}),$$

Η βασική αδυναμία του Adagrad είναι ότι στον παρονομαστή αθροίζονται θετικές ποσότητες, και το συνολικό άθροισμα αυξάνεται διαρκώς κατά τη διάρκεια της εκπαίδευσης. Έτσι ο βαθμός εκμάθησης σταδιακά συρρικνώνεται και τελικά γίνεται απειροελάχιστος, σε τέτοιο βαθμό που ο αλγόριθμος αδυνατεί να εκμαιεύσει περαιτέρω γνώση. Οι παρακάτω αλγόριθμοι σκοπεύουν στο να εξαλείψουν το ελάττωμά αυτό.

## RMSprop

Η λύση που του RMSprop [24] είναι πραγματικά απλή. Αντί να προστίθενται τα τετράγωνα όλων των προηγούμενων μερικών παραγώγων, ο RMSprop περιορίζει το άθροισμα αυτό για τις τελευταίες  $w$  παραγώγους. Οι προηγούμενες  $w$  παράμετροι δεν αποθηκεύονται, αλλά ορίζεται αναδρομικά το άθροισμά τους ως ένας εκθετικά φθίνων μέσος όρος

$$g_t = \nabla_{\theta} J(\theta_t)$$

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

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

## Adam

Η Adaptive Moment Estimation (Adam) [25] είναι μια ακόμα μέθοδος που υπολογίζει τους βαθμούς απόδοσης για κάθε μια από τις παραμέτρους  $\theta$ . Εκτός από την αποθήκευση ενός εκθετικά φθίνοντος μέσου όρου των τετραγώνων των προηγούμενων μερικών παραγώγων  $v_t$ , ο ADAM κρατάει επιπλέον ένα εκθετικά φθίνων μέσο όρο των προηγούμενων  $m_t$  μερικών παραγώγων, όπως και η ορμή.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$

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

όπου  $m_t$  και  $v_t$  είναι προσεγγίσεις της πρώτης και της δεύτερη ροπής των παραγώγων αντίστοιχα. Τα διανύσματα αυτά αρχικοποιούνται με μηδενικές τιμές, κάτι που τα καθιστά προδιαθετιμένα για να παίρνουν τιμές κοντά στο μηδέν, ιδικά στις αρχικά βήματα, και όταν οι βαθμοί μείωσης είναι μικροί (δηλαδή τα  $\beta_1$  και  $\beta_2$  είναι κοντά στο 1). Οι πολώσεις αυτές αντιμετωπίζονται υπολογίζοντας τις διορθωμένες τιμές των εκτιμήσεων των  $m_t$  και  $v_t$ :

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

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

Προτείνονται οι τιμές  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  και  $\varepsilon = 10^{-8}$ .



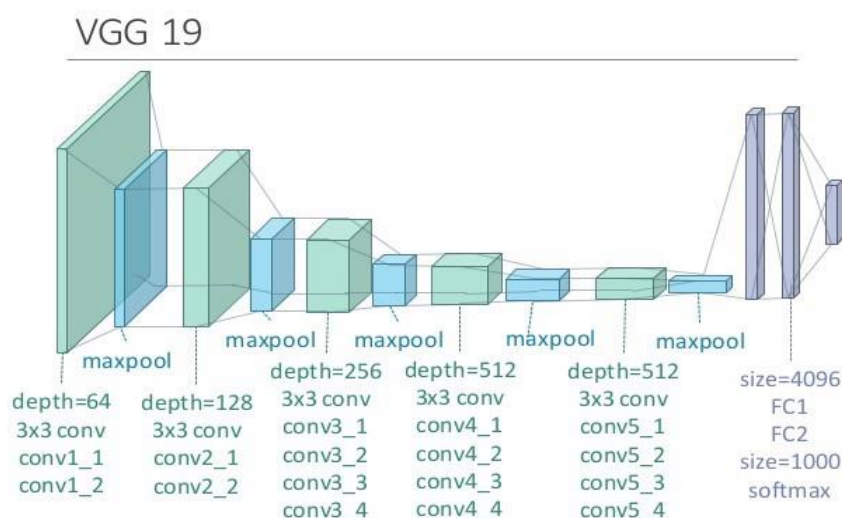
# Κεφάλαιο 3<sup>ο</sup>

## 3 ΒΑΘΙΑ ΣΥΝΕΛΙΚΤΙΚΑ ΜΟΝΤΕΛΑ

### 3.1 ΕΙΣΑΓΩΓΗ

Είδαμε στο προηγούμενο κεφάλαιο, πώς τα συνελικτικά δίκτυα αποτελούν μια *κατηγορία* των ακολουθιακών μοντέλων και καταφέρνουν, ενσωματώνοντας την πράξη της συνέλιξης να κυριαρχήσουν στην αναγνώριση και επεξεργασία εικόνας.

Σε απλά προβλήματα ταξινόμησης εικόνων, δηλαδή όταν θέλουμε να χωρίσουμε ένα σύνολο εικόνων σε υποσύνολα με βάση κάποιο κριτήριο, τα συνελικτικά δίκτυα μπορούν να τα καταφέρουν με κατάλληλη εκπαίδευση, π.χ. VGG-19 [26]. Όταν όμως το πρόβλημα που καλούμαστε να αντιμετωπίσουμε είναι πιο πολύπλοκο, τότε υπάρχει ανάγκη για εξαγωγή υψηλότερου επιπέδου χαρακτηριστικών, ώστε να αποσπάσουμε περισσότερη πληροφορία. Σε αυτό βοηθάει η προσθήκη κρυφών επιπέδων, κάτι που καθιστά ένα δίκτυο περισσότερο *βαθύ*.



Εικόνα 3.3 Επεξήγηση της αρχιτεκτονικής του μοντέλου VGG-19. Το Conv αναφέρεται στο συνελικτικό επίπεδο ενώ το FC στο πλήρως συνδεδεμένο επίπεδο.

Τα τελευταία χρόνια, που τα συνελικτικά δίκτυα χρησιμοποιούνται κατά κόρον, παρατηρήθηκε από τους ερευνητές ότι η προσθήκη επιπέδων σε ένα τέτοιο δίκτυο φέρει κινδύνους *υπερπροσαρμογής*, καθώς όσο τα δεδομένα περνούν ακολουθιακά τα διάφορα επίπεδα του δικτύου και οι χάρτες χαρακτηριστικών τους γίνονται πιο αναλυτικοί, τα βασικά και πιο γενικά χαρακτηριστικά τους τείνουν να “εξαφανιστούν” (“wash-out”), με αποτέλεσμα όταν τα δεδομένα φτάνουν προς την έξοδο, τα τελικά στρώματα του δικτύου να μην έχουν επαρκή πληροφορία για να προχωρήσουν σε σωστή εκτίμηση.

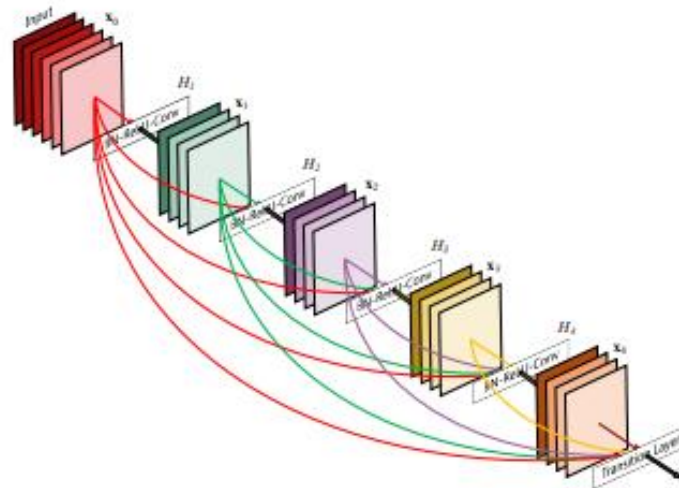
Έχουν προταθεί αρκετοί τρόποι αντιμετώπισης αυτού του φαινομένου. Το δίκτυο ResNet [27] για παράδειγμα, εφαρμόζει την τεχνική gradient descent μόνο στους παραγόμενους από κάθε επίπεδο χάρτες χαρακτηριστικών, και στη συνέχεια τους προσθέτει στο χάρτη χαρακτηριστικών της συνολικής εικόνας. Ωστόσο, με την πρόσθεση αυτή, είναι πιθανό να παρακωλύονται οι πληροφορίες που ρέουν στο δίκτυο. Με σκοπό να βελτιωθεί αυτή η ροή πληροφοριών στο δίκτυο μεταξύ των επιπέδων, προτάθηκε το παρακάτω πρότυπο συνδεσιμότητας.

## 3.2 DENSENET

Το DenseNet [28] προτείνει μια άλλη προσέγγιση για την αντιμετώπιση της αλλοίωσης (degradation) των χαρακτηριστικών της εικόνας. Η βασική του καινοτομία και η σημαντικότερη διαφορά του από τα ResNets συνίσταται στη **συνένωση** (concatenation) των χαρτών, αντί για την πρόσθεσή τους.

### Πυκνό μπλοκ (Dense block) :

Το δίκτυο αυτό εισάγει ένα νέο “επίπεδο” το οποίο ονομάζεται πυκνό μπλοκ. Η αρχιτεκτονική του επιπέδου αυτού φαίνεται στο σχήμα 3.2.



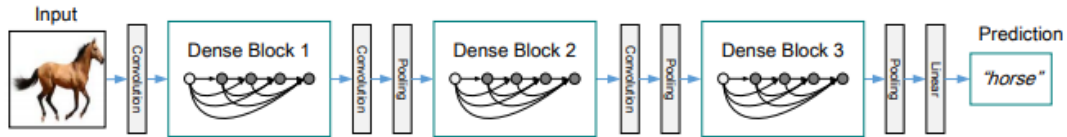
Εικόνα 3.4 Ένα πυκνό μπλοκ 5 επιπέδων με βαθμό αύξησης  $k = 4$  (growth rate). Κάθε επίπεδο παίρνει ως είσοδο όλους τους προηγούμενους χάρτες χαρακτηριστικών.

Κάθε ένα από τα πυκνά επίπεδα του πυκνού μπλοκ εξάγει κάποιο πλήθος χαρτών χαρακτηριστικών που θα δοθούν ως είσοδος σε όλα τα επόμενα πυκνά επίπεδα του δικτύου. Έτσι, το τελευταίο πυκνό επίπεδο πριν την έξοδο του δικτύου θα δεχθεί ως είσοδο τους χάρτες χαρακτηριστικών, που προέκυψαν από κάθε προηγούμενο, ενωμένους σε σειρά (concatenated). Το αποτέλεσμα αυτής της συνένωσης είναι ένα άνυσμα (tensor) με μία διάσταση περισσότερη από τις διαστάσεις εισόδου του ακολουθιακά προηγούμενου επιπέδου.

Ειδικότερα, έστω  $x_l$  η έξοδος του  $l - 1$  επιπέδου. Το  $x_l$  υπολογίζεται από την εφαρμογή του  $H_l(\cdot)$ , ενός μη γραμμικού μετασχηματιστή, στην έξοδο  $x_{l-1}$  του προηγούμενου επιπέδου.

$$x_l = H_l([x_{l-1}, x_{l-2}, \dots, x_0]),$$

όπου οι αγκύλες αντιπροσωπεύουν τη λειτουργία της συνένωσης (concatenation). Ο μη γραμμικός μετασχηματιστής  $H_i(\cdot)$  ορίζεται από τέσσερις διαδοχικές λειτουργίες, ένα επίπεδο Κανονικοποίησης Παρτίδων (Batch Normalization), ακολουθούμενο από μία μονάδα ReLU και ένα  $3 \times 3$  συνελκτικό επίπεδο. Παρακάτω παρουσιάζονται αναλυτικά οι δομές που χρησιμοποιούνται στο εν λόγω μοντέλο.



Εικόνα 3.5 Ένα Πυκνό Δίκτυο με τρία πυκνά μπλοκ. Τα επίπεδα μεταξύ δύο διαδοχικών μπλοκ αναφέρονται ως μεταβατικά επίπεδα και αλλάζουν το μέγεθος των χαρτών χαρακτηριστικών μέσω συνελίξεων και λειτουργιών συγκέντρωσης.

### Συγκέντρωση (Pooling) :

Η διαδικασία της συνένωσης είναι αδύνατη σε περίπτωση που αλλάζει το μέγεθος των χαρτών χαρακτηριστικών (feature maps). Ωστόσο, τα επίπεδα συγκέντρωσης είναι αναπόσπαστο κομμάτι ενός συνελκτικού δικτύου. Γι' αυτό το λόγο το δίκτυο διαιρείται σε πολλαπλά πυκνά μπλοκ και ανάμεσα σε αυτά τα μπλοκ παρεμβάλλονται επίπεδα που τα ονομάζουμε επίπεδα μετάβασης, τα οποία περιέχουν επίπεδα συγκέντρωσης.

### **Βαθμός αύξησης (Growth rate) :**

Η είσοδος του  $n$ -οστού επιπέδου δέχεται  $k * (n - 1) + m$  χάρτες χαρακτηριστικών, όπου  $m$  οι χάρτες χαρακτηριστικών της αρχικής εισόδου και  $k$  τα εκείνοι που παράγει το τρέχον επίπεδο. Για να αποτρέψουμε το δίκτυο να μεγαλώνει πολύ σε πλάτος θέτουμε χαμηλές τιμές στο  $k$ , πχ.  $k = 12$  και αναφερόμαστε στην παράμετρο αυτή ως βαθμός ανάπτυξης (**growth rate**). Παρά τις χαμηλές τιμές επιτυγχάνεται αρκετά καλά το επιθυμητό αποτέλεσμα. Η αιτιολόγηση του αποτελέσματος αυτού είναι ότι κάθε επίπεδο έχει πρόσβαση σε όλους τους προηγούμενους χάρτες χαρακτηριστικών του μπλοκ του και κατ' επέκταση στη «συγκεντρωτική γνώση» του δικτύου.

### **Επίπεδο συμφόρησης (Bottleneck Layer):**

Παρόλο που κάθε επίπεδο παράγει  $k$  χάρτες χαρακτηριστικών, τυπικά έχει πολύ περισσότερους για είσοδο. Για να μειωθεί λοιπόν αυτός ο αριθμός, τοποθετείται πριν από κάθε  $3 \times 3$  συνελκτικό επίπεδο, ένα  $1 \times 1$  συνελκτικό επίπεδο, που ονομάζεται επίπεδο συμφόρησης (bottleneck layer). Εκτός αν οριστεί διαφορετικά, κάθε  $1 \times 1$  συνελκτικό επίπεδο μειώνει της είσοδο σε  $4 * k$  χάρτες χαρακτηριστικών.

### **Συμπίεση (Compression):**

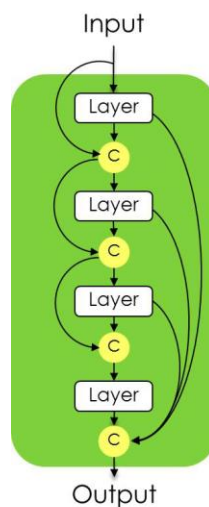
Για να καταστεί το μοντέλο ακόμα πιο συμπαγές, μπορεί να μειωθεί ο αριθμός των χαρτών χαρακτηριστικών στην έξοδο κάθε πυκνού μπλοκ. Αν για παράδειγμα ένα πυκνό μπλοκ έχει  $m$  χάρτες χαρακτηριστικών εξόδου, το ακολουθούμενο επίπεδο μετάβασης θα παράγει  $\theta * m$  χάρτες χαρακτηριστικών, όπου  $0 < \theta \leq 1$  η παράμετρος συμπίεσης.

### 3.3 ΤΟ ΠΛΗΡΩΣ ΣΥΝΔΕΔΕΜΕΝΟ ΠΥΚΝΟ ΔΙΚΤΥΟ (FC-DENSENET)

Αφού το Πυκνό Δίκτυο καταφέρνει να αντιμετωπίσει το πρόβλημα της αλλοίωσης των γενικών χαρακτηριστικών μιας εικόνας κατά την έξοδο, με το τέχνασμα της συνένωσης των χαρτών χαρακτηριστικών των δεδομένων στην πορεία τους μέσα στο Δίκτυο, είναι σχεδόν φυσική η εξέλιξή του σε ένα Πλήρως Συνελικτικό Βαθιά Συνδεδεμένο Δίκτυο. Στο σχήμα 3.5 φαίνεται πως το FC-DenseNet [29] επεκτείνει το κλασικό DenseNet αφού ουσιαστικά συνδέεται μέσω μόνο ενός ακόμη Πυκνού επιπέδου με ένα, τρόπο του λέγειν, κατοπτρικό του δίκτυο.

Το μοντέλο αυτό κατασκευάζεται από τη διαδρομή υποδειγματοληψίας (downsampling path), τη διαδρομή υπερδειγματοληψίας (upsampling path) και τις συνδέσεις μεταπήδησης (skip connections). Οι συνδέσεις μεταπήδησης βοηθούν κατά τη διαδικασία της υπερδειγματοληψίας ώστε να ανακτηθούν αναλυτικές πληροφορίες από τη διαδρομή της υποδειγματοληψίας, επαναχρησιμοποιώντας τους χάρτες χαρακτηριστικών. Θα αναλυθεί παρακάτω λεπτομερώς η αρχιτεκτονική του μοντέλου.

Η αρχιτεκτονική του DenseNet που περιεγράφηκε στο κεφάλαιο 3.1 αποτελεί το μονοπάτι υποδειγματοληψίας του FC-DenseNet. Είναι σημαντικό να σημειωθεί ότι η γραμμική αύξηση των αριθμών των χαρτών χαρακτηριστικών, αντισταθμίζεται από τη μείωση της χωρικής ανάλυσης κάθε χάρτη χαρακτηριστικών μετά από κάθε επίπεδο συγκέντρωσης. Το τελευταίο επίπεδο της διαδικασίας της υποδειγματοληψίας ονομάζεται συμφόρηση (bottleneck).

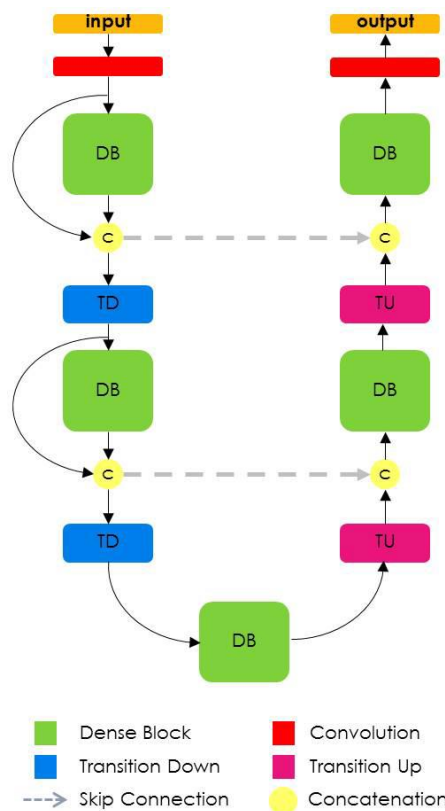


Εικόνα 3.6 Διάγραμμα ενός πυκνού μπλοκ τεσσάρων επιπέδων. Το πρώτο επίπεδο εφαρμόζεται στην είσοδο για να δημιουργήσει  $k$  χάρτες χαρακτηριστικών, οι οποίοι ενώνονται σειριακά με την είσοδο. Ένα δεύτερο επίπεδο εφαρμόζεται τότε για να δημιουργήσει άλλους  $k$  χάρτες χαρακτηριστικών, οι οποίοι και πάλι συνδέονται με τους προηγούμενους. Η διαδικασία αυτή επαναλαμβάνεται 4 φορές. Η έξοδος του μπλοκ είναι η συνένωση όλων των εξόδων των τεσσάρων επιπέδων και κατ' επέκταση εμπεριέχει  $4 * k$  χάρτες χαρακτηριστικών.

Όπως αναφέρθηκε προηγουμένως, το FC-DenseNet εισάγει μια διαδρομή υπερδειγματοληψίας, η οποία αποτελείται από συνέλιξη, λειτουργίες υπερδειγματοληψίας, και πιο συγκεκριμένα ανάστροφες συνέλιξεις ή λειτουργίες μη συγκέντρωσης, καθώς και συνδέσεις μεταπήδησης. Στα FC-DenseNets αντικαθίσταται η διαδικασία της συνέλιξης από ένα πυκνό μπλοκ και μια λειτουργία υπερδειγματοληψίας που αναφέρεται ως μετάβαση άνωθεν (transition up). Η διαδικασία αυτή αποτελείται από μία ανάστροφη συνέλιξη, η οποία υπερδειγματοληπτεί τους

προηγούμενους χάρτες χαρακτηριστικών. Οι χάρτες που προκύπτουν ενώνονται σε σειρά με αυτούς που προέρχονται από τις συνδέσεις μεταπήδησης, για να σχηματίσουν τελικά την είσοδο ενός νέου πυκνού μπλοκ. Αφού η διαδρομή υπερδειγματοληψίας αυξάνει τη χωρική ανάλυση των χαρτών χαρακτηριστικών, η γραμμική αύξηση του αριθμού των χαρακτηριστικών θα ήταν εξαιρετικά απαιτητική σε μνήμη, ειδικά για την πλήρη ανάλυση του τελευταίου επιπέδου.

Με σκοπό να ξεπεραστεί αυτός ο περιορισμός, η είσοδος ενός πυκνού μπλοκ δε συνενώνεται με την έξοδό του. Συνεπώς, η αναστροφή συνέλιξη εφαρμόζεται μόνο στους χάρτες χαρακτηριστικών που αποκτήθηκαν από το τελευταίο πυκνό μπλοκ και όχι σε όλους τους μέχρι τώρα χάρτες ενωμένους σε σειρά μεταξύ τους. Το τελευταίο πυκνό μπλοκ συνοψίζει τις πληροφορίες που περιέχονται σε όλα τα προηγούμενα πυκνά μπλοκ στην ίδια ανάλυση. Να σημειωθεί ότι κάποιες πληροφορίες από προηγούμενα πυκνά μπλοκ χάνονται κατά τη κάτωθεν μετάβαση εξαιτίας της λειτουργίας συγκέντρωσης. Παρ' όλα αυτά, οι πληροφορίες αυτές είναι διαθέσιμες στη διαδρομή υποδειγματοληψίας και μπορούν να περαστούν μέσω των συνδέσεων μεταπήδησης. Επομένως, τα πυκνά μπλοκ της διαδρομής υπερδειγματοληψίας υπολογίζονται χρησιμοποιώντας όλους τις διαθέσιμους χάρτες ενεργοποίησης, σε μια δεδομένη ανάλυση.



Εικόνα 3.7 Διάγραμμα της αρχιτεκτονικής για σημασιολογική κατάτμηση. Η αρχιτεκτονική είναι κατασκευασμένη από πυκνά μπλοκ. Το διάγραμμα αποτελείται από μία διαδρομή υποδειγματοληψίας με δύο Μεταβάσεις Κάτωθεν (TD) και μία διαδρομή υπερδειγματοληψίας με δύο Μεταβάσεις Άνωθεν (TU). Ένας κύκλος αντιπροσωπεύει τη συνένωση και τα βέλη αναπαριστούν τα μοτίβα σύνδεσης του δικτύου. Τα γκρι διακεκομμένα βέλη αναπαριστούν τις συνδέσεις μεταπήδησης, οι χάρτες χαρακτηριστικών από τη διαδρομή υποδειγματοληψίας συνενώνονται σε σειρά με αντίστοιχα της διαδρομής υπερδειγματοληψίας. Να σημειωθεί ότι, στη διαδρομή υποδειγματοληψίας, η είσοδος ενός πυκνού μπλοκ συνδέεται σε σειρά με την έξοδό του, οδηγώντας έτσι σε μία γραμμική αύξηση του αριθμού των χαρτών χαρακτηριστικών, ενώ στη διαδρομή υπερδειγματοληψίας δε συμβαίνει κάτι τέτοιο.



# Κεφάλαιο 4<sup>ο</sup>

## 4 ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΤΑΤΜΗΣΗΣ

---

### 4.1 ΔΕΔΟΜΕΝΑ ΚΑΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

Η βάση δεδομένων που χρησιμοποιήθηκε ήταν από τον διαγωνισμό MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling. Η αρχική μορφή των δεδομένων ήταν αρχεία σε ογκομετρική μορφή .nii (nifti) [30]. Περιέχει 2 φακέλους, έναν για το σύνολο εκπαίδευσης και έναν για το σύνολο αξιολόγησης, οι οποίοι είχαν πληροφορίες για 15 και 20 άτομα αντίστοιχα. Για κάθε ένα από τα άτομα υπάρχουν δύο nifti αρχεία, ένα για την αρχική μαγνητική τομογραφία και ένα για την επισημασμένη (labeled), καθώς και ένα .xml αρχείο με πληροφορίες για το όνομα, την τιμή, και το χρώμα κάθε κλάσης που υπάρχει στην επισημασμένη τομογραφία. Χρησιμοποιώντας τη βιβλιοθήκη nibabel της Python, τα παραπάνω μετατράπηκαν σε αρχεία .png. Καθότι οι διαστάσεις κάθε nifti αρχείου ήταν (256, 256, n), όπου n είναι ένας διαφορετικός αριθμός για κάθε αρχείο, η μετατροπή από την τρισδιάστατη μορφή, έγινε «τεμαχίζοντας» στον τελευταίο άξονα ώστε να προκύψουν εικόνες με διαστάσεις (256, 256), οι οποίες αναπαριστούν τις στεφανιαίες τομές του εγκεφάλου (coronal mri). Έτσι το τελικό σύνολο δεδομένων αποτελείται από 10340 εικόνες, 4403 για το σύνολο εκπαίδευσης και 5937 για αυτό της αξιολόγησης.

Οι εικόνες των αρχικών mri αποθηκεύτηκαν κάνοντας απλώς μία κανονικοποίηση στο διάστημα 0-255, ως grayscale. Στη συνέχεια, μετρήθηκαν όλες τις μοναδικές τιμές των εικονοστοιχείων στο σύνολο των επισημασμένων εικόνων και συγκρίθηκαν με τις μοναδικές τιμές αυτών στο σύνολο των .xml αρχείων. Ωστόσο παρατηρήθηκε ότι, ενώ οι συνολικές τιμές που υπήρχαν στα .xml αρχεία ήταν 201, οι κοινές ήταν 143. Το γεγονός αυτό οφείλεται στη πολύ μικρή περιοχή κάποιων κλάσεων σε σύγκριση με τον συνολικό όγκο του εγκεφάλου, κάτι που οδηγεί σε συγχώνευση αυτών με άλλες επικρατέστερες, καταλήγοντας έτσι σε 143 διαφορετικές κλάσεις. Δημιουργήθηκε όμως έτσι ένα πρόβλημα που αφορά το εύρος των τιμών, το οποίο ήταν από το 0 έως το 207. Για την εκπαίδευση του μοντέλου που παρουσιάζεται στην επόμενη παράγραφο, ήταν απαραίτητο οι κλάσεις να βρίσκονται στο διάστημα 0-142. Για τον λόγο αυτό, κατασκευάστηκε ένα ταίριασμα (mapping) των τιμών που υπήρχαν στο διάστημα 0-201 με αυτές στο διάστημα 0-142. Η αντιστοιχία αυτή μαζί με το όνομα κάθε κλάσης φαίνεται στο παράρτημα 1.

### 4.2 ΕΚΠΑΙΔΕΥΣΗ

#### 4.2.1 Αρχιτεκτονική του Μοντέλου

Στην παράγραφο αυτή θα αναλυθεί η διαδικασία εκπαίδευσης του μοντέλου FC\_DenseNet που περιγράφηκε στο κεφάλαιο 3. Πιο συγκεκριμένα, υλοποιήθηκε το FC\_DenseNet103, ή αλλιώς Tiramisù. Το μονοπάτι της υποδειγματοληψίας αποτελείται από 5 πυκνά μπλοκ και 5 μπλοκ κάτωθεν διάδοσης. Κάθε ένα από τα πυκνά μπλοκ έχουν 4, 5, 7, 10 και 12 επίπεδα. Αμέσως μετά υπάρχει το πυκνό μπλοκ κορεσμού με 15 επίπεδα. Τέλος, στο μονοπάτι της υπέρ-δειγματοληψίας έχουμε αντίστοιχα 5 μπλοκ μετάδοσης άνωθεν και 5 πυκνά μπλοκ, κάθε ένα από τα οποία έχει 12, 10, 7, 5 και 4. Στο τέλος υπάρχει ένα επίπεδο συνέλιξης 1x1, από το οποίο εξάγονται τόσοι χάρτες χαρακτηριστικών όσες και οι κλάσεις του προβλήματος, δηλαδή 143, και κανονικοποιούνται στο

διάστημα (0,1) με χρήση της softmax. Στο σχήμα 4.1 φαίνεται αναλυτικά η αρχιτεκτονική του μοντέλου.

Architecture
Input, m = 1
3 x 3 Convolution, m = 48
DB (4 layers) + TD, m = 112
DB (5 layers) + TD, m = 192
DB (7 layers) + TD, m = 304
DB (10 layers) + TD, m = 464
DB (12 layers) + TD, m = 656
DB (15 layers), m = 896
TU + DB (12 layers), m = 1088
TU + DB (10 layers), m = 816
TU + DB (7 layers), m = 578
TU + DB (5 layers), m = 384
TU + DB (4 layers), m = 256
1 x 1 Convolution, m = 143
Softmax

*Σχήμα 4.1 Αρχιτεκτονικές λεπτομέρειες του FC-DenseNet103 μοντέλου που χρησιμοποιήθηκε στα πειράματά μας. Το μοντέλο αυτό έχει κατασκευαστεί από 103 συνελκτικά επίπεδα. Το DB είναι το πυκνό μπλοκ (Dense Block), το TD είναι το μπλοκ κάτωθεν μετάβασης (Transition Down) και το TU αυτό της άνωθεν μετάβασης (Transition Up). Το m αντιστοιχεί στο συνολικό αριθμό χαρτών χαρακτηριστικών που εξάγονται στο τέλος κάθε πυκνού μπλοκ.*

Για την εκπαίδευση ενός μοντέλου το οποίο λαμβάνει ως είσοδο εικόνες, είναι αναγκαίο λόγω της υψηλής απαίτησης μνήμης, να χρησιμοποιείται μια γεννήτρια συνάρτηση. Το Keras έχει υλοποιημένο τον ImageDataGenerator ο οποίος όμως χρησιμεύει μόνο για κατηγοριοποίηση εικόνων και όχι κατάτμηση. Γι' αυτό το λόγο τροποποιήθηκε μία επέκταση του, ο SegDataGenerator<sup>1</sup>, ο οποίος λειτουργεί και για κατάτμηση εικόνων, καθώς σε κάθε κλήση του μπορεί να παράγει ταυτόχρονα εικόνες και για την είσοδο και για την έξοδο του εκάστοτε μοντέλου.

Σε όλα τα παρακάτω πειράματα δοκιμάστηκαν δύο συναρτήσεις βελτιστοποίησης του Keras, η SGD με παραμέτρους lr\_base = 0.0025 και momentum = 0.9 και η RMSprop με παραμέτρους lr=0.001 και decay=0.995. Έχοντας δοκιμάσει και τις δυο αυτές συναρτήσεις για το πρόβλημά μας καταλληλότερη ήταν η SGD. Οι κάρτες γραφικών που χρησιμοποιήθηκαν είναι η GeForce GTX 1080 Ti 11GB και η Titan Xp 12GB, ενώ η τιμή που επιλέχθηκε για το batch size ήταν 4, καθώς για μεγαλύτερες τιμές η μνήμη της κάρτας γραφικών δεν επαρκούσε. Τέλος, προσθέσαμε μια παράμετρο στο μοντέλο που ονομάστηκε Include\_top. Όταν η τιμή της είναι ψευδής δε γίνεται η χρήση της Softmax στο τελευταίο επίπεδο, κι αυτό διότι χρησιμοποιείται η Softmax μέσα στη συνάρτηση απώλειας.

<sup>1</sup> [https://github.com/ahundt/Keras-FCN/blob/densenet\\_atrous/utils/SegDataGenerator.py](https://github.com/ahundt/Keras-FCN/blob/densenet_atrous/utils/SegDataGenerator.py)



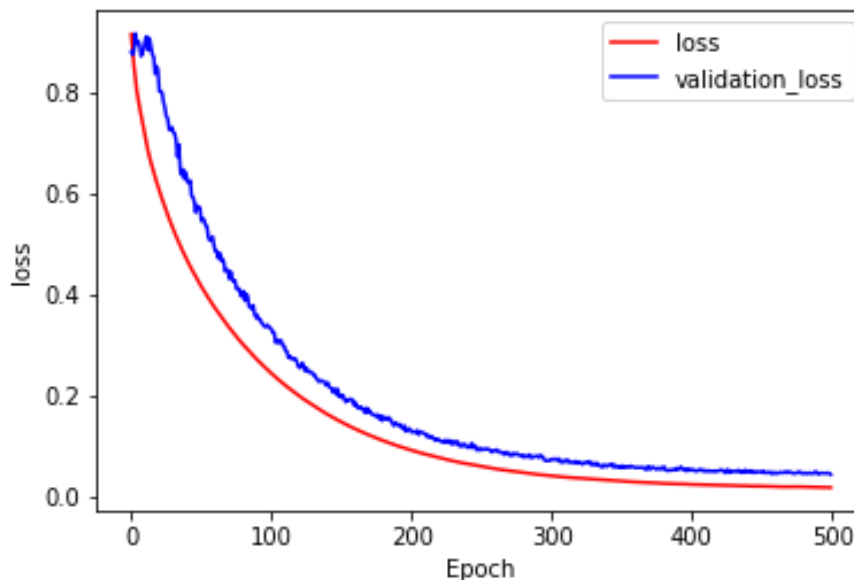
## 4.2.2 Εκπαίδευση

### Απλό Μοντέλο

Αρχικά το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας ως συνάρτηση σφάλματος την `sparse_categorical_crossentropy` του Keras. Ωστόσο, στα δεδομένα υπάρχει μεγάλη κυριαρχία της κλάσης μηδέν σε σχέση με τις υπόλοιπες. Για την ακρίβεια αποτελεί περίπου το 93% των συνολικών εικονοστοιχείων. Έτσι, με τη δεδομένη συνάρτηση απώλειας, το μοντέλο δεν κατάφερε να εκπαιδευτεί πάνω στα δεδομένα αυτά καθώς προέβλεπε συνεχώς την κλάση 0. Η εκπαίδευση έγινε για 100 εποχές και διήρκησε 1 ημέρα περίπου, ωστόσο από τις 10 πρώτες εποχές μπορούσε να διαπιστωθεί ότι τα αποτελέσματα δεν είναι τα επιθυμητά.

### Ignoring last label (ή ignoring label zero)

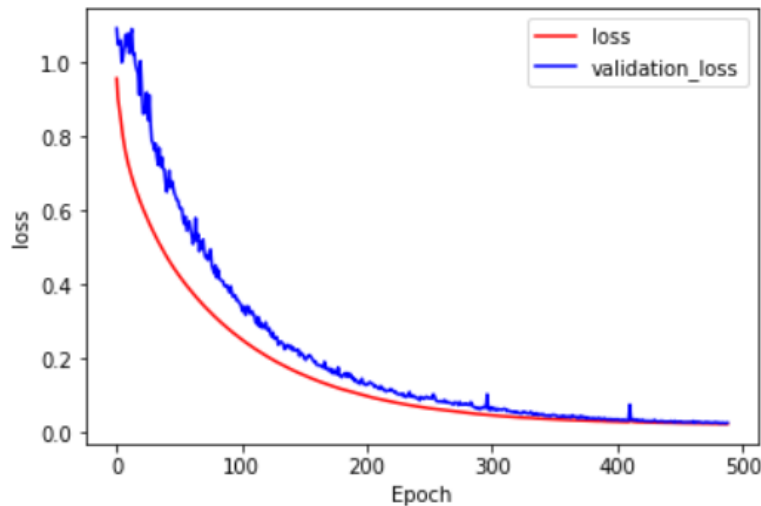
Εξαιτίας του προβλήματος των μηδενικών, ήταν αναγκαίο να δοκιμαστεί μια διαφορετική προσέγγιση για την επίλυση του προβλήματος. Χρησιμοποιήθηκε αυτή τη φορά ως συνάρτηση απώλειας η `softmax_sparse_crossentropy_ignoring_last_label`, η οποία εφαρμόζει τη συνάρτηση softmax στις προβλέψεις και μετά αυτήν της σταυρωειδούς εντροπίας ώστε να συγκριθεί η πρόβλεψη με τον στόχο και ταυτόχρονα αγνοεί την κλάση μηδέν. Έτσι ο υπολογισμός της τιμής της συνάρτησης σφάλματος και κατ' επέκταση η ενημέρωση των βαρών του μοντέλου, γίνεται χρησιμοποιώντας τις τιμές μόνο από τις υπόλοιπες κλάσεις. Με τη μέθοδο αυτή επιτυγχάνονται σαφώς καλύτερα και επιθυμητά αποτελέσματα. Η εκπαίδευση διήρκησε περίπου 5 μέρες και 500 εποχές.



Σχήμα 4.2 Συναρτήσεις απώλειας του συνόλου εκπαίδευσης (κόκκινο χρώμα) και του συνόλου επαλήθευσης (μπλε χρώμα) για τη μέθοδο `ignoring_last_label`.

## Class Weights

Στη συνέχεια, δοκιμάστηκε μια διαφορετική τεχνική η οποία δεν αγνοούσε την κλάση μηδέν, αλλά έκανε χρήση βαρών για κάθε κλάση, με σκοπό την εξισορρόπηση των κλάσεων. Πιο συγκεκριμένα, σε κλάσεις με μεγάλη συχνότητα ανατίθενται μικρές τιμές βαρών, ενώ σε αυτές με μικρή συχνότητα μεγάλες τιμές βαρών. Η συνάρτηση σφάλματος που χρησιμοποιήθηκε είναι η `softmax_sparse_crossentropy` η οποία είναι ίδια με αυτήν που αναφέρθηκε στην παράγραφο 4.2.2 λαμβάνοντας όμως υπόψιν όλες τις κλάσεις. Η εκπαίδευση διήρκησε και εδώ 5 περίπου μέρες και 500 εποχές, και όπως φαίνεται και στο σήμα 4.3 τα αποτελέσματα είναι πολύ ικανοποιητικά.



Σχήμα 4.3 Συναρτήσεις απώλειας του συνόλου εκπαίδευσης (κόκκινο χρώμα) και του συνόλου επαλήθευσης (μπλε χρώμα) για τη μέθοδο `ignoring_last_label`.

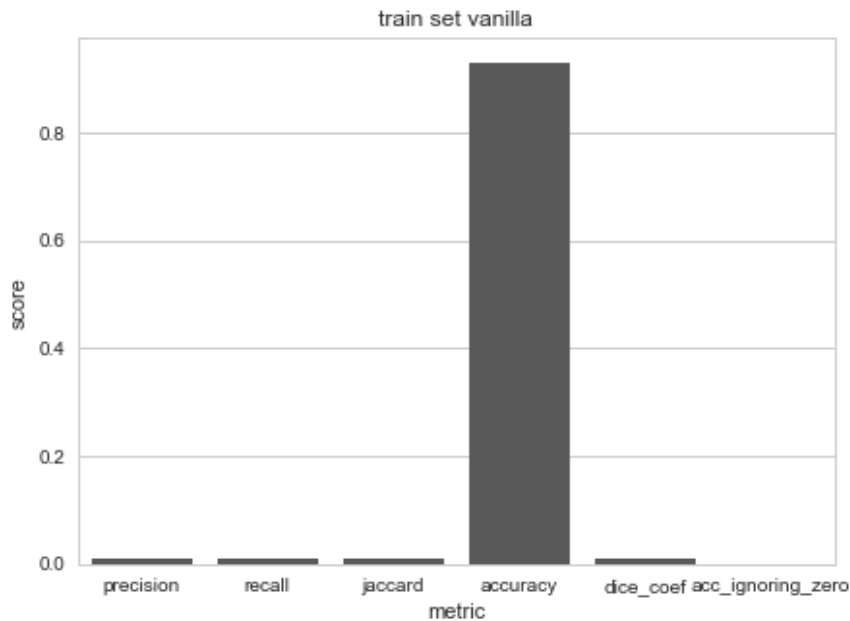
## 4.3 ΑΠΟΤΕΛΕΣΜΑΤΑ

### 4.3.1 Στατιστικά Αποτελέσματα

Ακολουθώς παρουσιάζονται τα αποτελέσματα των τριών μοντέλων στις μετρικές συναρτήσεις `precision`, `recall`, `jaccard coefficient`, `f1 score` (ή `dice coefficient`), `accuracy` και `accuracy_ignoring_zero`. Όλα τα αποτελέσματα είναι ισοσταθμισμένοι όροι ανά κλάση (`macro averaged`), δηλαδή υπολογίστηκε η τιμή της κάθε συνάρτησης για κάθε κλάση και το τελικό αποτέλεσμα είναι ο μέσος όρος τους. Με αυτή τη μέθοδο δίνεται ίδια σημασία τόσο στις κυρίαρχες κλάσεις όσο και στις πιο μικρές.

### Απλό μοντέλο

Παρατηρείται ότι η μετρική `accuracy` είναι αναξιόπιστη για την αξιολόγηση του μοντέλου. Όπως αναφέρθηκε στην προηγούμενη παράγραφο το μοντέλο αυτό προβλέπει μόνο την κλάση μηδέν, η οποία όμως αποτελεί το 93% των συνολικών εικονοστοιχείων στις κατατεταγμένες εικόνες. Έτσι η μετρική `accuracy` είναι τόσο υψηλή γιατί προβλέπει σωστά αυτήν την κλάση χωρίς όμως να επιτυγχάνεται ο γενικότερος στόχος του προβλήματος. Αυτό φαίνεται και από τις υπόλοιπες μετρικές, οι οποίες λαμβάνουν μηδενικές ή σχεδόν μηδενικές τιμές. Συγκεκριμένα, όλες εκτός από την `accuracy_ignoring_zero` πετυχαίνουν λίγο παραπάνω από το μηδέν επειδή αναγνωρίζουν σωστά την κλάση του υποβάθρου, ενώ η `accuracy_ignoring_zero` είναι ακριβώς μηδέν διότι την αγνοεί.



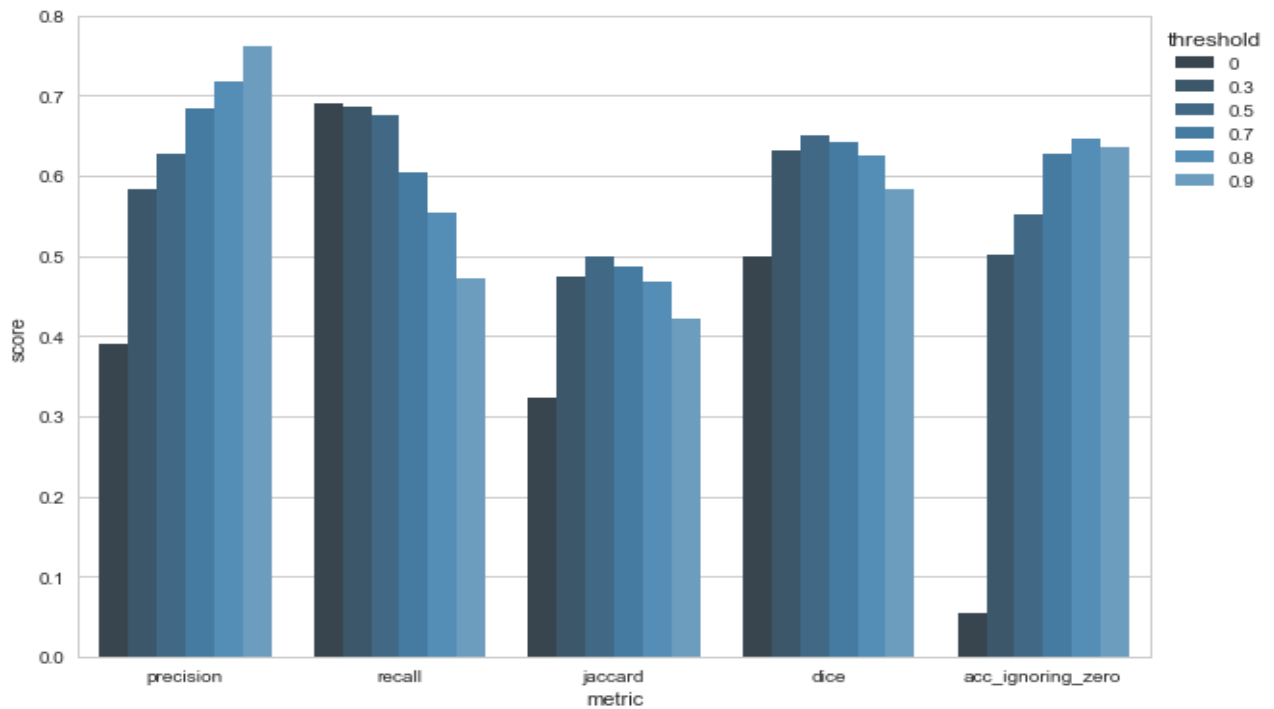
Σχήμα 4.4 Τα αποτελέσματα μετρικών συναρτήσεων για το απλό μοντέλο.

### **Ignoring label zero**

Με τη μέθοδο αυτή, όπως αναφέρθηκε, αγνοείται η κλάση μηδέν, δηλαδή δεν προβλέπεται καθόλου. Ωστόσο αυτό δεν είναι επιθυμητό διότι αποτελεί πολύ μεγάλο ποσοστό της επιφάνειας μιας εικόνας. Γι' αυτό το λόγο χρησιμοποιήσαμε κατώφλι (threshold). Η έξοδος του μοντέλου είναι 143 χάρτες χαρακτηριστικών, κάθε ένας από τους οποίους δείχνει την πιθανότητα το εικονοστοιχείο να πάρει μια συγκεκριμένη κλάση. Για παράδειγμα ο 1<sup>ος</sup> χάρτης χαρακτηριστικών δείχνει τις πιθανότητες για κάθε ένα από τα εικονοστοιχεία να ανήκουν στην κλάση 0, ο 2<sup>ος</sup> τις πιθανότητες να ανήκουν στην κλάση 1 κ.ο.κ. Έτσι, για να ληφθεί η τελική εικόνα παίρνεται κάθε φορά η θέση του χάρτη χαρακτηριστικών που έχει τη μέγιστη πιθανότητα. Το κατώφλι θέτει έναν περιορισμό σε αυτή τη διαδικασία, ώστε η θέση του μέγιστου να λαμβάνεται μόνο αν η τιμή του είναι μεγαλύτερη από αυτή του κατωφλίου, διαφορετικά θέτει στο εικονοστοιχείο την τιμή 0. Στο Σχήμα 4.5 παρουσιάζονται οι μετρικές εκτός της accuracy, για διάφορες τιμές κατωφλίου που εφαρμόστηκαν στο test set.

Όσο αυξάνεται το κατώφλι τόσο αυξάνεται και το ποσοστό της μετρικής precision, η οποία μας δείχνει πόσα από τα εικονοστοιχεία που προβλέφθηκαν σε μια κλάση ανήκουν όντως σε αυτήν. Μεγαλύτερο κατώφλι σημαίνει πιο αυστηρές επιλογές, δηλαδή το μοντέλο κατανέμει στις κλάσεις τα εικονοστοιχεία για τα οποία είναι πολύ βέβαιο. Άρα είναι αναμενόμενη αυτή η αύξηση που παρατηρείται.

Ακριβώς η αντίθετη συμπεριφορά παρατηρείται για τη μετρική recall, η οποία δείχνει από τα εικονοστοιχεία που πραγματικά ανήκουν σε μία κλάση, πόσα το μοντέλο προέβλεψε σωστά. Όσο πιο μεγάλο είναι το κατώφλι να μεν το μοντέλο είναι πιο σίγουρο για τις επιλογές τους, αλλά θα προβλέπει σε κάθε κλάση λιγότερα εικονοστοιχεία από όσα πραγματικά θα έπρεπε. Έτσι κι εδώ η συμπεριφορά είναι η αναμενόμενη.



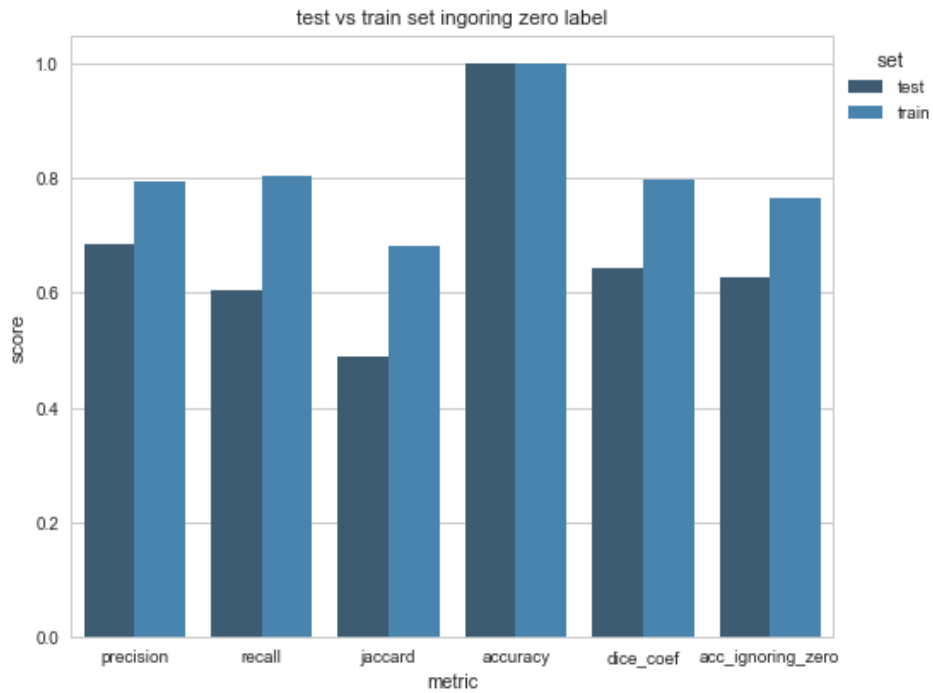
Σχήμα 4.5 Τα αποτελέσματα μετρικών συναρτήσεων για τις διάφορες τιμές κατωφλίου του μοντέλου `ignoring_label_zero`.

Από τα αποτελέσματα των μετρικών `jaccard` και `dice`, οι οποίες θεωρούνται από τις καταλληλότερες για το πρόβλημα της κατάτμησης, είναι εμφανές ότι οι μεσαίες τιμές κατωφλίου δίνουν τα καλύτερα αποτελέσματα. Και σε αυτές τις περιπτώσεις μπορεί να ειπωθεί ότι τα αποτελέσματα είναι αναμενόμενα, αφού οι δυο αυτές συναρτήσεις είναι συνδυασμός των δυο προηγούμενων, οπότε τα προβλήματα που παρουσιάζονται στο `precision` και το `recall` εξισορροπούνται στις μεσαίες τιμές.

Τέλος, θα αναφερθούμε στην μετρική `accuracy_ignoring_zero`, η οποία υπολογίζει την ορθότητα αγνοώντας την κλάση 0. Έτσι έχουμε μια αντιπροσωπευτική εικόνα για το πόσο καλά προβλέπονται τις υπόλοιπες κλάσεις το μοντέλο. Παρατηρείται ότι αυτή η μετρική επιτυγχάνει καλύτερα αποτελέσματα καθώς το κατώφλι αυξάνεται.

Λαμβάνοντας υπόψιν τα αποτελέσματα από όλες τις μετρικές, επιλέχθηκε ως καλύτερο κατώφλι αυτό με τιμή 0.7.

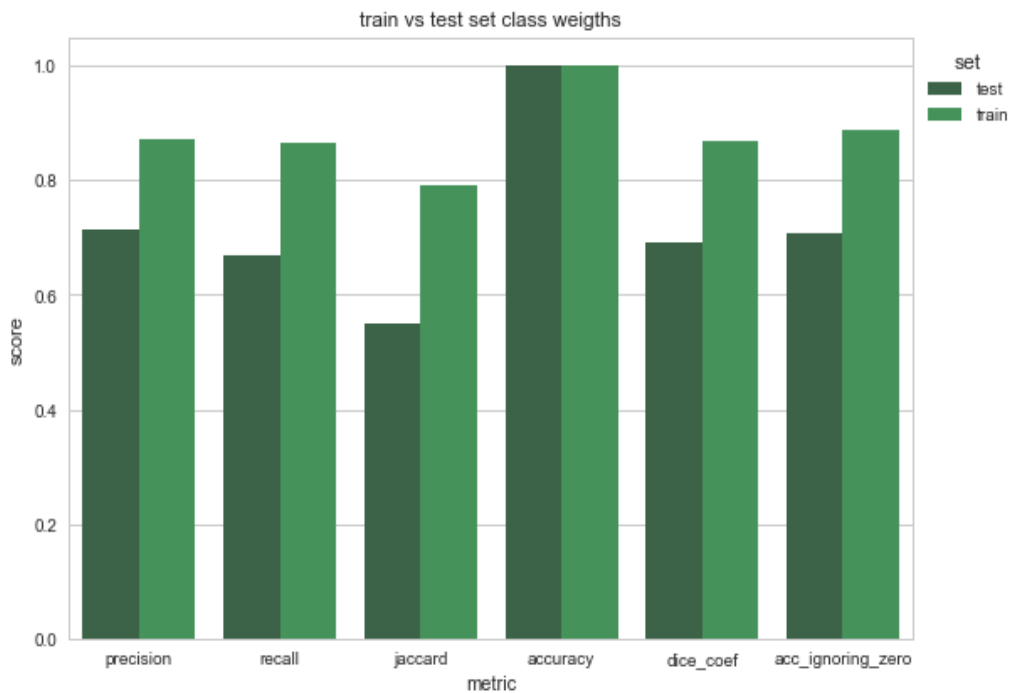
Στο επόμενο διάγραμμα φαίνεται η σύγκριση του `train set` σε σχέση με το `test set`. Όπως αναμένεται, οι τιμές στο πρώτο είναι λίγο υψηλότερες από αυτές του δεύτερου.



Σχήμα 4.6 Τα αποτελέσματα μετρικών συναρτήσεων για το σύνολο εκπαίδευσης (train set) και το σύνολο δοκιμής (test set) για τη μέθοδο ignoring label zero.

### Class weights

Παρουσιάζονται παρακάτω οι μετρικές σε test set και train set για τη μέθοδο με τα βάρη.

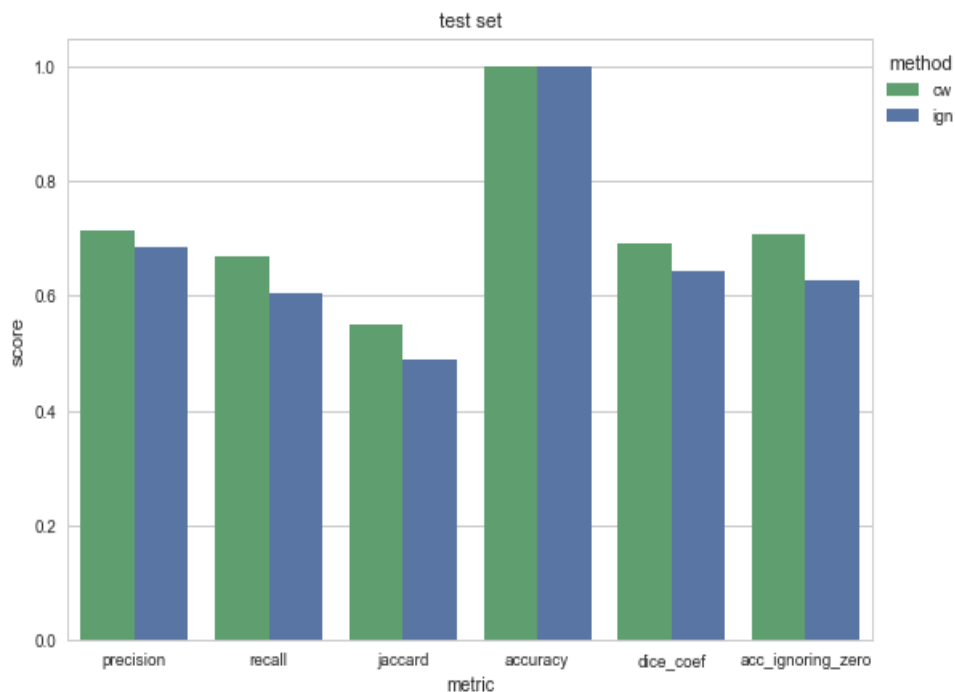


Σχήμα 4.7 Τα αποτελέσματα μετρικών συναρτήσεων για το σύνολο εκπαίδευσης (train set) και το σύνολο δοκιμής (test set) για τη μέθοδο class weights.

Είναι χρήσιμο να σημειωθεί, ότι τόσο σε αυτό το διάγραμμα όσο και στο αντίστοιχο της προηγούμενης μεθόδου, η μετρική της ορθότητας (accuracy) είναι 100% και για τα δύο σύνολα δεδομένων. Έτσι, συμπεραίνεται και από εδώ ότι τα αποτελέσματά της δεν μπορούν να δώσουν καμία αξιόπιστη πληροφορία για την απόδοση του μοντέλου. Στις υπόλοιπες μετρικές φαίνεται η διαφορά των αποτελεσμάτων τους στα δύο σύνολα, όπου παρατηρείται σαφώς καλύτερη απόδοση στο train set.

### Σύγκριση ignoring label zero vs class weights

Τέλος, παρουσιάζεται μια σύγκριση των αποτελεσμάτων των δύο μεθόδων που χρησιμοποιήθηκαν στο test set. Αν εξαιρέσουμε τη μετρική της ακρίβειας, η οποία όπως αναφέρθηκε δεν μπορεί να δώσει κάποια χρήσιμη πληροφορία, σε όλες τις υπόλοιπες η μέθοδος με τα βάρη υπερिशύει.

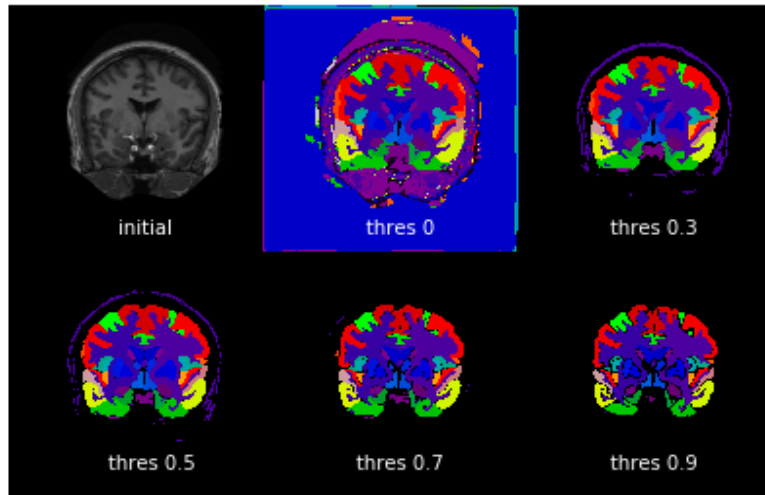


Σχήμα 4.8 Σύγκριση των αποτελεσμάτων των μετρικών συναρτήσεων για το σύνολο δοκιμής (test set) των μεθόδων ignoring label zero και class weights.

### 4.3.2 Εμπειρικά Αποτελέσματα

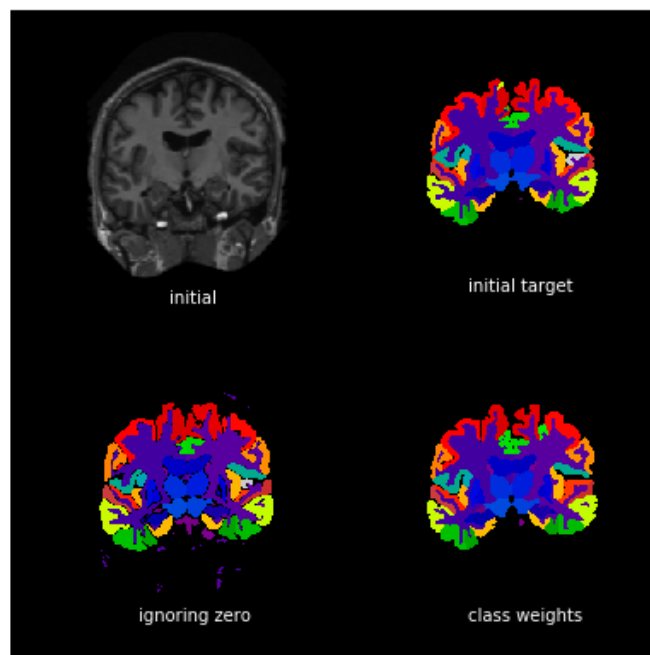
Για την καλύτερη ερμηνεία των αποτελεσμάτων, παρουσιάζονται παρακάτω οι προβλέψεις των μοντέλων ως εικόνες. Στο Σχήμα 4.9 παρουσιάζονται οι προβλέψεις του μοντέλου ignoring\_label\_zero, για τις διάφορες τιμές κατωφλίου όπως φαίνονται κάτω από κάθε εικόνα.

Στην πρόβλεψη με κατώφλι μηδέν, όταν ουσιαστικά δε χρησιμοποιείται κατώφλι, παρατηρείται ότι η κλάση μηδέν που αντιστοιχεί στο υπόβαθρο, και φαίνεται με μαύρο χρώμα, δεν προβλέπεται καθόλου. Αυτό είναι αναμενόμενο διότι, αφού το μοντέλο δεν προβλέπει ποτέ μηδέν, χωρίς τη χρήση κατωφλίου η τιμή αυτή δε θα δοθεί σε κανένα εικονοστοιχείο. Καθώς αυξάνεται το κατώφλι, το φαινόμενο αυτό εξαφανίζεται, όμως η κάθε περιοχή αρχίζει και συρρικνώνεται. Αυτές οι δύο παρατηρήσεις, συνάδουν απόλυτα με τη λειτουργία των μετρικών precision και recall, που αναφέρθηκαν στην προηγούμενη παράγραφο.



Σχήμα 4.9 Οι προβλέψεις του μοντέλου με χρήση της μεθόδου *ignoring label zero* για τις διάφορες τιμές κατωφλίου σε μία εικόνα από το σύνολο δοκιμής (*test set*).

Τέλος, στο επόμενο σχήμα παρουσιάζεται μια εποπτική σύγκριση των δύο μεθόδων. Στο σχήμα διακρίνονται τόσο οι προβλέψεις που προέκυψαν χρησιμοποιώντας τα δύο μοντέλα, καθώς και η αρχική εικόνα με την αντίστοιχη κατατετημημένη της από τα δεδομένα που χρησιμοποιήθηκαν.



Σχήμα 4.10 Εποπτική σύγκριση των μεθόδων *ignoring label zero* και *class weights*.

Επιβεβαιώνονται και εδώ τα αποτελέσματα των μετρικών συναρτήσεων. Η πρόβλεψη της μεθόδου με τα βάρη είναι σχεδόν πανομοιότυπη με την εικόνα στόχο. Αντιθέτως, η μέθοδος που αγνοεί το υπόβαθρο, αν και πετυχαίνει καλά αποτελέσματα, είναι εμφανές ότι περιέχει περισσότερο «θόρυβο» απ' όσο χρειάζεται.





# Κεφάλαιο 5<sup>ο</sup>

## 5 ΕΦΑΡΜΟΓΗ ΤΗΣ ΚΑΤΑΤΜΗΣΗΣ ΣΕ ΙΑΤΡΙΚΕΣ ΕΙΚΟΝΕΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

---

### 5.1 ΣΥΛΛΟΓΗ ΚΑΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Σκοπός της παρούσας εργασίας είναι η εφαρμογή του μοντέλου που περιγράφηκε στο κεφάλαιο 4, με σκοπό να επιτευχθεί ο διαχωρισμός πασχόντων, από κάποια νευροεκφυλιστική ασθένεια, και μη πασχόντων ατόμων από τις εγκεφαλικές τους τομογραφίες. Πιο συγκεκριμένα μελετήθηκαν τομογραφίες ασθενών με την ασθένεια του Alzheimer. Η ασθένεια του Alzheimer πρόκειται για μια μη θεραπεύσιμη και θανατηφόρα νευροεκφυλιστική νόσο, στην οποία οφείλεται η πλειοψηφία των περιπτώσεων άνοιας. Τα εγκεφαλικά κύτταρα των ασθενών με τη συγκεκριμένη νόσο, εκφυλίζονται και πεθαίνουν, προκαλώντας μια σταθερή μείωση στη μνήμη και τη διανοητική ικανότητα. Εδώ και πολλές δεκαετίες έχει γίνει πληθώρα ερευνών για να βρεθεί η ακριβής επίπτωση που έχει η νόσος στον ανθρώπινο εγκέφαλο και έχει παρατηρηθεί σημαντική ατροφία στις περιοχές του ιππόκαμπου (hippocampus), της πάρα-ιπποκάμπειας έλικας (parahippocampal gyrus), του ενδορινικού φλοιού (entorhinal cortex) και της αμυγδαλής (amygdala) [31] [32].

Τα δεδομένα τα οποία χρησιμοποιήθηκαν για τα πειράματα είναι υποσύνολο της βάσης δεδομένων ADNI (Alzheimer's disease Neuroimaging Initiative) [33]. Η βάση αυτή περιέχει δεδομένα από ιατρικές εξετάσεις ασθενών που έχουν εμφανίσει συμπτώματα της νόσου Alzheimer. Για τη συγκεκριμένη μελέτη αξιοποιήθηκαν οι T1 τομές από ασθενείς που είχαν χαρακτηριστεί είτε Normal είτε AD, δηλαδή ασθενείς χωρίς τη νόσο και ασθενείς με τη νόσο του Alzheimer αντίστοιχα. Δε λήφθηκαν υπόψη οι επιδόσεις των ασθενών στα σχετικά τεστ (FAQ, GDS-C, Global CDR, MMSE, NPI-Q). Για την καλύτερη δυνατή ποιότητα εικόνας επιλέχθηκαν προεπεξεργασμένα δεδομένα, ενώ για κάθε επίσκεψη ενός ασθενή ήταν διαθέσιμα τα δεδομένα σε κάθε στάδιο της προεπεξεργασίας (MPR-R, GradWarp, N3, Scaled/Scaled 2). Επιλέχθηκαν τα αρχεία MPR-R; GradWarp; N3; Scaled για τη μέγιστη ποιότητα, ενώ αγνοήθηκαν τα αρχεία MPR-R; GradWarp; N3; Scaled 2 για λόγους συνοχής των δεδομένων: Αφ' ενός ήταν πολύ λιγότερα, αφ' ετέρου δεν υπήρχε κάποια εμφανής διαφορά με τα Scaled.

Από την παραπάνω διαδικασία επιλογής προέκυψε το τελικό σετ δεδομένων που αξιοποιήθηκε για την κυρίως επεξεργασία (πίνακας 5.1.). Παρατηρείται ότι υπάρχουν διαφορετικές διαστάσεις στις εικόνες, κάτι που πιθανόν να οφείλεται σε λήψεις από διαφορετικούς τομογράφους. Για λόγους συνοχής στη μεθοδολογία επεξεργασίας, επιλέχθηκαν μόνο τα αρχεία με διαστάσεις (256, 256, 166), τα οποία αφορούν σε 195 ασθενείς με τη νόσο του Alzheimer και σε 347 ασθενείς χωρίς αυτήν. Όπως αναφέρθηκε και στο κεφάλαιο 4, η εκπαίδευση του FC\_DenseNet έγινε με στεφανιαίες τομές μαγνητικών τομογραφιών, οι οποίες στα δεδομένα του ADNI προκύπτουν αποδομώντας τα nifti αρχεία στη 2<sup>η</sup> διάσταση δημιουργώντας έτσι εικόνες με διαστάσεις (256, 166). Οι εικόνες αυτές κανονικοποιήθηκαν στο διάστημα 0-255 και οι διαστάσεις τους μετασχηματίστηκαν σε (256, 256), χρησιμοποιώντας το Antialias, ένα υψηλής ποιότητας φίλτρο βασισμένο στη συνέλιξη.

AD		Normal	
Διαστάσεις	Αριθμός εικόνων	Διαστάσεις	Αριθμός εικόνων
(192, 192, 160)	186	(192, 192, 160)	335
(192, 192, 176)	1	(240, 256, 160)	61
(240, 256, 160)	49	(248, 256, 160)	1
(256, 256, 160)	1	(256, 256, 160)	2
(256, 256, 162)	1	(256, 256, 166)	347
(256, 256, 166)	195		

Πίνακας 5.1: Διαστάσεις και αριθμός εικόνων μετά από την μετατροπή των δεδομένων σε .png.

Για λόγους πληρότητας, ένας μαγνητικός τομογράφος λαμβάνει δεδομένα από το μπροστινό μέρος του κεφαλιού μέχρι το πίσω, με αποτέλεσμα οι πρώτες και οι τελευταίες εικόνες να μην περιέχουν καμία πληροφορία για τον εγκέφαλο. Για το λόγο αυτό δεν αξιοποιήθηκαν οι πρώτες 30 και οι τελευταίες 36 εικόνες κάθε ασθενούς. Με βάση τα παραπάνω τα τελικά δεδομένα αποτελούνται από 37050 εικόνες AD και 65930 Normal.

## 5.2 ΜΕΘΟΔΟΛΟΓΙΑ

Οι περιοχές του εγκεφάλου που επηρεάζονται σε ασθενείς με τη νόσο του Alzheimer είναι, όπως αναφέρθηκε στην προηγούμενη παράγραφο, ο ιππόκαμπος, η πάρα-ιπποκάμπεια έλικα, η αμυγδαλή και ο ενδορινικός φλοιός. Για να εξαχθούν πληροφορίες από τη μαγνητική τομογραφία κάθε ασθενούς, χρησιμοποιήθηκε αρχικά το μοντέλο FC\_DenseNet, που εκπαιδεύτηκε όπως περιγράφηκε στο κεφάλαιο 4, και έτσι για κάθε εικόνα προβλέφθηκε η αντίστοιχη καταταμημένη της. Οι τέσσερις περιοχές προς εξέταση, αντιστοιχούν σε 8 κλάσεις του μοντέλου, μία για το αριστερό και μία για το δεξί μέρος του εγκεφάλου. Πιο συγκεκριμένα οι κλάσεις αυτές αναφέρονται στα δεδομένα του MultiAtlas ως Right Hippocampus, Left Hippocampus, Right PHG, Left PHG, Right Amygdala, Left Amygdala, Right Ent και Left Ent. Για κάθε έναν από τους ασθενείς, υπολογίστηκε ο αριθμός των εικονοστοιχείων (pixels) από το σύνολο των καταταμημένων εικόνων του εγκεφάλου του, που αντιστοιχούν σε κάθε μία από τις κλάσεις αυτές καθώς και τα συνολικά εικονοστοιχεία που δεν παίρνουν την τιμή μηδέν, δηλαδή το background. Τα αποτελέσματα αυτά αποθηκεύτηκαν σε ένα excel αρχείο, όπου υπολογίστηκαν τα ποσοστά κάθε κλάσης σε σχέση με τον συνολικό όγκο του εγκεφάλου, δηλαδή τον αριθμό των μη μηδενικών εικονοστοιχείων. Χρησιμοποιώντας την εφαρμογή Jupyter Notebook, δημιουργήθηκε ένας πίνακας με τα παραπάνω ποσοστά για κάθε ασθενή, καθώς και ένας πίνακας με δυαδικές τιμές ώστε να χαρακτηρίζεται κάθε ασθενής ως AD (τιμή 1) ή Normal (τιμή 0).

## 5.3 ΑΠΟΤΕΛΕΣΜΑΤΑ

Μετά τον υπολογισμό του μεγέθους των περιοχών με την απαρίθμηση των εικονοστοιχείων της καθεμίας, έπρεπε να διαπιστωθεί αν τα μεγέθη αυτά συμβαδίζουν με τη βιβλιογραφία. Αν δηλαδή, οι περιοχές στους πάσχοντες ήταν μικρότερες από αυτές στους μη πάσχοντες και αν η διαφορά αυτή ήταν στατιστικά σημαντική. Ακολουθήθηκαν τρεις μεθοδολογίες: Η πρώτη αφορούσε τη συσχέτιση μίας προς μίας των περιοχών βάσει μεγέθους, με την κατηγορία του συγκεκριμένου εξεταζόμενου (AD ή Normal). Η δεύτερη απαιτούσε την εκπαίδευση ταξινομητών στα δεδομένα αυτά προκειμένου να διαπιστωθεί αν ένα μοντέλο θα μπορούσε να μάθει να

διαχωρίζει τους ασθενείς στις δυο κατηγορίες, μόνο από τα μεγέθη των αντίστοιχων περιοχών. Τέλος, η τρίτη περιλάμβανε την αναπαράσταση και την εμπειρική αξιολόγηση των δεδομένων.

### 5.3.1 Τεστ Συσχέτισης Περιοχών με τη Μεταβλητή Εξόδου

Προκειμένου να διαπιστωθεί η συσχέτιση της κάθε κλάσης με τη δυαδική μεταβλητή, χρησιμοποιήθηκε ο συντελεστής συσχέτισης point biserial [34]. Τα αποτελέσματα φαίνονται στον πίνακα 5.2. Παρατηρείται ότι εκτός από την Right Ent όλες οι υπόλοιπες παρουσιάζουν συσχέτιση, άλλες σε μεγαλύτερο και άλλες σε μικρότερο βαθμό, με κυρίαρχες τις Left PHG, Left Hippo και Left/Right Amygdala. Οι συσχετίσεις θα αναδειχθούν περισσότερο και στην επόμενη παράγραφο.

	Left Amygdala	Left Ent	Left Hippo	Left PHG	Right Amygdala	Right Ent	Right Hippo	Right PHG
r_value	<b>-0.3518</b>	-0.2618	<b>-0.4466</b>	<b>-0.5709</b>	<b>0.4206</b>	-0.0233	-0.1809	0.1586
p_value	<b>3.09e-17</b>	2.23e-10	<b>6.24e-28</b>	<b>3.29e-48</b>	<b>1.19e-24</b>	<b>0.0002</b>	2.25e-05	<b>0.5883</b>

Πίνακας 5.2 R\_value και p\_value για τον συντελεστή συσχέτισης point biserial.

### 5.3.2 Εκπαίδευση Ταξινομητών

Με σκοπό να αξιολογηθεί όσο το δυνατόν καλύτερα η κατηγοριοποίηση που μπορεί να επιτευχθεί, τα δεδομένα χωρίστηκαν σε test set και train set και εφαρμόστηκαν οι τιμές των ποσοστών κάθε μεταβλητής του πίνακα ποσοστών που αναφέρθηκε στο κεφάλαιο 5.2, ως είσοδος σε διάφορους ταξινομητές. Η μεγαλύτερη ακρίβεια επιτεύχθηκε με τον Random Forest\*\*. Πιο συγκεκριμένα, εκτελέστηκαν διάφορα πειράματα σε κάθε ένα από τα οποία κρατήθηκε για το test set το 10%, το 20%, το 25%, το 33% και το 50% των δεδομένων ενώ το εναπομείναν ποσοστό αποτέλεσε το train set. Να σημειωθεί ότι ο διαχωρισμός που αναφέρθηκε έχει γίνει ανά ασθενή.

Σε όλα τα πειράματα ταξινομήσεων με εφαρμογή του ταξινομητή Random Forest η ακρίβεια που επιτεύχθηκε ήταν πάνω από 98%. Στη συνέχεια προκειμένου να εξεταστεί η ικανότητα του μοντέλου να παράγει ακριβή αποτελέσματα με λιγότερη διαθέσιμη πληροφορία αλλά και να αξιολογηθεί η επίδραση του κάθε χαρακτηριστικού στην αποτελεσματικότητα του διαχωρισμού, επαναλήφθηκε η διαδικασία, δίνοντας κάθε φορά ως είσοδο μόνο δύο από τα οχτώ χαρακτηριστικά του πίνακα 5.2. και χρησιμοποιώντας το 10% των εξεταζόμενων για το test set. Στον πίνακα 5.3 παρουσιάζεται ότι ο Random Forest πετυχαίνει πολύ καλή ακρίβεια για τους συνδυασμούς που περιέχουν τα Left και Right Amygdala, Left PHG και Left Hippo, ενώ αντιθέτως δεν τα πάει καθόλου καλά με συνδυασμούς των Left και Right Ent.

Στον πίνακα 5.4 επισημαίνονται οι πιο σημαντικές περιοχές/χαρακτηριστικά (πράσινο) στην εφαρμογή του ταξινομητή Random Forest για τον ακριβή διαχωρισμό πασχόντων και μη πασχόντων της νόσου Alzheimer.

Με σκοπό να παραχθεί μια επιπλέον επιβεβαίωση της επίδρασης και σημαντικότητας των περιοχών/χαρακτηριστικών για την ταξινόμηση διαχωρισμού, χρησιμοποιήθηκε επιπρόσθετα ένα γραμμικό μοντέλο που εκπαιδεύεται με ποινή την απόλυτη τιμή της νόρμας (L1 norm penalty) των βαρών, το μοντέλο Lasso. Τα αποτελέσματα από την εφαρμογή του μοντέλου παρουσιάζονται στον πίνακα 5.5.

Οι πιο σημαντικές περιοχές/χαρακτηριστικά και για τα 2 μοντέλα ήταν οι δύο Αμυγδαλές (κυρίως η αριστερή) και η αριστερή πάρα-ιπποκάμπεια έλικα (PHG). Τα αποτελέσματα αυτά συμβαδίζουν απόλυτα με τις συσχετίσεις του πίνακα 5.2, καθώς και με τα ζεύγη του πίνακα 5.3.

	<b>Right Hippo</b>	<b>Left Ent</b>	<b>Right Ent</b>	<b>Left Amygdala</b>	<b>Right Amygdala</b>	<b>Right PHG</b>	<b>Left PHG</b>
<b>Left Hippo</b>	0.8181	0.7636	0.6909	<b>0.9636</b>	0.8545	<b>0.9272</b>	0.8545
<b>Right Hippo</b>		<b>0.6</b>	<b>0.6363</b>	<b>0.9090</b>	0.8	0.8	<b>0.9454</b>
<b>Left Ent</b>			<b>0.5090</b>	0.8545	0.7818	<b>0.6181</b>	0.8
<b>Right Ent</b>				<b>0.9818</b>	<b>0.8909</b>	<b>0.6181</b>	0.8545
<b>Left Amygdala</b>					<b>1.0</b>	<b>0.9454</b>	<b>0.9818</b>
<b>Right Amygdala</b>						0.8363	0.8727
<b>Right PHG</b>							<b>0.8909</b>

Πίνακας 5.3 Πίνακας με τα σκορ ακρίβειας του ταξινομητή Random Forest για όλους του συνδυασμούς ανά 2 περιοχών/χαρακτηριστικών.

Left Amygdala	Left Ent	Left Hippo	Left PHG	Right Amygdala	Right Ent	Right Hippo	Right PHG
0,3726	0,027	0,0721	0,167	0,2189	0,0303	0,0537	0,0584

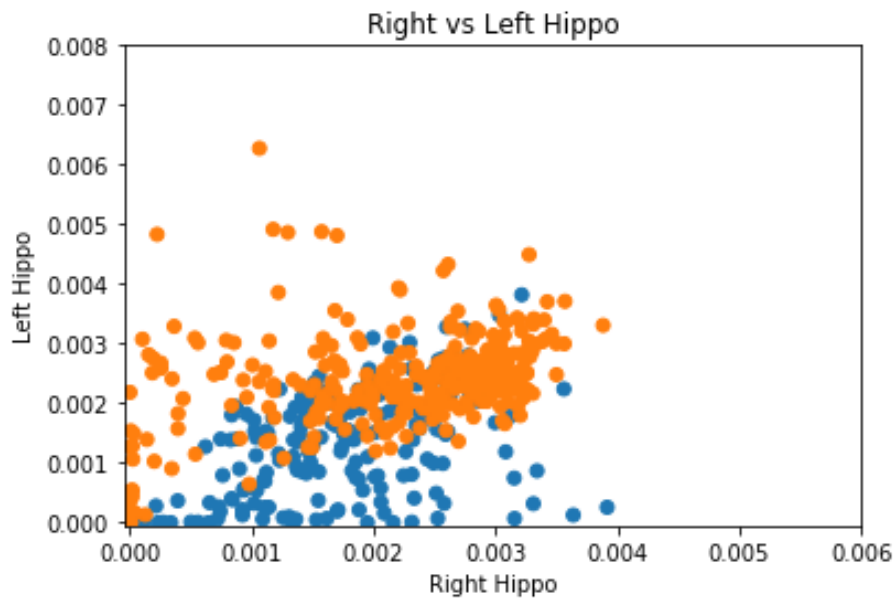
Πίνακας 5.4 Σπουδαιότητα των χαρακτηριστικών του ταξινομητή Random Forest

Left Amygdala	Left Ent	Left Hippo	Left PHG	Right Amygdala	Right Ent	Right Hippo	Right PHG
0,3409	0,0219	0,0681	0,1824	0,2314	0,0488	0,0505	0,0558

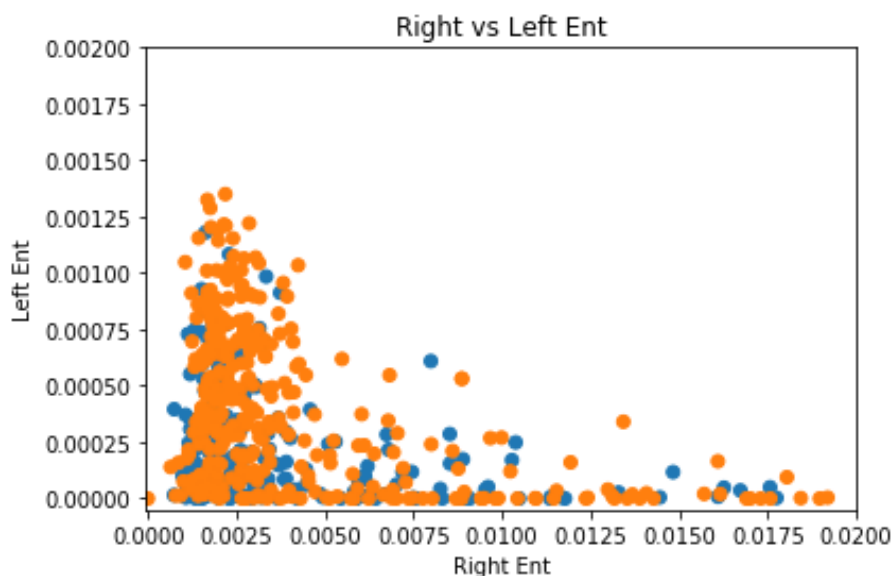
Πίνακας 5.5 Συντελεστές Lasso

### 5.3.3 Εμπειρική Αξιολόγηση

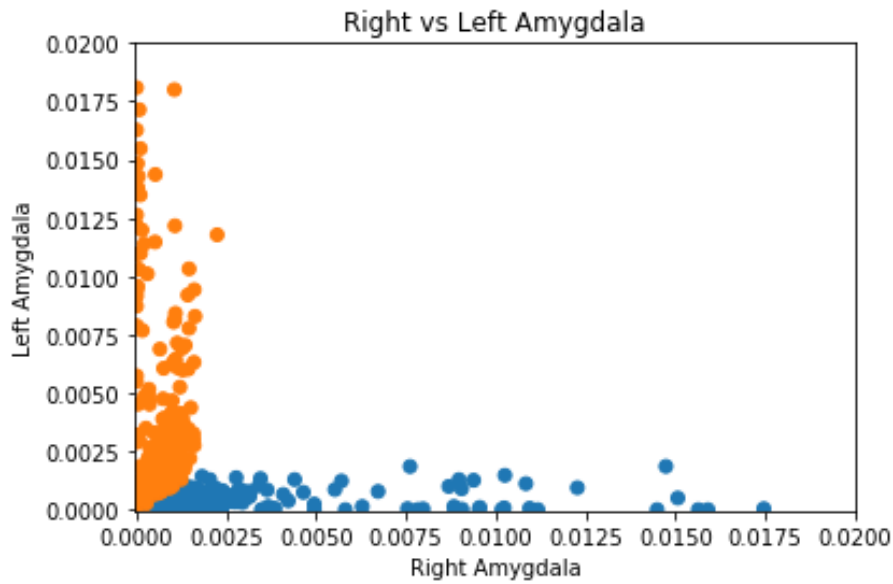
Ακολουθώς παρουσιάζονται διαγράμματα διασποράς για διάφορους συνδυασμούς των οχτώ χαρακτηριστικών, όπου και επιβεβαιώνονται οπτικά τα αποτελέσματα και οι παρατηρήσεις που προέκυψαν στην ανάλυση των προηγούμενων παραγραφών. Στους άξονες βρίσκονται οι τιμές των ποσοστών για την εκάστοτε περιοχή, όπως προέκυψαν από τη διαδικασία που περιγράφηκε στην παράγραφο 5.2. Με μπλε χρώμα αναπαριστώνται οι ασθενείς με τη νόσο του Alzheimer (AD) και με πορτοκαλί αυτοί χωρίς (Normal). Η επιθυμητή αναπαράσταση θα είχε την πορτοκαλί ομάδα (AD) με μεγαλύτερες τιμές και στους 2 άξονες από την μπλε (Normal) και ιδανικά θα ήταν και γραμμικά διαχωρίσιμες οι 2 ομάδες. Αρχικά εξετάστηκαν οι ίδιες περιοχές για αριστερό και δεξί ημισφαίριο.



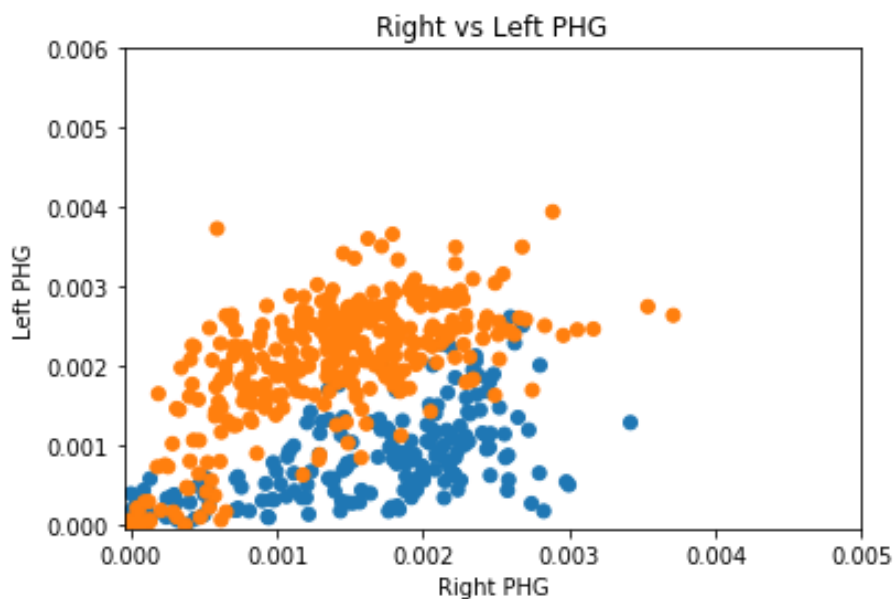
Σχήμα 5.1 Διάγραμμα διασποράς αριστερού μέρους ιππόκαμπου σε σχέση με το δεξί. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



Σχήμα 5.2 Διάγραμμα διασποράς αριστερού μέρους ενδορινικού φλοιού σε σχέση με το δεξί. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



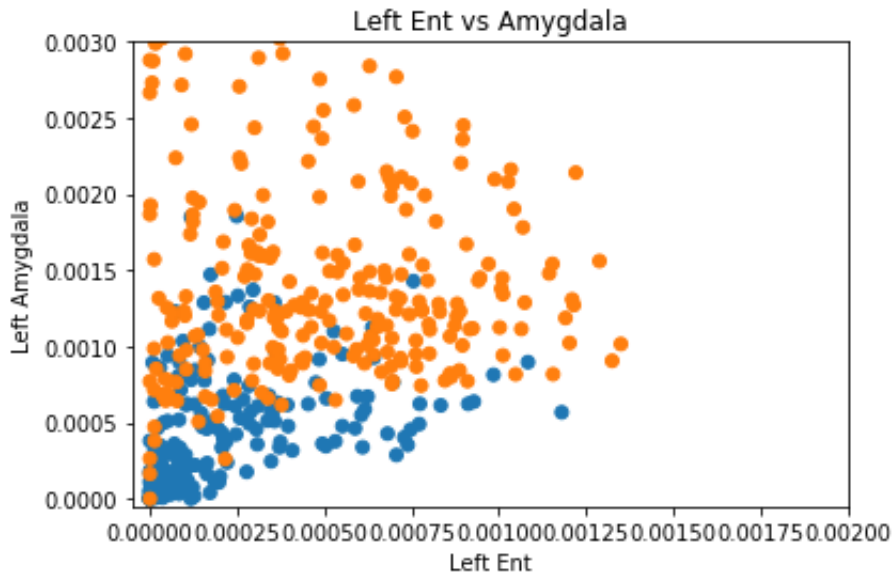
Σχήμα 5.3 Διάγραμμα διασποράς αριστερού μέρους αμυγδαλής σε σχέση με το δεξί. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



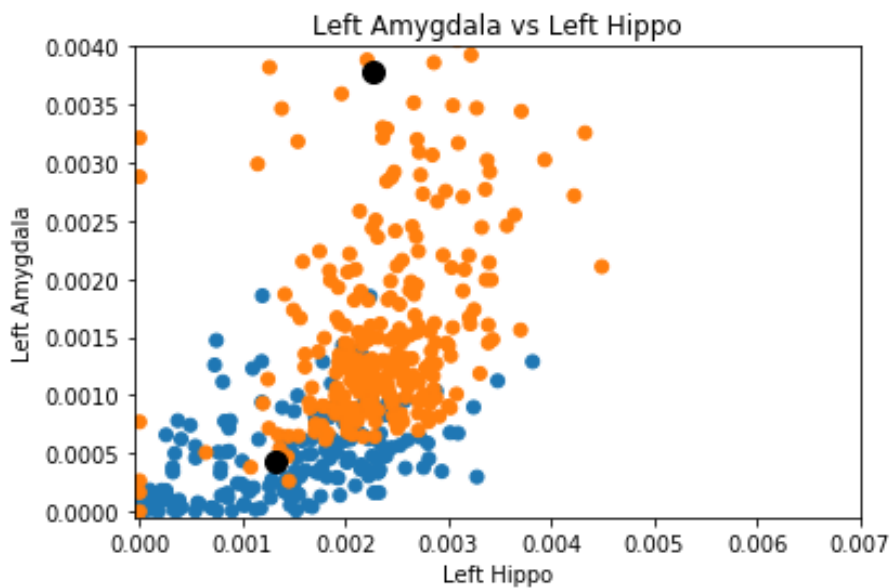
Σχήμα 5.4 Διάγραμμα διασποράς αριστερού μέρους πάρα-ιπποκάμπειας έλικας σε σχέση με το δεξί. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

Στα παραπάνω διαγράμματα παρατηρείται ότι με εξαίρεση την περιοχή του ενδορινικού φλοιού, σε όλες τις άλλες περιοχές υπάρχει εμφανής διαχωρισμός μεταξύ των AD και των Normal. Είναι επίσης εμφανές ότι οι αριστερές περιοχές είναι αυτές που παρουσιάζουν συρρίκνωση στους πάσχοντες, ενώ οι δεξιές έχουν το ίδιο μέγεθος και για τις δύο κατηγορίες εξεταζόμενων. Η μόνη περιοχή στην οποία σημειώνεται διαφορά και στο δεξί μέρος είναι αυτή της αμυγδαλής, όπου οι ασθενείς έχουν εμφανώς μεγαλύτερη τιμή, το οποίο θα αναλυθεί ακολούθως.

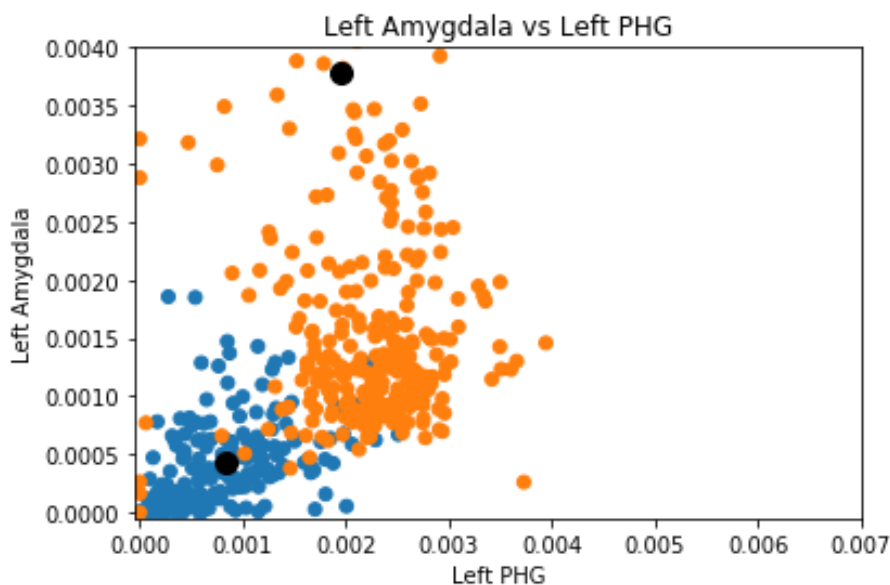
Επειδή οι αριστερές περιοχές φαίνεται να ατροφούν περισσότερο στους πάσχοντες, δηλαδή έχουν μικρότερες τιμές ποσοστών σε σχέση με τους μη πάσχοντες, στη συνέχεια παρουσιάζονται διαγράμματα ανά δύο περιοχές/χαρακτηριστικά μονό του αριστερού ημισφαιρίου.



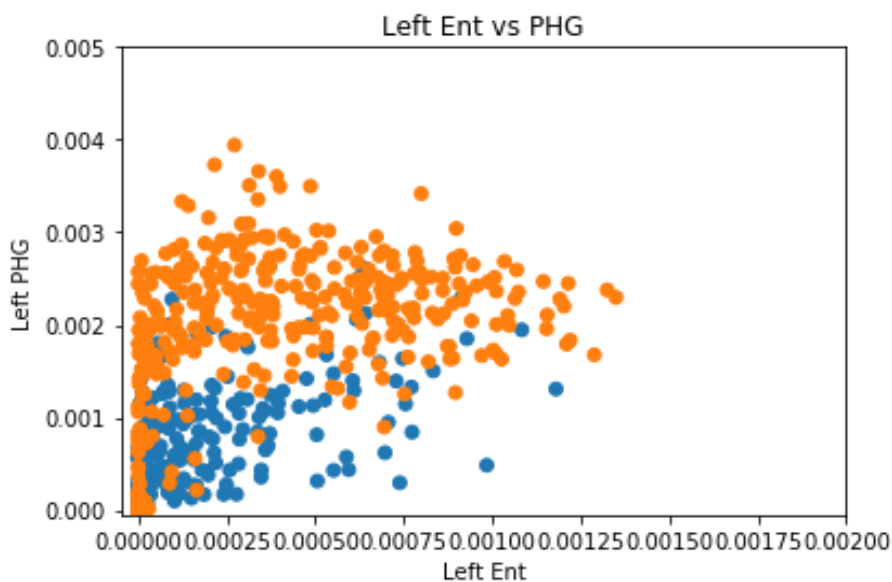
Σχήμα 5.5 Διάγραμμα διασποράς αριστερού μέρους αμυγδαλής σε σχέση με αυτό του αριστερού ενδορινικού φλοιού. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



Σχήμα 5.6 Διάγραμμα διασποράς αριστερού μέρους ιππόκαμπου σε σχέση με αυτό της αριστερής αμυγδαλής. Λόγω διαφοράς στο εύρος τιμών των δυο χαρακτηριστικών παρουσιάζεται εδώ μια μεγέθυνση του πραγματικού διαγράμματος στις τιμές που είναι διακριτός ο διαχωρισμός. Τα μαύρα σημεία παριστάνουν τη μέση τιμή για κάθε μια από τις δύο κατηγορίες ασθενών. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

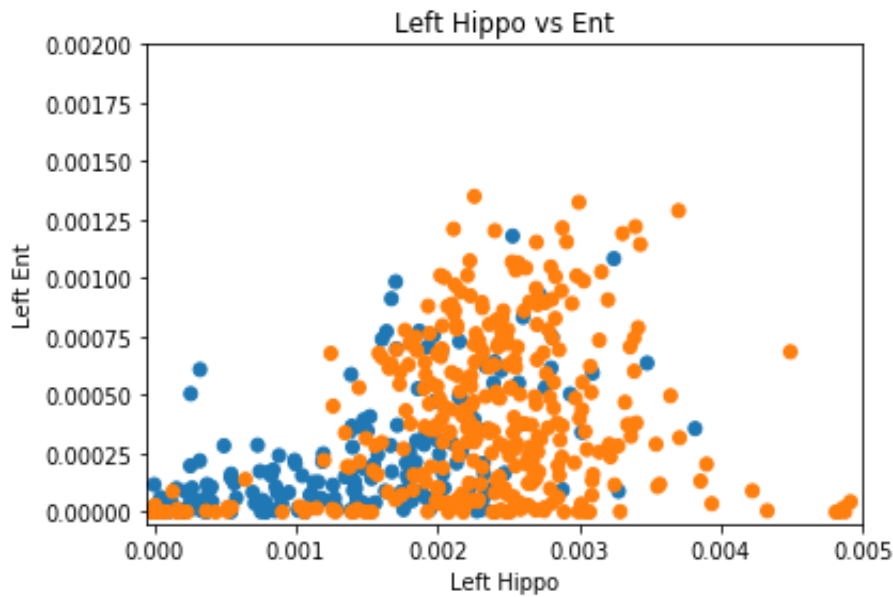


Σχήμα 5.7 Διάγραμμα διασποράς αριστερού μέρους αμυγδαλής σε σχέση με αυτό της αριστερής παρά-ιπποκάμπειας έλικας. Λόγω διαφοράς στο εύρος τιμών των δυο χαρακτηριστικών παρουσιάζεται εδώ μια μεγέθυνση του πραγματικού διαγράμματος στις τιμές που είναι διακριτός ο διαχωρισμός. Τα μαύρα σημεία παριστάνουν τη μέση τιμή για κάθε μια από τις δύο κατηγορίες ασθενών. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

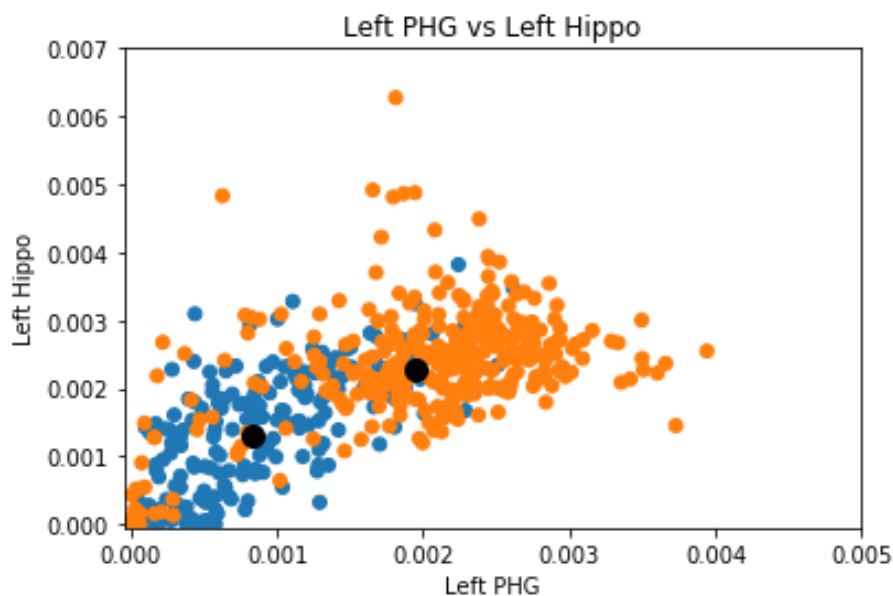


Σχήμα 5.8 Διάγραμμα διασποράς αριστερού μέρους παρά-ιπποκάμπειας έλικας σε σχέση με αυτό του αριστερού ενδορινικού φλοιού. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



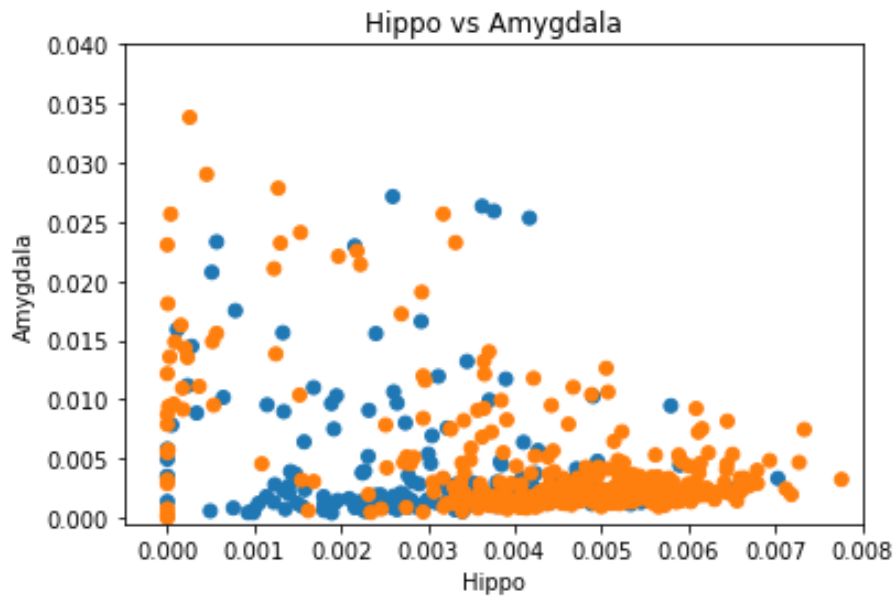


Σχήμα 5.9 Διάγραμμα διασποράς αριστερού μέρους ιππόκαμπου σε σχέση με αυτό του αριστερού ενδορινικού φλοιού. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

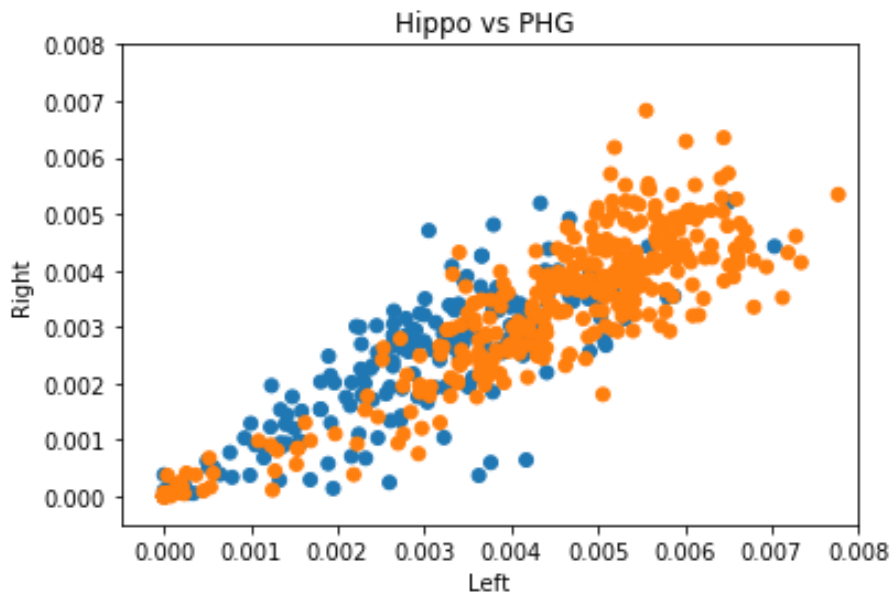


Σχήμα 5.10 Διάγραμμα διασποράς αριστερού μέρους ιππόκαμπου σε σχέση με αυτό της αριστερής παρά-ιπποκάμπειας έλικας. Τα μαύρα σημεία παριστάνουν τη μέση τιμή για κάθε μια από τις δύο κατηγορίες ασθενών. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

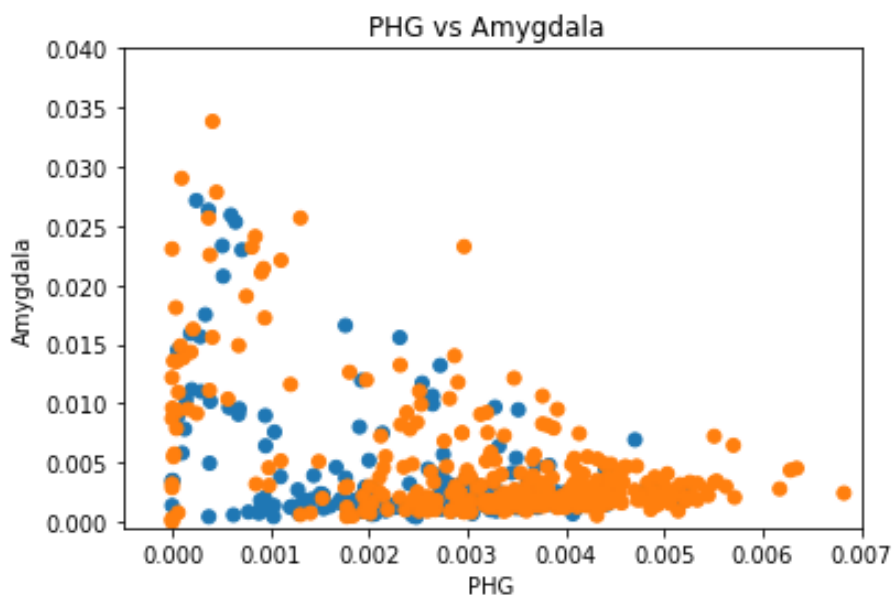
Όπως φαίνεται στα παραπάνω διαγράμματα, με εξαίρεση την αριστερή περιοχή του ενδορινικού φλοιού, σε όλες τις υπόλοιπες αριστερές περιοχές είναι εμφανής η συρρίκνωσή για τους πάσχοντες σε σχέση με τους μη πάσχοντες. Παρακάτω παρουσιάζονται και τα διαγράμματα ανά δυο περιοχές/χαρακτηριστικά στο σύνολό τους (άθροισμα αριστερής και δεξιάς περιοχής).



Σχήμα 5.11 Διάγραμμα διασποράς της συνολικής περιοχής του ιππόκαμπου σε σχέση με αυτήν της αμυγδαλής. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

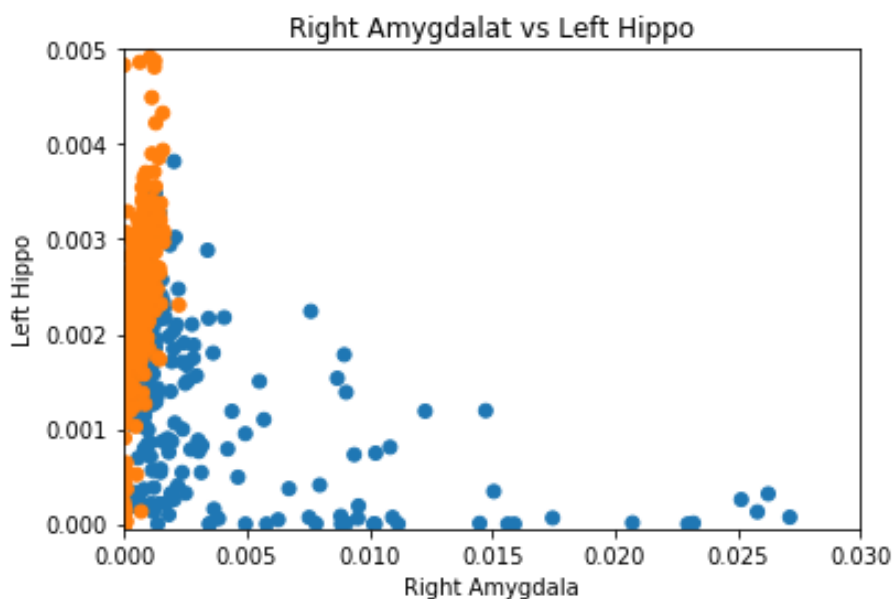


Σχήμα 5.12 Διάγραμμα διασποράς της συνολικής περιοχής ιππόκαμπου σε σχέση με αυτή της πάρα-ιπποκάμπειας έλικας. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

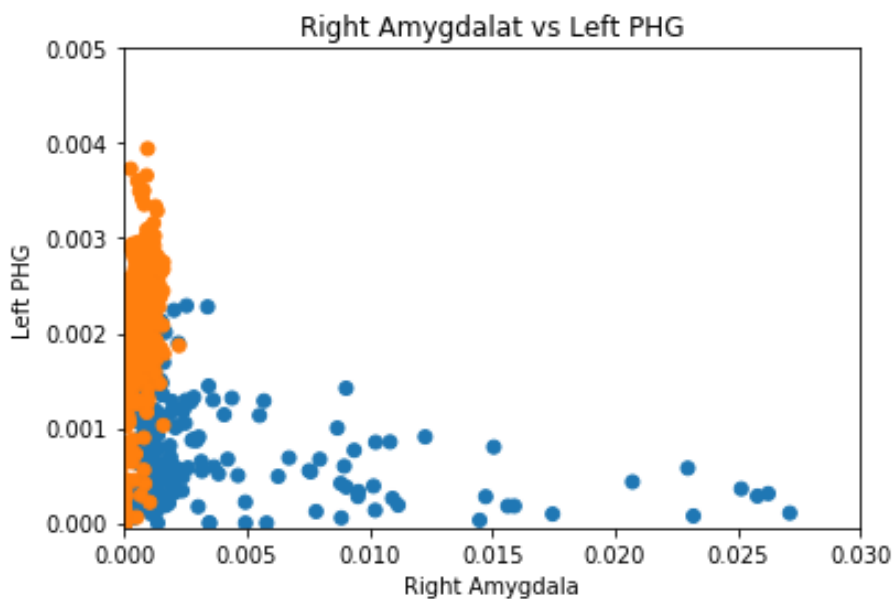


Σχήμα 5.13 Διάγραμμα διασποράς της συνολικής περιοχής της αμυγδαλής σε σχέση με αυτήν της παρά-ιπποκάμπειας έλικα. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

Με μια ενδελεχή παρατήρηση των προηγούμενων διαγραμμάτων προκύπτει ότι τα συμπεράσματα που εξάγονται είναι απολύτως αναμενόμενα με βάση τα προηγούμενα αποτελέσματα.. Πιο συγκεκριμένα έχει ενδιαφέρον να παρατηρήσουμε τα διαγράμματα στα οποία υπάρχει η περιοχή της αμυγδαλής. Όπως αναφέρθηκε και προηγουμένως, το δεξί μέρος της αμυγδαλής παρουσιάζει συρρίκνωση για τα Normal άτομα, και όχι για τους AD. Για το λόγο αυτό και στα σχήματα 5.11 και 5.13, στα οποία παρουσιάζεται το δεξί και το αριστερό μέρος της αμυγδαλής ενωμένα, δε διακρίνεται κάποιος διαχωρισμός. Για να διαπιστωθεί καλύτερα η συμπεριφορά του δεξιού μέρους της αμυγδαλής παρουσιάζονται παρακάτω δυο ακόμη διαγράμματα της.

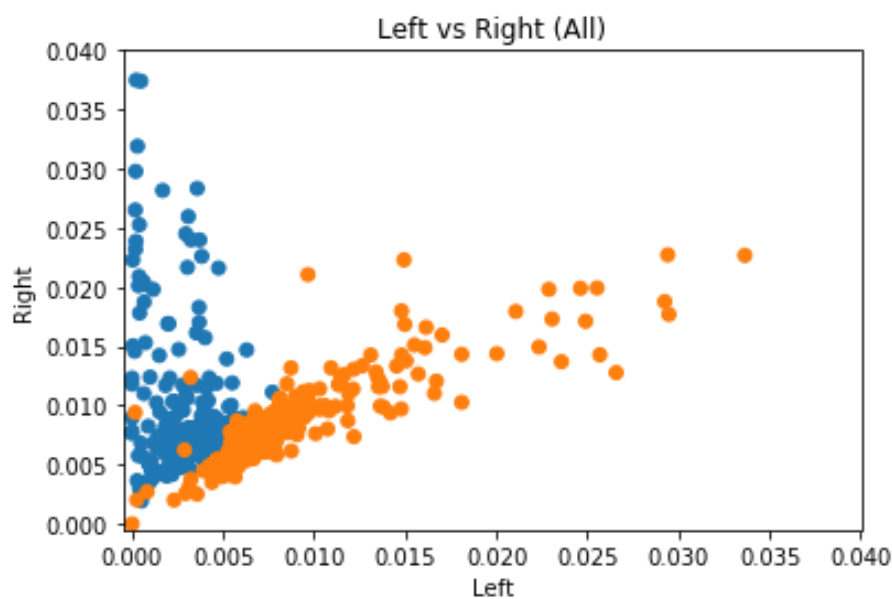


Σχήμα 5.14 Διάγραμμα διασποράς της δεξιάς περιοχής της αμυγδαλής σε σχέση με την αριστερή περιοχή του ιππόκαμπου. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.



Σχήμα 5.15 Διάγραμμα διασποράς της δεξιάς περιοχής της αμυγδαλής σε σχέση με την αριστερή περιοχή της πάρα-ιπποκάμπειας έλικας. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

Στα προηγούμενα διαγράμματα, καθώς και στο διάγραμμα 5.3 που παρουσιάστηκε σε προηγούμενη παράγραφο, γίνεται εμφανές ότι οι δύο κατηγορίες ασθενών είναι απόλυτα διαχωρίσιμες. Ωστόσο δε θεωρούμε αντιπροσωπευτική για το πείραμά μας την δεξιά περιοχή της αμυγδαλής, διότι με βάση τη βιβλιογραφία θα έπρεπε να υπάρχει συρρίκνωση για τους AD ασθενείς και όχι τους Normal. Η πιο πιθανή αιτιολόγηση για τη συμπεριφορά της συγκεκριμένης περιοχής είναι ότι οι ασθενείς που έχουν χαρακτηριστεί ως Normal πάσχουν από κάποια άλλη νόσο του επηρεάζει σε μεγάλο βαθμό το δεξί μέρος της αμυγδαλής.



Σχήμα 5.16 Διάγραμμα διασποράς όλων των αριστερών περιοχών σε σχέση με τις δεξιές. Με πορτοκαλί χρώματα φαίνονται οι Normal και με μπλε οι AD.

Στο τελευταίο σχήμα φαίνεται ότι η ατροφία εξαιτίας της νόσου συντελείται κατά κύριο λόγο στο αριστερό μέρος των περιοχών που εξετάστηκαν, όπως άλλωστε είχε παρατηρηθεί και νωρίτερα με βάση τα διαγράμματα των επιμέρους περιοχών. Επίσης, από τα διαγράμματα των σχημάτων 5.6, 5.7 και 5.10 είναι εμφανές ότι η αριστερή περιοχή της αμυγδαλής, της πάρα-ιπποκάμπειας έλικας και του ιππόκαμπου είναι αυτές που παρουσιάζουν τη μεγαλύτερη ατροφία. Στα συγκεκριμένα διαγράμματα έχουν σημειωθεί ως μαύρα σημεία οι μέσοι όροι για τα τους AD και Normal ασθενείς και μπορεί να διακριθεί, πως τόσο στον κάθετο όσο και στον οριζόντιο άξονα, οι ασθενείς της νόσου του Alzheimer λαμβάνουν μικρότερες τιμές από τους μη ασθενείς.



# Κεφάλαιο 6<sup>ο</sup>

## 6 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΟΙ ΣΤΟΧΟΙ

---

### 6.1 ΣΥΜΠΕΡΑΣΜΑΤΑ

Από την ποσοτική και ποιοτική αξιολόγηση των αποτελεσμάτων που παρήχθησαν στο πλαίσιο εκπόνησης της παρούσας εργασίας συμπεραίνεται ότι οι στόχοι που τέθηκαν αρχικά, για την αξιοποίηση τεχνικών Μηχανικής Μάθησης για την πρόβλεψη της νόσου Alzheimer, χρησιμοποιώντας εικόνες μαγνητικής τομογραφίας εγκεφάλου, επιτεύχθηκαν με υψηλά ποσοστά ακρίβειας. Πιο συγκεκριμένα συμπεραίνεται ότι το μοντέλο FC-DenseNet απέδωσε αποτελέσματα ακριβούς κατάτμησης εγκεφαλικών εικόνων και σε συνδυασμό με έναν απλό ταξινομητή μηχανικής μάθησης επιτεύχθηκε ο διαχωρισμός πασχόντων και μη από τη νόσο Alzheimer με ποσοστά ακρίβειας πάνω από 95%.

Ειδικότερα συμπεράσματα που εξήχθησαν αφορούν κατά κύριο λόγο την αρχιτεκτονική του δικτύου FC-DenseNet103 (tiramisu) και την υλοποίηση του στο συγκεκριμένο πλαίσιο. Συγκεκριμένα συμπεραίνεται ότι:

- Το συγκεκριμένο αρχιτεκτονικά μοντέλο αποδείχθηκε κατάλληλο, αποτελεσματικό και ακριβές όσον αφορά το πρόβλημα της κατάτμησης. Καταφέρνει να προβλέπει πολύ σωστά την κατάτμηση εικόνων και πιο συγκεκριμένα εικόνων εγκεφαλικής τομογραφίας, οι οποίες έχουν την ιδιαιτερότητα να συγκεντρώνουν την πληροφορία τους σε πολύ μικρό ποσοστό της συνολικής τους επιφάνειας.
- Λόγω της υπερίσχυσης της κλάσης υποβάθρου σε μία τέτοια εικόνα, η αρχική προσέγγιση απέτυχε στο να εκπαιδεύσει το μοντέλο, το οποίο προβλέπει σε όλα τα εικονοστοιχεία την τιμή μηδέν, δηλαδή την κλάση του υποβάθρου.
- Η μέθοδος που αγνοούσε την κλάση υποβάθρου βελτίωσε σημαντικά την απόδοση. Το μοντέλο εκπαιδεύεται με αρκετά μεγάλη επιτυχία χωρίς ωστόσο να πετυχαίνει ιδανικά αποτελέσματα.
- Για την παραπάνω μέθοδο ήταν απαραίτητη η χρήση κατώφλιου. Αποδείχθηκε ότι χαμηλές τιμές κατώφλιου υπερεκτιμούν τις κλάσεις, αποδίδοντας τους μεγαλύτερες περιοχές από ότι τους αναλογούν, ενώ οι πολύ υψηλές τιμές κάνουν το αντίθετο, αποδίδουν δηλαδή σε κάθε κλάση λιγότερα εικονοστοιχεία από όσα τους αναλογούν. Έτσι, και με βάση τα αποτελέσματα, καλύτερο κατώφλι αποδείχθηκε το 0.7.
- Η μέθοδος η οποία αναθέτει βάρη στις κλάσεις βελτιώνει ακόμα περισσότερο την απόδοση του μοντέλου και πετυχαίνει αποτελέσματα υψηλής ακρίβειας. Στις υπερέχουσες κλάσεις ανατίθενται μικρά βάρη, ενώ αντίθετα στις κλάσεις που εμφανίζονται σπάνια ανατίθενται μεγάλα βάρη, και με τον τρόπο αυτό επιτυγχάνεται εξισορρόπηση των δεδομένων.
- Οι μετρικές συναρτήσεις που χρησιμοποιήθηκαν δεν ήταν όλες αντιπροσωπευτικές για την απόδοση του μοντέλου. Η μετρική της ορθότητας (accuracy) πρακτικά δεν εξέφραζε την αληθινή διάσταση της απόδοσης του μοντέλου καθώς είναι ένα καθολικό μέτρο που το μέγεθος του επηρεάζεται περισσότερο από τις πολυπληθέστερες σε εικονοστοιχεία

κλάσεις, που σε αυτή την περίπτωση αφορούν την κλάση του υποβάθρου, που δεν αποτελεί κατηγορία ενδιαφέροντος. Οι μετρικές της ακρίβειας (precision) και της ανάκλησης (recall) κρίνονται λιγότερο αξιόπιστες αν εξεταστούν κάθε μια ατομικά, για αυτό προτείνεται να παρατηρηθούν σε συνδυασμό. Γι' αυτό το λόγο καταλληλότερες θεωρήθηκαν οι μετρικές, τομή προς ένωση (jaccard coefficient) και f1 (dice coefficient), οι οποίες συνδυάζουν τις δύο προηγούμενες, καθώς και η ορθότητα αγνοώντας την κλάση υποβάθρου (accuracy ignoring last label) η οποία μας δίνει σημαντικές πληροφορίες για την απόδοση του μοντέλου μόνο στις κλάσεις του εγκεφάλου.

Όσον αφορά το κομμάτι της εφαρμογής της κατάτμησης σε εικόνες εγκεφαλικής τομογραφίας διαπιστώθηκαν τα εξής :

- Η κατάτμηση των εικόνων εγκεφαλικής τομογραφίας του συνόλου δεδομένων του ADNI, έγινε επιτυχημένα, άρα το εκπαιδευμένο μοντέλο δείχνει να διαθέτει τη δυνατότητα γενίκευσης.
- Η κατάτμηση των εγκεφαλικών περιοχών βοήθησε αρκετά στην κατηγοριοποίηση των δεδομένων. Χρησιμοποιώντας τις περιοχές ενδιαφέροντος, και πιο συγκεκριμένα τα ποσοστά αναλογίας τους ως προς το σύνολο του εγκεφάλου, είδαμε ότι οι ασθενείς (AD) και μη (Normal) από τη νόσο του Alzheimer, μπορούν να διαχωριστούν με βάση τα ποσοστά αυτά.
- Η στατιστική ανάλυση των δεδομένων έδειξε ότι σε ασθενείς με Alzheimer ατροφούν κυρίως οι περιοχές της αμυγδαλής, της παραιπιποκάμπειας έλικας και του ιππόκαμπου.
- Η ανάλυση των δεδομένων έδειξε επίσης πως το αριστερό μέρος του εγκεφάλου ατροφεί περισσότερο σε σχέση με το δεξί σε ασθενείς με Alzheimer.

## 6.2 ΜΕΛΛΟΝΤΙΚΟΙ ΣΤΟΧΟΙ

Τέλος παρατίθενται ορισμένοι μελλοντικοί στόχοι για περαιτέρω ανάπτυξη της παρούσας έρευνας και εφαρμογής. Πιο συγκεκριμένα προτείνεται::

- Εφαρμογή του μοντέλου σε νέα δεδομένα για καλύτερη αξιολόγηση και δυνατότητα μεγαλύτερης γενίκευσης.
- Συνδυασμός των μεθόδων αγνόησης υποβάθρου και βαρών στις κλάσεις για καλύτερη προσαρμογή. Πιο συγκεκριμένα, αρχικοποίηση των βαρών με αυτά που έχουν προκύψει από τη μέθοδο αγνόησης υποβάθρου και μετέπειτα εκπαίδευση του μοντέλου με τη μέθοδο της εφαρμογής βαρών στις κλάσεις.
- Επέκταση του μοντέλου για εκπαίδευση σε τρισδιάστατα δεδομένα. Αντί η είσοδος να είναι μια εγκεφαλική τομή, θα είναι ένα σύνολο διαδοχικών εγκεφαλικών τομών και με τον τρόπο αυτό το μοντέλο θα έχει πληροφορίες για γειτονικά εικονοστοιχεία ως προς το βάθος, κάτι που δεν ισχύει με την δισδιάστατη μέθοδο.
- Διερεύνηση όλων των εγκεφαλικών περιοχών για την ταξινόμηση των ασθενών.
- Δοκιμή μοντέλου για ταξινόμηση ατόμων ήπιας νοητικής εξασθένησης (Mild Cognitive Impairment - MCI). Τα άτομα αυτά χωρίζονται σε 2 κατηγορίες και πιο συγκεκριμένα σε αυτούς που μετά από 18 μήνες εμφάνισαν τη νόσο και σε αυτούς που δεν την εμφάνισαν. Έτσι, αντί για την ταξινόμηση σε δύο κατηγορίες, θα επιχειρηθεί η ταξινόμηση σε τέσσερις.
- Δοκιμή του μοντέλου σε εικόνες πασχόντων από άλλες νευροεκφυλιστικές νόσους, όπως είναι η νόσος του Parkinson.



# Παράρτημα 1

Initial Class Num	Final Class Num	Class Name	Initial Class Num	Final Class Num	Class Name
0	0	Background	133	72	Left ITG inferior temporal gyrus
4	1	3rd Ventricle	134	73	Right LiG lingual gyrus
11	2	4th Ventricle	135	74	Left LiG lingual gyrus
15	3	5th Ventricle	136	75	Right LOrG lateral orbital gyrus
23	4	Right Accumbens Area	137	76	Left LOrG lateral orbital gyrus
30	5	Left Accumbens Area	138	77	Right MCGG middle cingulate gyrus
31	6	Right Amygdala	139	78	Left MCGG middle cingulate gyrus
32	7	Left Amygdala	140	79	Right MFC medial frontal cortex
35	8	Brain Stem	141	80	Left MFC medial frontal cortex
36	9	Right Caudate	142	81	Right MFG middle frontal gyrus
37	10	Left Caudate	143	82	Left MFG middle frontal gyrus
38	11	Right Cerebellum Exterior	144	83	Right MOG middle occipital gyrus
39	12	Left Cerebellum Exterior	145	84	Left MOG middle occipital gyrus
40	13	Right Cerebellum White Matter	146	85	Right MOrG medial orbital gyrus
41	14	Left Cerebellum White Matter	147	86	Left MOrG medial orbital gyrus
42	15	Right Cerebral Exterior	148	87	Right MPoG postcentral gyrus medial segment
43	16	Left Cerebral Exterior	149	88	Left MPoG postcentral gyrus medial segment
44	17	Right Cerebral White Matter	150	89	Right MPrG precentral gyrus medial segment
45	18	Left Cerebral White Matter	151	90	Left MPrG precentral gyrus medial segment
46	19	CSF	152	91	Right MSFG superior frontal gyrus medial segment
47	20	Right Hippocampus	153	92	Left MSFG superior frontal gyrus medial segment
48	21	Left Hippocampus	154	93	Right MTG middle temporal gyrus
49	22	Right Inf Lat Vent	155	94	Left MTG middle temporal gyrus
50	23	Left Inf Lat Vent	156	95	Right OCP occipital pole
51	24	Right Lateral Ventricle	157	96	Left OCP occipital pole
52	25	Left Lateral Ventricle	160	97	Right OFuG occipital fusiform gyrus
55	26	Right Pallidum	161	98	Left OFuG occipital fusiform gyrus
56	27	Left Pallidum	162	99	Right OpIFG opercular part of the inferior frontal gyrus
57	28	Right Putamen	163	100	Left OpIFG opercular part of the inferior frontal gyrus
58	29	Left Putamen	164	101	Right OrIFG orbital part of the inferior frontal gyrus
59	30	Right Thalamus Proper	165	102	Left OrIFG orbital part of the inferior frontal gyrus
60	31	Left Thalamus Proper	166	103	Right PCgG posterior cingulate gyrus
61	32	Right Ventral DC	167	104	Left PCgG posterior cingulate gyrus
62	33	Left Ventral DC	168	105	Right PCu precuneus
63	34	Right vessel	169	106	Left PCu precuneus
64	35	Left vessel	170	107	Right PHG parahippocampal gyrus
65	36	Left Insula	171	108	Left PHG parahippocampal gyrus
66	37	Right Insula	172	109	Right PIns posterior insula
69	38	Optic Chiasm	173	110	Left PIns posterior insula
71	39	Cerebellar Vermal Lobules I-V	174	111	Right PO parietal operculum
72	40	Cerebellar Vermal Lobules VI-VII	175	112	Left PO parietal operculum
73	41	Cerebellar Vermal Lobules VIII-X	176	113	Right PoG postcentral gyrus
74	42	Vitamin E Tablet	177	114	Left PoG postcentral gyrus
75	43	Left Basal Forebrain	178	115	Right POrg posterior orbital gyrus
76	44	Right Basal Forebrain	179	116	Left POrg posterior orbital gyrus
100	45	Right ACgG anterior cingulate gyrus	180	117	Right PP planum polare
101	46	Left ACgG anterior cingulate gyrus	181	118	Left PP planum polare
102	47	Right AIns anterior insula	182	119	Right PrG precentral gyrus
103	48	Left AIns anterior insula	183	120	Left PrG precentral gyrus
104	49	Right AOrG anterior orbital gyrus	184	121	Right PT planum temporale
105	50	Left AOrG anterior orbital gyrus	185	122	Left PT planum temporale
106	51	Right AnG angular gyrus	186	123	Right SCA subcallosal area
107	52	Left AnG angular gyrus	187	124	Left SCA subcallosal area
108	53	Right Calc calcarine cortex	190	125	Right SFG superior frontal gyrus
109	54	Left Calc calcarine cortex	191	126	Left SFG superior frontal gyrus
112	55	Right CO central operculum	192	127	Right SMC supplementary motor cortex
113	56	Left CO central operculum	193	128	Left SMC supplementary motor cortex
114	57	Right Cun cuneus	194	129	Right SMG supramarginal gyrus
115	58	Left Cun cuneus	195	130	Left SMG supramarginal gyrus
116	59	Right Ent entorhinal area	196	131	Right SOG superior occipital gyrus
117	60	Left Ent entorhinal area	197	132	Left SOG superior occipital gyrus
118	61	Right FO frontal operculum	198	133	Right SPL superior parietal lobule
119	62	Left FO frontal operculum	199	134	Left SPL superior parietal lobule
120	63	Right FRP frontal pole	200	135	Right STG superior temporal gyrus
121	64	Left FRP frontal pole	201	136	Left STG superior temporal gyrus
122	65	Right FuG fusiform gyrus	202	137	Right TMP temporal pole
123	66	Left FuG fusiform gyrus	203	138	Left TMP temporal pole
124	67	Right GRe gyrus rectus	204	139	Right TrIFG triangular part of the inferior frontal gyrus
125	68	Left GRe gyrus rectus	205	140	Left TrIFG triangular part of the inferior frontal gyrus
128	69	Right IOG inferior occipital gyrus	206	141	Right TTrG transverse temporal gyrus
129	70	Left IOG inferior occipital gyrus	207	142	Left TTrG transverse temporal gyrus
132	71	Right ITG inferior temporal gyrus			

Πίνακας με τις τιμές των αρχικών κλάσεων και αυτών στις οποίες τις αντιστοιχίζουμε, καθώς και οι ονομασίες τους.

## 7 ΒΙΒΛΙΟΓΡΑΦΙΑ

---

- [1] L. Coe, *The Telegraph: A History of Morse's Invention and Its Predecessors in the United States*, 2003.
- [2] A. Hodges, *Alan Turing: The Enigma*, Walker Books, 2000.
- [3] T. M. Mitchell, «The Discipline of Machine Learning, volume 9.,» *Carnegie Mellon University, School of Computer Science, Machine Learning Department*, 2006.
- [4] S. J. Russell και P. Norvig, «Intelligent agents,» σε *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995, pp. 31-52.
- [5] P. C. CHEN και T. PAVLIDIS, «Image Segmentation as an Estimation Problem\*,» *Department of Electrical Engineering and Computer Science, New Jersey*, 1980.
- [6] S. Yuheng και Y. Hao, «Image Segmentation Algorithms Overview,» *SiChuan University, SiChuan, ChengDu*, 2017.
- [7] Y. Luo, K. Liu και D. N. Davis, «A multi-agent decision support system for stock trading,» *IEEE Network*, pp. 20 - 27, 2002 .
- [8] Igor Kononenko, «Machine learning for medical diagnosis: history, state of the art and perspective,» *Artificial Intelligence in Medicine*, pp. 89-109, August 2001.
- [9] G. D. Magoulas και A. Prentza, *Machine Learning in Medical Applications*, 2001.
- [10] L. N. Robert και C. E. Ellis, «Alzheimer's Disease and Parkinson's Disease,» *The NEW ENGLAND JOURNAL of MEDICINE*, April 2003.
- [11] S. Feng, L. Bing, Z. Yuan, Y. Chunshui και J. Tianzi, «Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies,» *Hippocampus*, March 2009.
- [12] Y. LeCun, Y. Bengio και G. Hinton, «Deep learning,» *Nature* 521, pp. 436-444, 28 May 2015.
- [13] M. Campbell, A. J. J. Hoane και F.-h. Hsu, «Deep Blue,» *Elsevier*, 2002.
- [14] KERAS, «Keras Documentation,» [Ηλεκτρονικό]. Available: <https://keras.io/>.
- [15] F. ROSENBLATT, «THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN,» *Psychological Review*, 1958.
- [16] K. Hirasawa, M. Ohbayashi, M. Koga και M. Harada, «Forward propagation universal learning network,» *Proceedings of International Conference on Neural Networks (ICNN'96)*, 1996.
- [17] Y. Bengio, P. Simard και P. Frasconi, «Learning long-term dependencies with gradient descent is difficult,» *IEEE Transactions on Neural Networks*, pp. 157 - 166.

- [18] A. L., A. Maas, Y. Hannun και A. Y. Ng, «Rectifier Nonlinearities Improve Neural Network Acoustic Models,» *International conference on Machine Learning, Atlanta, Georgia, USA, 2013.*
- [19] D. E. Rumelhart, G. E. Hinton και R. J. Williams, «Learning representations by back-propagating errors,» *nature*, 1986.
- [20] P.-T. d. Boe, D. P. Kroese, S. MannorReuven και Y. Rubinstein, «A Tutorial on the Cross-Entropy Method,» *Ann Oper Res*, 2005.
- [21] S. Ruder, «An overview of gradient descent optimization algorithms,» 2016. [Ηλεκτρονικό]. Available: <http://ruder.io/optimizing-gradient-descent/index.html>.
- [22] N. Qian, «On the momentum term in gradient descent learning algorithms,» *Neural Networks : The Official Journal of the International Neural Network Society 12(1)*, 1999.
- [23] J. Duchi, E. Hazan και Y. Singer, «Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,» *Journal of Machine Learning Research*, 2011.
- [24] G. E. Hinton, 2013. [Ηλεκτρονικό]. Available: [https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf).
- [25] D. P. Kingma και J. L. Ba, «Adam: a method for stochastic optimization,» *arXiv preprint arXiv:1412.6980*, 2014.
- [26] K. Simonyan και A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition,» *CoRR, abs/1409.1556*, 2014.
- [27] K. He, X. Zhang, S. Ren και J. Sun, «Deep residual learning for image.,» *Proceedings of the IEEE conference on computer vision and pattern recognition,,* pp. 770-778, 2016.
- [28] G. Huang, Z. Liu, L. v. d. Maaten και K. Q. Weinberger, «Densely connected convolutional networks,» *CoRR, abs/1608.06993*, 2016.
- [29] S. J'egou, M. Drozdal, D. Vazquez, A. Romero και Y. Bengio, «The one hundred layer tiramisu. Fully convolutional densenets for semantic segmentation.,» *arXiv:1611.09326v3*, 2017.
- [30] «Brainder,» 2012. [Ηλεκτρονικό]. Available: <https://brainder.org/2012/09/23/the-nifti-file-format/>.
- [31] D. P. Devanand, G. Pradhaban, A. K. X. Liu, S. D. Santi, S. Segal, H. Rusinek, G. H. Pelton, L. S. Honig, R. Mayeux, Y. Stern, M. H. Tabert και M. J. d. Leon, «Hippocampal and entorhinal atrophy in mild cognitive impairment,» *Neurology*, 2007.
- [32] M. P. Laakso, H. Soininen, K. Partanen, E. -L. Helkala, P. Hartikainen, P. Vainio, M. Hallikainen, T. Hänninen και P. J. R. Sr, «Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: Correlation with memory functions,» *Journal of Neural Transmission - Parkinson's Disease and Dementia Section*, 1995.
- [33] P. RC, B. L. Aisen PS, D. MC, H. D. Gamst AC, J. C. Jr, J. WJ, S. LM, T. AW, T. JQ και W. MW, «Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization.,» *Neurology*, pp. 201-9, 2010.

- [34] S. D. Gupta, «Point biserial correlation coefficient and its generalization,» *Psychometrika*, p. 393–408, 1960.
- [35] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim και S. Serikawa, «Motor Anomaly Detection for Unmanned Aerial Vehicles Using Reinforcement Learning,» *IEEE Internet of Things Journal*, 2017.
- [36] S. Ruder, «An overview of gradient descent optimization,» *Insight Centre for Data Analytics, NUI Galway*, 2017.
- [37] F. Rosenblatt, «Principles of neurodynamics; perceptrons and the theory of brain mechanisms.,» *Technical report, CORNELL AERONAUTICAL LAB INC*, 1961.