



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Χρήση τεχνικών βαθιάς μάθησης και
επεξηγήσιμης τεχνητής νοημοσύνης για
την ανίχνευση της νόσου του Alzheimer
σε εικόνες μαγνητικής τομογραφίας
εγκεφάλου

Διπλωματική Εργασία
του
Ιάσονα Σκυλίτση

Επιβλέποντες
Αθανάσιος Βουλόδημος, Επίκουρος Καθηγητής
Ευτύχιος Πρωτοπαπαδάκης, Υπότροφος ΕΣΠΑ

Αθήνα, Ιούλιος 2022



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Χρήση τεχνικών βαθιάς μάθησης και επεξηγήσιμης τεχνητής νοημοσύνης για την ανίχνευση της νόσου του Alzheimer σε εικόνες μαγνητικής τομογραφίας εγκεφάλου

Όνοματεπώνυμο Ιάσων Σκυλίτσης

Αριθμός Μητρώου 71151097

Επιβλέποντες: Αθανάσιος Βουλόδημος, Επίκουρος Καθηγητής

Ευτύχιος Πρωτοπαπαδάκης, Υπότροφος ΕΣΠΑ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Ευτύχιος Πρωτοπαπαδάκης
Υπότροφος ΕΣΠΑ

Χρήστος Τρούσσας
Επίκουρος Καθηγητής

Παναγιώτα Τσελέντη
ΕΔΙΠ

Αθήνα, Ιούλιος 2022



UNIVERSITY OF WEST ATTICA

School of Engineering

Department of Informatics and Computer Engineering

Deep learning and Explainable Artificial Intelligence techniques for Alzheimer's disease detection on magnetic resonance images

Full Name Iason Skylitsis

Identification Number 711151097

Supervisors: Athanasios Voulodimos, Assistant Professor

Eftychios Protopapadakis, Adjunct Lecturer

Athens, July 2022

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος **Σκυλίτσης Ιάσων** του Άγγελου, με αριθμό μητρώου **711151097** φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του **Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών**, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου»

Ο Δηλών,

Ιάσων Σκυλίτσης



Στη μνήμη του πατέρα μου
In loving memory of my father

Περίληψη

Η νόσος του Alzheimer είναι μια προοδευτική νευροεκφυλιστική νόσος και αποτελεί μια από τις σημαντικότερες προκλήσεις του 21^{ου} αιώνα για τον υγειονομικό τομέα. Τα συμπτώματα της νόσου αφορούν διαταραχές στη μνήμη, την διάθεση και την προσωπικότητα και επηρεάζουν την ποιότητα ζωής τόσο αυτών που νοσούν, όσο και των κοντινών τους ανθρώπων που τους φροντίζουν.

Η διάγνωση της νόσου προϋποθέτει την κλινική εξέταση των ασθενών από ειδικούς γιατρούς σε συνδυασμό με μια σειρά από ειδικές εξετάσεις και διαγνωστικά τεστ. Η απεικονιστική εξέταση μέσω μαγνητικής τομογραφίας εγκεφάλου αποτελεί αναπόσπαστο κομμάτι της κλινικής αξιολόγησης ασθενών με υποψία νόσου του Alzheimer. Οι δείκτες της μαγνητικής τομογραφίας βοηθούν στην έγκαιρη και ακριβή διάγνωση καθώς και στη μέτρηση της προοδευτικότητας της νόσου.

Η ερμηνεία των ιατρικών εικόνων πραγματοποιείται κυρίως σε κλινικό περιβάλλον από ειδικευμένους γιατρούς ή ακτινολόγους. Αυτή είναι μια χρονοβόρος και απαιτητική διαδικασία λόγω του μεγάλου αριθμού παραλλαγών στην παθολογία, αλλά και επιρρεπής σε σφάλματα λόγω του ανθρώπινου παράγοντα.

Στην παρούσα εργασία θα υλοποιηθούν μέθοδοι βαθιάς μάθησης και επεξηγήσιμης τεχνητής νοημοσύνης με σκοπό την αναγνώριση της νόσου σε εικόνες μαγνητικής τομογραφίας εγκεφάλου. Η αναγνώριση της νόσου του Alzheimer με μεθόδους βαθιάς μάθησης θα μπορούσε να υποβοηθήσει την διαδικασία λήψης αποφάσεων των γιατρών, ενώ η επεξηγησιμότητα αποτελεί βασική προϋπόθεση για την εμπιστοσύνη και την αποδοχή τέτοιων συστημάτων από τους ασθενείς.

Λέξεις Κλειδιά

Βαθιά μάθηση, Ερμηνεύσιμη Τεχνητή Νοημοσύνη, Μαγνητική Τομογραφία Εγκεφάλου, Μεταφορά Μάθησης, Συνελκτικά Νευρωνικά Δίκτυα, Ιατρικές Εικόνες, Αναγνώριση της νόσου του Alzheimer

Abstract

Alzheimer's Disease is a neurodegenerative disease and one of the great health-care challenges of the 21st century. The symptoms of Alzheimer's disease include memory loss, mood and personality changes and have an impact not only to the ones suffering, but also to their loved ones.

Diagnosis of Alzheimer disease involves the clinical examination of the patients and various tests for brain function. Imaging has a key role in the clinical assessment of patients with suspected Alzheimer's disease. Magnetic Resonance Imaging (MRI) markers help measure disease progression as well as in early and accurate diagnosis.

The interpretation of medical images is mainly carried out in a clinical setting by qualified doctors or radiologists. This is a time-consuming and demanding process due to the large number of variations in the pathology, but also susceptible to human error.

In this thesis, deep learning and explainable artificial intelligence methods will be implemented in order to detect Alzheimer's disease in brain MRI. The automated detection of Alzheimer's disease using deep learning methods could support the decision-making process of doctors, while explainability is a basic prerequisite for the trust and acceptance of such systems by patients.

Key words

Deep Learning, Explainable Artificial Intelligence, Magnetic Resonance Imaging, Transfer Learning, Convolutional Neural Networks, CNN, Medical Images, Alzheimer Detection

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον κ. Βουλόδημο, ο οποίος προσπαθούσε συνεχώς να καλλιεργήσει την επιστημονική μας σκέψη και να μεταδώσει τις γνώσεις του. Τον ευχαριστώ θερμά για την βοήθεια και τις συμβουλές που μου έδωσε καθ' όλη την διάρκεια των σπουδών μου.

Επίσης ευχαριστώ τον κ. Πρωτοπαπαδάκη για την εισαγωγή σε θέματα υπολογιστικής όρασης και μηχανικής μάθησης, καθώς και για την προθυμία του να λύσει οποιαδήποτε απορία μου.

Θα ήθελα ακόμη να ευχαριστήσω την οικογένεια μου και ιδιαίτερα την μητέρα μου η οποία αποτελεί τον πιο ένθερμο υποστηρικτή κάθε μου προσπάθειας. Την ευχαριστώ για την αμέριστη συμπαράσταση και στήριξη της.

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τους φίλους μου, που ήταν εκεί για μένα στις χαρές και στις λύπες. Ειδική μνεία στην Ο.Π.Ι.Α.Β.

Τέλος αυτή η εργασία είναι αφιερωμένη στη μνήμη του πατέρα μου, που τα τελευταία χρόνια έπασχε από άνοια. Η επιλογή του θέματος προέκυψε από την επιθυμία μου να κατανοήσω καλύτερα τη νόσο του Alzheimer και να βοηθήσω όσο μπορώ στην συλλογική ερευνητική προσπάθεια για την αντιμετώπιση της.

Περιεχόμενα

Περίληψη	8
Abstract	10
Ευχαριστίες	12
Περιεχόμενα.....	13
Κατάλογος Εικόνων	15
Κατάλογος Πινάκων	16
1. Εισαγωγή	17
1.1 Μια σύνοψη της νόσου Alzheimer	17
1.2 Ανάλυση Ιατρικών Εικόνων με τεχνικές Βαθιάς Μάθησης (Deep Learning)	21
1.3 Στόχος της εργασίας.....	22
1.4 Δομή εργασίας	22
2. Θεωρητικό Υπόβαθρο.....	23
2.1 Μεταφορά Μάθησης (Transfer Learning)	23
2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI).....	26
3. Σχεδιασμός και Υλοποίηση	29
3.1 Δομή του συστήματος.....	29
3.1.1 Δομή του συστήματος διάγνωσης του Alzheimer	29
3.1.2 Δομή του συστήματος επεξήγησης των αποτελεσμάτων.....	29
3.2 Δεδομένα.....	31
3.2.1 Προ-επεξεργασία των Δεδομένων	32
3.2.2 Μετατροπή σε 2D και διαχωρισμός των δεδομένων (dataset split).....	34
3.3 Μοντέλα.....	38
3.3.1 VGG16.....	38
3.3.2 ResNet50.....	39
3.3.3 Xception	40
3.3.4 EfficientNetV2M	41
3.4 Εκπαίδευση των μοντέλων.....	42
3.5 Μετρικές αξιολόγησης των αποτελεσμάτων	44
3.5.1 Confusion Matrix (Πίνακας Σύγκρισης)	44
3.5.2 Καμπύλη ROC - AUC	45
3.5.3 Precision – Recall Curve (Καμπύλη Ακρίβειας – Ανάκλησης)	46
4. Πειραματικά Αποτελέσματα	47
4.1 Slice Level Split.....	47
4.1.1 Μετρικές Αξιολόγησης	47
4.1.2 Πίνακες σύγκρισης (Confusion Matrix).....	49

4.1.3 Καμπύλη ROC (Receiver Operating Characteristic Curve).....	51
4.1.4 Καμπύλη Ακρίβειας-Ανάκλησης (Precision-Recall Curve)	53
4.2 Διαχωρισμός σε επίπεδο υποκειμένου (Subject level split).....	56
4.2.1 Υγιείς έναντι νοσούντων (CN vs AD)	56
4.2.2 Υγιείς έναντι νοσούντων με ήπια γνωστική διαταραχή (CN vs MCI).....	59
4.2.3 Νοσούντες με ήπια γνωστική διαταραχή έναντι νοσούντων με Alzheimer (MCI vs AD) ..	62
5. Αποτελέσματα Επεξηγησιμότητας (Slice Level Split)	65
5.1. Κλάση CN.....	65
5.1.1. Σωστές ταξινομήσεις.....	65
5.1.2. Λανθασμένες ταξινομήσεις.....	66
5.2. Κλάση MCI.....	67
5.2.1. Σωστές Ταξινομήσεις.....	67
5.2.2 Λανθασμένες Ταξινομήσεις.....	69
5.3. Κλάση AD.....	70
5.3.1. Σωστές Ταξινομήσεις.....	70
5.3.2. Λανθασμένες Ταξινομήσεις.....	71
6. Επίλογος.....	73
6.1 Σύνοψη.....	73
6.2 Τελικά Συμπεράσματα	74
6.3 Μελλοντικές Προεκτάσεις.....	75
Βιβλιογραφία	76
Παράρτημα Α: Αποτελέσματα Επεξηγησιμότητας (Subject Level Split)	82
A.1 Σωστές ταξινομήσεις CN.....	82
A.2 Λανθασμένες ταξινομήσεις CN	83
A.3 Σωστές Ταξινομήσεις MCI	84
A.4 Λανθασμένες Ταξινομήσεις MCI	85
A.5 Σωστές ταξινομήσεις AD.....	86
A.6 Λανθασμένες Ταξινομήσεις AD	87

Κατάλογος Εικόνων

Εικόνα 1.1 Προτεινόμενη προσέγγιση για τη διάγνωση και τη διαχείριση της ήπιας γνωστικής διαταραχής ¹¹	18
Εικόνα 1.2: Προοδευτική ατροφία (έσω κροταφικών λοβών) σε ένα ηλικιωμένο άτομο γνωστικά φυσιολογικό (CN), ένα άτομο με αμνησιακή ήπια γνωστική διαταραχή (aMCI), και ένα άτομο με νόσο του Alzheimer (AD). ¹⁴	19
Εικόνα 1.3: Αλλαγές του εγκεφάλου στην νόσο Alzheimer ⁶	20
Εικόνα 1.4: Απεικόνιση πλακών αμυλοειδούς με χρώση αργύρου από περιοχή εγκεφαλικού φλοιού ασθενούς με νόσο του Alzheimer ¹⁵	20
Εικόνα 1.5: Ενδοκυττάριας εναποθέσεις νευροϊνδικών σωρών σε περιοχή του φλοιού του εγκεφάλου ασθενούς με νόσο του Alzheimer ανοσοχρωματισμένα για tau ¹⁵	20
Εικόνα 2.1: Μεταφορά μάθησης βάση του ορισμού τομέα και ζητούμενου ³⁵	24
Εικόνα 2.2: Διαδικασία της προσαρμογής (fine-tuning) ³⁷	25
Εικόνα 2.3: Χάρτες Ενεργοποίησης Κλάσης ⁴⁹	28
Εικόνα 2.4: Αρχιτεκτονική του Grad-CAM	28
Εικόνα 3.1: Δομή του συστήματος ταξινόμησης	30
Εικόνα 3.2: Δομή του συστήματος επεξήγησης των αποτελεσμάτων	30
Εικόνα 3.3: Αναπαράσταση Voxel ⁵²	32
Εικόνα 3.4: Βήματα προεπεξεργασίας των δεδομένων	32
Εικόνα 3.5: Διαδικασία καταχώρησης (register) στον atlas	33
Εικόνα 3.6 Εγκάρσιο, στεφανιαίο και οβελιαίο επίπεδο, αντιστοίχως, μιας καταχωρημένης εικόνας στην οποία έχει αφαιρεθεί το κρανίο	34
Εικόνα 3.7: Διάγραμμα της επεξεργασίας των εικόνων και του διαχωρισμού σε δεδομένα εκπαίδευσης και ελέγχου ⁶⁶	35
Εικόνα 3.8: Αποτελέσματα μετατροπής των MRI σε 2D εικόνες	36
Εικόνα 3.9: Αρχιτεκτονική VGG16 ⁷⁴	38
Εικόνα 3.10: Η δομή ενός Residual block ⁴¹	39
Εικόνα 3.11: Depthwise separable convolution στο Xception ⁷⁵	40
Εικόνα 3.12: Κλιμάκωση πλάτος, βάθους και ομοιόμορφη κλιμάκωση ⁴⁴	41
Εικόνα 3.13: Τυχαίοι μετασχηματισμοί data augmentation	43
Εικόνα 3.14: Μεταφορά μάθησης και fine-tuning στο πρόβλημα του Alzheimer Classification	43
Εικόνα 3.15: Confusion Matrix ⁸⁵	44
Εικόνα 3.16: Επεξήγηση της καμπύλης ROC ⁹⁰	46
Εικόνα 3.17: Επεξήγηση της καμπύλης Precision-Recall ⁹⁰	46
Εικόνα 4.1: Πίνακας Σύγκρισης VGG16	49
Εικόνα 4.2: Πίνακας Σύγκρισης ResNet50	49
Εικόνα 4.3: Πίνακας Σύγκρισης Xception	50
Εικόνα 4.4: Πίνακας Σύγκρισης EfficientNetV2M	50
Εικόνα 4.5: Καμπύλη ROC για το μοντέλο VGG16	51
Εικόνα 4.6: Καμπύλη ROC για το μοντέλο ResNet50	52
Εικόνα 4.7: Καμπύλη ROC για το μοντέλο Xception	52
Εικόνα 4.8: Καμπύλη ROC για το μοντέλο EfficientNetV2M	53
Εικόνα 4.9: Καμπύλη Precision-Recall για το μοντέλο VGG16	53
Εικόνα 4.10: Καμπύλη Precision-Recall για το μοντέλο ResNet50	54

Εικόνα 4.11: Καμπύλη Precision-Recall για το μοντέλο Xception	54
Εικόνα 4.12: Καμπύλη Precision-Recall για το μοντέλο EfficientNetV2M.....	55
Εικόνα 4.13: Συγκριτική αξιολόγηση μετρικών για CN vs AD	56
Εικόνα 4.14: Πίνακες σύγχυσης για CN vs AD.....	57
Εικόνα 4.15: Συγκριτική αξιολόγηση καμπυλών ROC για CN vs AD.....	58
Εικόνα 4.16: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για CN vs AD	58
Εικόνα 4.17: Συγκριτική αξιολόγηση μετρικών για CN vs MCI.....	59
Εικόνα 4.18: Πίνακες σύγχυσης για CN vs MCI.....	60
Εικόνα 4.19: Συγκριτική αξιολόγηση καμπυλών ROC για CN vs MCI.....	61
Εικόνα 4.20: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για CN vs MCI.....	61
Εικόνα 4.21: Συγκριτική αξιολόγηση μετρικών για MCI vs AD.....	62
Εικόνα 4.22 Πίνακες σύγχυσης για MCI vs AD.....	63
Εικόνα 4.23: Συγκριτική αξιολόγηση καμπυλών ROC για MCI vs AD.....	64
Εικόνα 4.24: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για MCI vs AD.....	64

Κατάλογος Πινάκων

Πίνακας 3.1: Κατανομή των ιατρικών εικόνων σε κλάσεις.....	31
Πίνακας 3.2: Κατανομή των subject splitted εικόνων σε κλάσεις.....	37
Πίνακας 3.3:Κατανομή των εικόνων σε δεδομένα εκπαίδευση, επικύρωσης και ελέγχου.....	37
Πίνακας 3.4: Πλήθος παραμέτρων των μοντέλων που χρησιμοποιήθηκαν.....	38

1. Εισαγωγή

1.1 Μια σύνοψη της νόσου Alzheimer

Η νόσος του Alzheimer είναι η κύρια αιτία της άνοιας και αποτελεί μια από τις σημαντικότερες προκλήσεις του 21^{ου} αιώνα για τον υγειονομικό τομέα.

Τον Δεκέμβριο του 2013, η G8 δήλωσε ότι η άνοια πρέπει να γίνει παγκόσμια προτεραιότητα, με σκοπό την ανάπτυξη θεραπείας για την νόσο μέχρι το 2025.¹ Περίπου 40 εκατομμύρια άνθρωποι, ηλικίας κυρίως άνω των 60 ετών, έχουν άνοια παγκοσμίως και αυτός ο αριθμός προβλέπεται να διπλασιάζεται κάθε 20 χρόνια τουλάχιστον μέχρι το 2050.² Στην Ελλάδα το 2007 περίπου 140 χιλιάδες ασθενείς έπασχαν από την νόσο και υπολογίζεται ότι το 2040 ο αριθμός αυτός θα φτάσει τις 560 χιλιάδες.³

Στις 3 Νοεμβρίου του 1906, ο Alois Alzheimer παρουσίασε για πρώτη φορά τα κλινικά και νευροπαθολογικά χαρακτηριστικά της νόσου, η οποία στη συνέχεια πήρε το όνομα του.⁴ Η νόσος Alzheimer είναι μια προοδευτική νευροεκφυλιστική νόσος, η οποία χαρακτηρίζεται από την συσσώρευση δυο παθολογικών πρωτεϊνών, του β αμυλοειδούς (amyloid β) και της τ πρωτεΐνης (tau protein), τα οποία οδηγούν προοδευτικά στην απώλεια νευρώνων και την ατροφία του εγκεφάλου.⁵

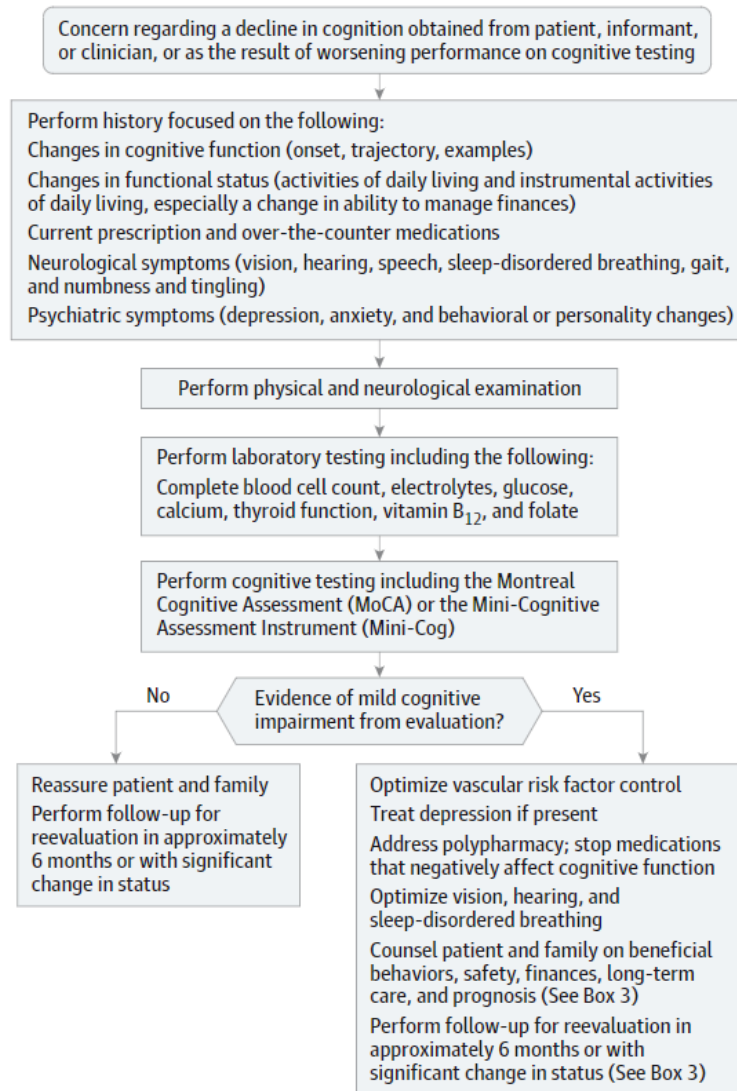
Η έναρξη της νόσου Alzheimer είναι αργή και προοδευτική, με τα συμπτώματα να εκδηλώνονται με διαφορετικό τρόπο σε κάθε άτομο και να επιδεινώνονται με την πρόοδο της νόσου. Τα συμπτώματα της νόσου μπορεί να είναι από ήπια μέχρι πολύ σοβαρά. Τα ήπια συμπτώματα περιλαμβάνουν προβλήματα μνήμης και ανάκλησης γεγονότων που έγιναν πρόσφατα (συνήθως αποτελεί το πιο πρώιμο σύμπτωμα), λανθασμένη τοποθέτηση αντικειμένων, απώλεια προσανατολισμού στο χώρο και χρόνο, δυσκολία στην ολοκλήρωση πολύπλοκων εργασιών όπως η πληρωμή λογαριασμών και μεταβολές στην προσωπικότητα και την διάθεση. Η εξέλιξη της νόσου οδηγεί σε επιδείνωση των διαταραχών στην μνήμη, τη διάθεση και την προσωπικότητα, που μπορούν να εκδηλωθούν σαν εχθρότητα ή παράνοια, μη αναγνώριση οικείων προσώπων, και ανάγκη βοήθειας για καθημερινές δραστηριότητες όπως το ντύσιμο ή το μπάνιο. Τέλος τα σοβαρότερα συμπτώματα της νόσου περιλαμβάνουν την απώλεια της ικανότητας ομιλίας και την πλήρη εξάρτηση από τους άλλους για όλες τις καθημερινές λειτουργίες.⁶

Η διάγνωση της νόσου Alzheimer συνήθως πραγματοποιείται από τον γιατρό πρωτοβάθμιας φροντίδας ενός ατόμου. Η διάγνωση βασίζεται στο ιατρικό και οικογενειακό ιστορικό του ασθενή, σε συνδυασμό με το ψυχιατρικό ιστορικό καθώς και το ιστορικό αλλαγών στο γνωσιακό και συμπεριφορικό επίπεδο. Είναι σύνηθες, ο γιατρός να ζητά επιπλέον πληροφορίες από κάποιο μέλος της οικογένειας ή άλλο κοντινό πρόσωπο, ούτως ώστε να σχηματίσει μια πιο ολοκληρωμένη εικόνα για τον ασθενή και την κατάσταση του.⁷ Επιπλέον, καθοριστικό παράγοντα για την αποτελεσματική διάγνωση, αποτελεί ο έλεγχος της νοητικής λειτουργίας του ασθενή μέσω ειδικά διαμορφωμένων διαγνωστικών τεστ όπως το Mini-Mental State Examination (MMSE) και το Montreal Cognitive Assessment (MoCA). Αυτά τα τεστ είναι ιδιαίτερα αποτελεσματικά εργαλεία στην διαδικασία της διάγνωσης και είναι σημαντικό να χρησιμοποιούνται σε πληθυσμούς με αυξημένο κίνδυνο για άνοια, όπως οι ηλικιωμένοι ασθενείς και εκείνοι που εκφράζουν παράπονα για διαταραχές στη μνήμη.^{8,9}

Πιο συγκεκριμένα το MoCA¹⁰, σχεδιάστηκε ειδικά για την ανίχνευση της ήπιας μορφής της νόσου, η οποία ονομάζεται Ήπια Γνωστική Διαταραχή (Mild Cognitive Impairment, MCI), και χρειάζεται περίπου 10 λεπτά για να ολοκληρωθεί. Το συγκεκριμένο τεστ παρουσιάζει ευαισθησία (sensitivity) από 80% έως 100% και ειδικότητα (specificity) από 50% έως 76% στην ανίχνευση MCI. Από την άλλη το MMSE παρουσιάζει τιμές ευαισθησίας από 45% έως 60% και ειδικότητας από 65% έως 90% στη ανίχνευση της MCI. Ένα επιπλέον τεστ που χρησιμοποιείται συχνά είναι το Mini-Cog, το οποίο παρέχει αρκετά αξιόπιστα αποτελέσματα και μπορεί να ολοκληρωθεί σε λιγότερο από 3 λεπτά.¹¹

Όλοι οι ασθενείς με υποψία MCI θα πρέπει να υποβληθούν σε πλήρες ιστορικό και φυσική εξέταση με επίκεντρο τη γνωστική λειτουργία, τη λειτουργική κατάσταση, τα φάρμακα, τις νευρολογικές ή ψυχιατρικές ανωμαλίες και τις εργαστηριακές εξετάσεις. Οι κύριοι στόχοι είναι η διάκριση της MCI

από τη φυσιολογική γήρανση ή άνοια και η αναγνώριση πιθανών αναστρέψιμων μορφών MCI που οφείλονται σε άλλες καταστάσεις π.χ. κατάθλιψη, παρενέργειες φαρμάκων, νόσος του θυρεοειδούς και ανεπάρκεια βιταμίνης B12/φυλλικού οξέος.¹¹ Στην **Εικόνα 1.1** παρουσιάζεται ένα διάγραμμα ροής που περιγράφει μια προτεινόμενη προσέγγιση για την διάγνωση και διαχείριση ενός ασθενή που παρουσιάζει ήπια γνωστική διαταραχή.



Εικόνα 1.1 Προτεινόμενη προσέγγιση για τη διάγνωση και τη διαχείριση της ήπιας γνωστικής διαταραχής¹¹

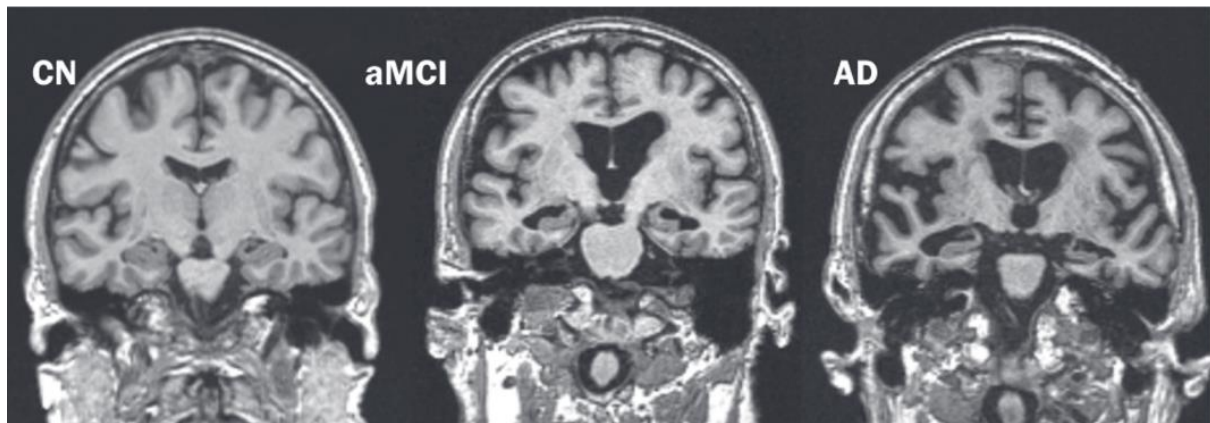
Οι ασθενείς που παρουσιάζουν ενδείξεις γνωστικής εξασθένησης θα πρέπει να υποβληθούν σε απεικονιστικές εξετάσεις. Αυτό συνήθως πραγματοποιείται με αξονική τομογραφία εγκεφάλου (computed tomography, CT) ή μαγνητική τομογραφία εγκεφάλου (magnetic resonance imaging, MRI). Αυτό έχει ως σκοπό να αξιολογηθούν μεταβολές στον εγκεφαλο που συνάδουν με τη νόσο Alzheimer και να αποκλειστούν εναλλακτικές αιτίες της γνωστικής εξασθένησης.^{1,6,12}

Η απεικονιστική εξέταση μέσω μαγνητικής τομογραφίας εγκεφάλου αποτελεί αναπόσπαστο κομμάτι της κλινικής αξιολόγησης ασθενών με υποψία νόσου Alzheimer. Αρκετές μελέτες έχουν δείξει ότι οι εκτιμήσεις μέσω MRI για βλάβη ή απώλεια ιστού σε χαρακτηριστικά ευάλωτες περιοχές του εγκεφάλου, όπως ο ιππόκαμπος και ο ενδορινικός φλοιός, είναι προγνωστικές για την εξέλιξη από MCI σε Alzheimer's Disease (AD). Επιπλέον, έχει τεκμηριωθεί η κλινική χρησιμότητα της μαγνητικής τομογραφίας στη διαφοροποίηση της AD από άλλες παθολογίες, όπως ο νευροεκφυλισμός που

οφείλεται σε αγγειακούς ή άλλους παθολογικούς λόγους που δεν σχετίζονται με το Alzheimer. Τέλος μέσω της μαγνητικής τομογραφίας μπορεί να γίνει εκτίμηση της εξέλιξης της νόσου π.χ. μέσω του ποσοστού ατροφίας που παρατηρείται στον εγκέφαλο.¹³

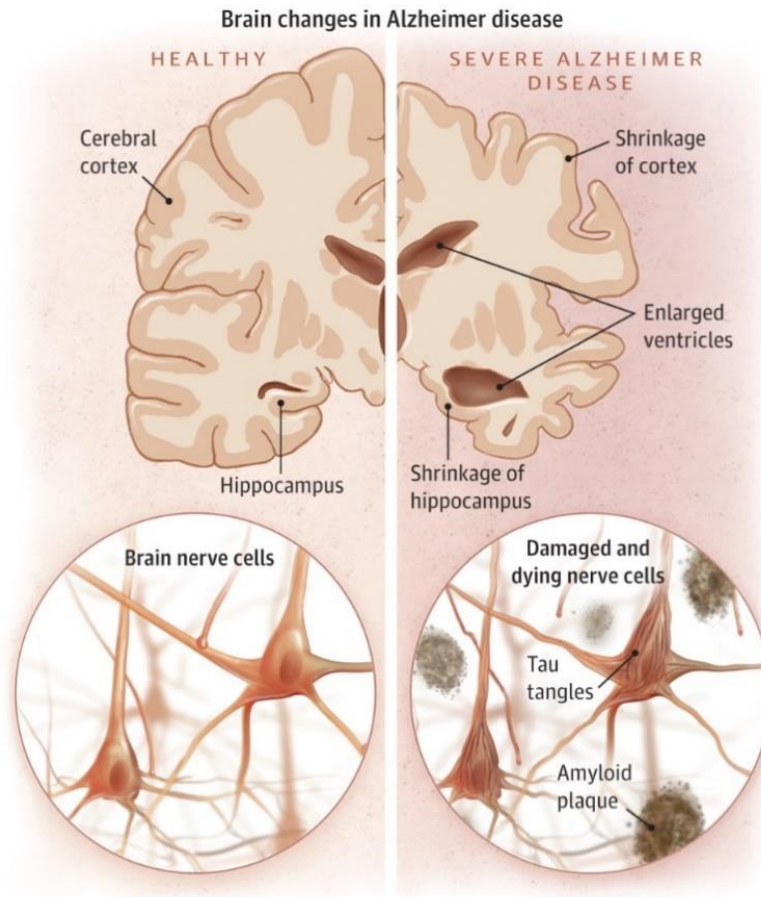
Οι δείκτες της μαγνητικής τομογραφίας βοηθούν στην έγκαιρη και ακριβή διάγνωση καθώς και στη μέτρηση της εξέλιξης της νόσου. Η παρουσία ατροφίας στις περιοχές του κροταφικού λοβού του εγκεφάλου αποτελεί έναν από τους υποψήφιους δείκτες για την έγκαιρη διάγνωση της νόσου στο στάδιο MCI. Επίσης τα ποσοστά ατροφίας ολόκληρου του εγκεφάλου και του ιππόκαμπου αποτελούν ευαίσθητους και ισχυρούς δείκτες της εξέλιξης του νευροεκφυλισμού και χρησιμοποιούνται συχνά, σε συνδυασμό με κλινικές μετρήσεις, για τον έλεγχο των αποτελεσμάτων σε κλινικές δοκιμές πιθανών θεραπειών τροποποίησης της νόσου. Τέλος οι μετρήσεις που αφορούν την ατροφία του φλοιού του εγκεφάλου και οι αυτοματοποιημένες προσεγγίσεις ταξινόμησης που αξιολογούν το συνολικό μοτίβο της ατροφίας φαίνεται να δείχνουν πολλά υποσχόμενες για τη διάγνωση της AD.¹³

Στην **Εικόνα 1.2** παρουσιάζονται ενδεικτικά παραδείγματα σαρώσεων MRI σε γνωστικά φυσιολογικά άτομα (Cognitive Normal, CN) και σε ασθενείς με MCI ή AD. Όπως φαίνεται στην εικόνα, υπάρχει αυξημένη έσω κροταφική ατροφία (πιο συγκεκριμένα παρατηρείται ατροφία στην περιοχή του ιππόκαμπου και διεύρυνση των κοιλιών του εγκεφάλου) στις σαρώσεις που αφορούν το MCI και AD σε σύγκριση με αυτή που αφορά το CN.¹⁴

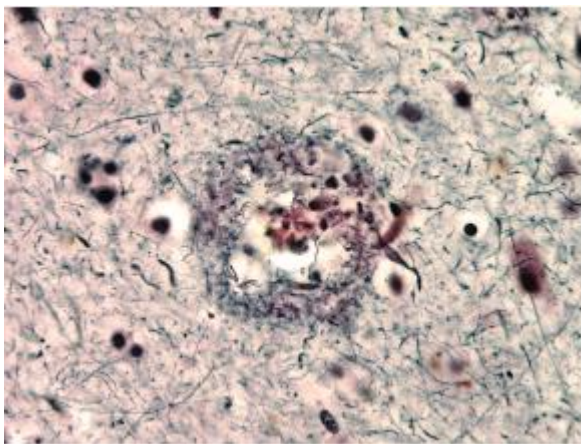


Εικόνα 1.2: Προοδευτική ατροφία (έσω κροταφικών λοβών) σε ένα ηλικιωμένο άτομο γνωστικά φυσιολογικό (CN), ένα άτομο με αμνησιακή ήπια γνωστική διαταραχή (aMCI), και ένα άτομο με νόσο του Alzheimer (AD).¹⁴

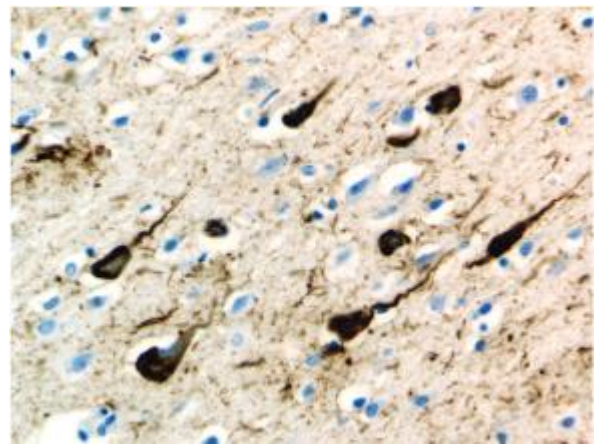
Ο εγκέφαλος των ασθενών με νόσο του Alzheimer παρουσιάζει νευρωνική απώλεια και συρρίκνωση των μεγάλων νευρώνων του φλοιού με συναπτική απώλεια. Αυτή η απώλεια κυττάρων έχει ως αποτέλεσμα τη γενικευμένη ατροφία του φλοιού του εγκεφάλου, η οποία παρατηρείται σε μεγάλα δείγματα ιστού με διεύρυνση των αυλάκων, ατροφία των ελίκων, λέπτυνση του φλοιού και αντιροπιστική διεύρυνση των κοιλιών, που παρατηρούνται στον εγκέφαλο στα τελευταία στάδια της νόσου (**Εικόνα 1.3**). Η νόσος του Alzheimer χαρακτηρίζεται ιστολογικά από την εναπόθεση συσσωματωμάτων πρωτεϊνών, οι οποίες σχηματίζουν δυο ιδιαίτερες δομές υπό το μικροσκόπιο: τις πλάκες αμυλοειδούς (amyloid plaques) και τους νευροϊνιδικούς σωρούς (neurofibrillary tangles). (**Εικόνες 1.4 και 1.5**). Οι πλάκες αμυλοειδούς αποτελούνται από εναποθέσεις ενός πεπτιδίου που ονομάζεται β-αμυλοειδές. Οι πλάκες αυτές βρίσκονται ανάμεσα στα νευρικά κύτταρα και σχετίζονται με τις βιοχημικές διαδικασίες επεξεργασίας της πρόδρομης πρωτεΐνης του αμυλοειδούς (amyloid precursor protein). Οι νευροϊνιδιακοί σωροί (neurofibrillary tangles), αποτελούνται από την πρωτεΐνη tau (ή αλλιώς πρωτεΐνη “τ”) και βρίσκονται εντός των νευρικών κυττάρων. Η πρωτεΐνη tau είναι πολύ σημαντική για τη λειτουργία των νευρώνων και εμπλέκεται σε πολλές νευροεκφυλιστικές νόσους. Αυτές οι αλλοιώσεις ξεκινούν από συγκεκριμένα σημεία του εγκεφάλου και επεκτείνονται σε παρακείμενες περιοχές καθώς η νόσος εξελίσσεται.¹⁵



Εικόνα 1.3: Αλλαγές του εγκεφάλου στην νόσο Alzheimer⁶



Εικόνα 1.4: Απεικόνιση πλακών αμυλοειδούς με χρώση αργύρου από περιοχή εγκεφαλικού φλοιού ασθενούς με νόσο του Alzheimer¹⁵



Εικόνα 1.5: Ενδοκυττάρια εναποθέσεις νευροϊνιδικών σωρών σε περιοχή του φλοιού του εγκεφάλου ασθενούς με νόσο του Alzheimer ανοσοχρωματισμένα για tau¹⁵

1.2 Ανάλυση Ιατρικών Εικόνων με τεχνικές Βαθιάς Μάθησης (Deep Learning)

Τις τελευταίες δεκαετίες τεχνικές ιατρικής απεικόνισης, όπως η αξονική τομογραφία (CT), η μαγνητική τομογραφία (MRI), η τομογραφία εκπομπής ποζιτρονίων (Positron Emission Tomography, PET), η μαστογραφία, το υπερηχογράφημα και οι ακτίνες Χ, χρησιμοποιούνται για την έγκαιρη ανίχνευση, διάγνωση και θεραπεία διαφόρων ασθενειών. Η ερμηνεία των ιατρικών εικόνων πραγματοποιείται κυρίως σε κλινικές, από ειδικευμένο προσωπικό όπως ακτινολόγους και γιατρούς. Παρόλα αυτά, λόγω μεγάλου αριθμού παραλλαγών στην παθολογία σε συνδυασμό με την πιθανή κούραση του ειδικευμένου ανθρώπινου δυναμικού, ερευνητές και γιατροί έχουν αρχίσει να επωφελούνται από παρεμβάσεις υποβοηθούμενες από υπολογιστή. Αν και ο ρυθμός προόδου στην υπολογιστική ανάλυση ιατρικών εικόνων δεν ήταν τόσο γρήγορος όσο αυτός στις τεχνολογίες ιατρικής απεικόνισης, η κατάσταση βελτιώθηκε με την εισαγωγή τεχνικών μηχανικής μάθησης (machine learning).

Για την εφαρμογή μεθόδων μηχανικής μάθησης σε εργασίες ανάλυσης ιατρικών εικόνων, καθοριστικό ρόλο παίζει η εύρεση ή εκμάθηση χρήσιμων χαρακτηριστικών ικανών να αναγνωρίζουν μοτίβα που παρουσιάζονται στα δεδομένα. Για δεκαετίες, η κατασκευή ενός συστήματος αναγνώρισης προτύπων ή μηχανικής μάθησης απαιτούσε προσεκτικό σχεδιασμό των χαρακτηριστικών σε συνδυασμό με σημαντική τεχνογνωσία του τομέα εφαρμογής. Τα χαρακτηριστικά αυτά σχεδιάζονταν, κατά κανόνα, από ειδικούς του εκάστοτε τομέα εφαρμογής και κατά συνέπεια όποιος δεν ήταν ειδικός στο πεδίο εφαρμογής θα δυσκολευόταν να εφαρμόσει μεθόδους μηχανικής μάθησης. Ωστόσο, η βαθιά μάθηση (deep learning)¹⁶ ξεπέρασε αυτό το εμπόδιο ενσωματώνοντας το κομμάτι της εξαγωγής χαρακτηριστικών (feature engineering) στο κομμάτι της εκπαίδευσης. Δηλαδή, αντί η εξαγωγή των χαρακτηριστικών να γίνεται χειροκίνητα, τα μοντέλα βαθιάς μάθησης χρειάζονται μόνο ένα σύνολο δεδομένων με μικρή προεπεξεργασία, εάν αυτή κρίνεται απαραίτητη, και στη συνέχεια μπορούν και ανακαλύπτουν τις αναπαραστάσεις που εμπεριέχουν πληροφορία με αυτοδίδακτο τρόπο. Κατά συνέπεια, το βάρος της εξαγωγής χαρακτηριστικών έχει μετατοπιστεί από τους ανθρώπους στους υπολογιστές, επιτρέποντας στους μη ειδικούς στη μηχανική μάθηση να χρησιμοποιούν αποτελεσματικά τη βαθιά μάθηση για τη δική τους έρευνα ή/και εφαρμογές, ειδικά στην ανάλυση ιατρικών εικόνων.¹⁷

Η άνευ προηγουμένου επιτυχία της βαθιάς μάθησης οφείλεται κυρίως στους ακόλουθους παράγοντες: (α) πρόοδος στις κεντρικές μονάδες επεξεργασίας (CPU) και στις μονάδες επεξεργασίας γραφικών (GPU), (β) στη διαθεσιμότητα τεράστιου όγκου δεδομένων (big data) και (γ) εξελίξεις στους αλγόριθμους εκπαίδευσης.¹⁸⁻²⁰ Η εξαγωγή των χαρακτηριστικών στα δίκτυα βαθιάς μάθησης γίνεται ιεραρχικά, δηλαδή τα χαρακτηριστικά υψηλότερου επιπέδου προέρχονται από χαρακτηριστικά χαμηλότερου επιπέδου. Έτσι τα δίκτυα βαθιάς μάθησης έχουν καταφέρει κορυφαίες επιδόσεις σε μια ποικιλία εφαρμογών τεχνητής νοημοσύνης και μεγάλων διαγωνισμών.^{21,22} Πιο συγκεκριμένα, η πρόοδος της όρασης υπολογιστών, οδήγησε στην χρήση βαθιάς μάθησης στην ανάλυση ιατρικών εικόνων με εφαρμογές στην κατάτμηση εικόνας (image segmentation)²³, την καταχώρηση εικόνας (image registration)²⁴, την συγχώνευση εικόνας (image fusion)²⁵, το σχολιασμό εικόνας (image annotation)²⁶, τη διάγνωση με τη βοήθεια υπολογιστή (Computer Aided Diagnosis, CADx) και πρόγνωση^{27,28}, την ανίχνευση βλάβης (lesion detection)²⁹ και μικροσκοπική ανάλυση εικόνας³⁰.

1.3 Στόχος της εργασίας

Στόχος της παρούσας εργασίας είναι η ανάπτυξη μοντέλων βαθιάς μάθησης με σκοπό την αναγνώριση της νόσου του Alzheimer από εικόνες μαγνητικής τομογραφίας εγκεφάλου. Πιο συγκεκριμένα, θα πραγματοποιηθούν συγκριτικές μελέτες 4 διαφορετικών δικτύων βαθιάς μάθησης τα οποία έχουν εκπαιδευτεί χρησιμοποιώντας την μέθοδο της μετάδοσης μάθησης (transfer learning). Η απόδοση των μοντέλων θα εξεταστεί χρησιμοποιώντας γνωστές μετρικές αξιολόγησης καθώς και διαγραμματικές τεχνικές.

Ο δεύτερος στόχος της εργασίας αυτής είναι η εφαρμογή μεθόδων ερμηνεύσιμης τεχνητής νοημοσύνης (Explainable AI) με σκοπό την επεξήγηση των αποτελεσμάτων που προκύπτουν από αυτά τα μοντέλα. Αυτό θα επιτευχθεί χρησιμοποιώντας μια επέκταση των Χαρτών Ενεργοποίησης Κλάσης (Class Activation Maps, CAM), το Grad-CAM.

1.4 Δομή εργασίας

Η παρούσα διπλωματική εργασία είναι οργανωμένη σε 6 κεφάλαια. Στο παρόν κεφάλαιο γίνεται μια εισαγωγή στην νόσο του Alzheimer και στην ανάλυση ιατρικών εικόνων με τεχνικές βαθιάς μάθησης. Στο Κεφάλαιο 2 γίνεται αναφορά σε βασικές έννοιες, απαραίτητες για την κατανόηση αυτής της εργασίας, όπως η μεταφορά μάθησης και η επεξηγήσιμη τεχνητή νοημοσύνη. Στο Κεφάλαιο 3 περιγράφεται ο σχεδιασμός και η υλοποίηση του πειραματικού μέρους της εργασίας, παρουσιάζοντας αναλυτικά την δομή του συστήματος, τα δεδομένα που χρησιμοποιήθηκαν, τα μοντέλα που υλοποιήθηκαν και τις μετρικές με τις οποίες έγινε η πειραματική αξιολόγηση. Στο Κεφάλαιο 4 γίνεται παρουσίαση και συγκριτική αξιολόγηση των αποτελεσμάτων που προέκυψαν από το σύστημα αναγνώρισης της νόσου του Alzheimer. Στο Κεφάλαιο 5 παρουσιάζονται σε μορφή χαρτών θερμότητας τα αποτελέσματα που προέκυψαν εφαρμόζοντας τεχνικές επεξηγήσιμης τεχνητής νοημοσύνης μαζί με τα συμπεράσματα που προκύπτουν από αυτή την ανάλυση. Τέλος στο Κεφάλαιο 6 πραγματοποιείται μια σύνοψη της δουλειάς της διπλωματικής μαζί με τα τελικά συμπεράσματα που προέκυψαν, αλλά και προτάσεις για πιθανές μελλοντικές προεκτάσεις.

2. Θεωρητικό Υπόβαθρο

Οι μέθοδοι βαθιάς μάθησης είναι ιδιαίτερα αποτελεσματικές όταν ο αριθμός των εκπαιδευτικών παραδειγμάτων είναι μεγάλος. Για παράδειγμα, στο ImageNet Large Scale Visual Recognition Challenge (ILSVRC), ήταν διαθέσιμες περισσότερες από ένα εκατομμύριο εικόνες με τις αντίστοιχες ετικέτες τους.³¹ Ωστόσο, στις περισσότερες ιατρικές εφαρμογές υπάρχουν πολύ λιγότερες εικόνες (δηλαδή <1.000). Συνεπώς, μια από τις σημαντικότερες προκλήσεις στην εφαρμογή μεθόδων βαθιάς μάθησης σε ιατρικές εικόνες, είναι η αποφυγή της υπερεκπαίδευσης (overfitting). Υπερεκπαίδευση έχουμε όταν το μοντέλο μας γίνεται πολύ καλό στα δεδομένα εκπαίδευσης, αλλά δεν μπορεί να κάνει γενίκευση (generalization). Δηλαδή ενώ η επίδοση του μοντέλου είναι άριστη στα δεδομένα εκπαίδευσης, στα δεδομένα ελέγχου (τα οποία δεν έχει δει προηγουμένως), είναι ανεπαρκής.³²

Για να ξεπεραστεί αυτή η πρόκληση, οι ερευνητικές ομάδες έχουν επινοήσει διάφορες στρατηγικές όπως (α) τη λήψη είτε δισδιάστατων (2D) είτε τρισδιάστατων (3D) τμημάτων της εικόνας, αντί για εικόνες πλήρους μεγέθους, ως είσοδο, προκειμένου να μειωθεί η διάσταση εισόδου και συνεπώς ο αριθμός των παραμέτρων του μοντέλου. (β) επέκταση του συνόλου δεδομένων δημιουργώντας τεχνητά δείγματα μέσω γραμμικών (affine) μετασχηματισμών, δηλαδή προσαύξηση των δεδομένων (data augmentation), και στη συνέχεια εκπαίδευση του δικτύου τους από την αρχή με το επαυξημένο σύνολο δεδομένων. (γ) χρήση μοντέλων βαθιάς μάθησης που έχουν εκπαιδευτεί σε ένα τεράστιο αριθμό φυσικών εικόνων στην όραση υπολογιστών, για την εξαγωγή των χαρακτηριστικών χωρίς περαιτέρω εκπαίδευση, και στη συνέχεια εκπαίδευση του τελικού ταξινομητή ή του στρώματος εξόδου με τα δεδομένα του προβλήματος. (δ) αρχικοποίηση των παραμέτρων του μοντέλου με εκείνες των προεκπαιδευμένων μοντέλων από μη ιατρικές ή φυσικές εικόνες και, στη συνέχεια, βελτιστοποίηση των παραμέτρων του δικτύου στα δεδομένα του προβλήματος.¹⁷

Στην εργασία αυτή χρησιμοποιήθηκαν όλες οι παραπάνω τεχνικές με σκοπό την δημιουργία ενός μοντέλου το οποίο θα παράγει καλά και αξιόπιστα αποτελέσματα και θα είναι ικανό να γενικεύει, ταξινομώντας σωστά δεδομένα τα οποία δεν έχει δει προηγουμένως.

2.1 Μεταφορά Μάθησης (Transfer Learning)

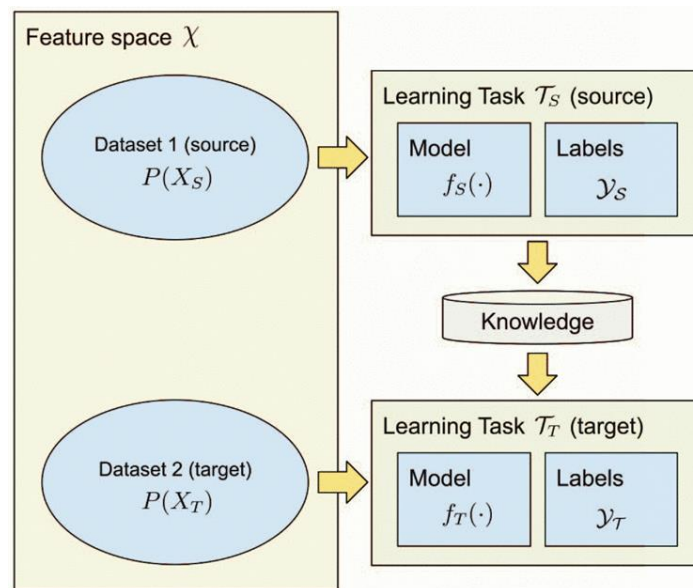
Η μελέτη της Μεταφοράς μάθησης έχει ως βασική ιδέα το γεγονός ότι οι άνθρωποι μπορούμε και εφαρμόζουμε έξυπνα τη γνώση που έχουμε μάθει προηγουμένως, για να επιλύουμε καινούρια προβλήματα γρηγορότερα ή με καλύτερες λύσεις. Για παράδειγμα, μπορούμε να πούμε ότι η εκμάθηση της αναγνώρισης των μήλων μπορεί να βοηθήσει στην αναγνώριση των αχλαδιών. Ομοίως, η εκμάθηση ενός ηλεκτρονικού οργάνου μπορεί να διευκολύνει την εκμάθηση του πιάνου. Το θεμελιώδες κίνητρο για τη μεταφορά μάθησης στον τομέα της μηχανικής μάθησης συζητήθηκε στο συνέδριο NIPS το 1995 στην ημερίδα «Learning to Learn»³³ («Μαθαίνοντας να μαθαίνεις»), το οποίο επικεντρώθηκε στην ανάγκη για δια βίου μεθόδους μηχανικής μάθησης που διατηρούν και επαναχρησιμοποιούν γνώσεις που έχουν μάθει προηγουμένως. Η έρευνα για τη μεταφορά μάθησης ξεκίνησε να προσελκύει ολοένα και μεγαλύτερο ερευνητικό ενδιαφέρον από το 1995 μέχρι σήμερα.³⁴

Αντλώντας έμπνευση από τις ικανότητες των ανθρώπων να μεταφέρουν γνώση σε διάφορους τομείς, η μεταφορά μάθησης στοχεύει στην αξιοποίηση της γνώσης από έναν σχετικό τομέα (που ονομάζεται τομέας πηγής) για τη βελτίωση της μαθησιακής απόδοσης ή την ελαχιστοποίηση του αριθμού των εκπαιδευτικών παραδειγμάτων που απαιτούνται στον τομέα στόχο. Ένας τομέας (domain) D σχετίζεται περισσότερο με την κατανομή των εκπαιδευτικών δεδομένων. Ορίζεται από ένα χώρο χαρακτηριστικών χ και μια οριακή κατανομή πιθανότητας $P(X)$, όπου $X = \{x_1, \dots, x_n\} \in \chi$. Ένα ζητούμενο (task) T ορίζεται από το σύνολο όλων των πιθανών ετικετών Y και ένα μοντέλο ή προγνωστική συνάρτηση $f(\cdot)$ η οποία προβλέπει την αντίστοιχη ετικέτα με βάση ένα δεδομένο τομέα

D. Το ζητούμενο συμβολίζεται ως $T = \{\mathcal{X}, f(\cdot)\}$ και μπορεί να γίνει γνωστό μέσω της εκπαιδευτικής διαδικασίας, χρησιμοποιώντας τα δεδομένα εκπαίδευσης, τα οποία αποτελούνται από μια συλλογή ζευγών $\{x_i, y_i\}$, όπου $x_i \in X$ και $y_i \in Y$. Η συνάρτηση $f(\cdot)$ χρησιμοποιείται για την πρόβλεψη μιας ετικέτας y ή $f(x)$ από ένα παράδειγμα x . Έχοντας ένα τομέα πηγής D_S και ένα ζητούμενο μάθησης T_S , ένα τομέα στόχου D_T και ένα ζητούμενο μάθησης T_T , η μετάδοση μάθησης ορίζεται ως μια διαδικασία η οποία θα βοηθήσει στη βελτίωση της συνάρτησης $f_T(\cdot)$ στο ζητούμενο μάθησης του στόχου με βάση τη γνώση που αποκτήθηκε από το D_S και το T_S , όπου $D_S \neq D_T$, ή $T_S \neq T_T$.³⁵

Σε γενικές γραμμές, σύμφωνα με την ασυμφωνία μεταξύ των τομέων, η μεταφοράς μάθησης μπορεί περαιτέρω να χωριστεί σε δύο κατηγορίες: ομοιογενή και ετερογενή μεταφορά μάθησης. Οι μέθοδοι ομοιογενούς μεταφοράς μάθησης αναπτύσσονται και προτείνονται για την διαχείριση περιπτώσεων όπου οι τομείς έχουν τον ίδιο χώρο χαρακτηριστικών. Η ετερογενής μάθηση μεταφοράς αναφέρεται στη διαδικασία μεταφοράς γνώσης σε καταστάσεις όπου οι τομείς έχουν διαφορετικούς χώρους χαρακτηριστικών.³⁴

Στην **Εικόνα 2.1**, γίνεται αναπαράσταση της μεταφοράς μάθησης με βάση τον ορισμό του τομέα και του ζητούμενου. Σε αυτήν την περίπτωση, τόσο ο τομέας πηγής όσο και ο τομέας προορισμού μοιράζονται τον ίδιο χώρο χαρακτηριστικών, έχουμε δηλαδή ομοιογενή μάθηση, ενώ οι οριακές κατανομές πιθανοτήτων είναι διαφορετικές. Για παράδειγμα, κάθε τομέας έχει το δικό του σύνολο δεδομένων που αποτελείται από διαφορετικές, αλλά σχετικές, εικόνες που αφορούν δυο διαφορετικές εργασίες. Κάθε ζητούμενο έχει το δικό του μοντέλο και ετικέτες.



Εικόνα 2.1: Μεταφορά μάθησης βάση του ορισμού τομέα και ζητούμενου³⁵

Με την πρόοδο στον τομέα της βαθιάς μάθησης, η έρευνα που σχετίζεται με τη μεταφορά μάθησης έχει επικεντρωθεί σε αυτό το είδος αρχιτεκτονικής νευρωνικών δικτύων. Χρησιμοποιώντας τη μεταφορά μάθησης είναι δυνατή η προσαρμογή ενός μοντέλου που έχει εκπαιδευτεί σε ένα διαφορετικό τομέα για διαφορετικό ζητούμενο πηγής στο επιθυμητό ζητούμενο του στόχου.

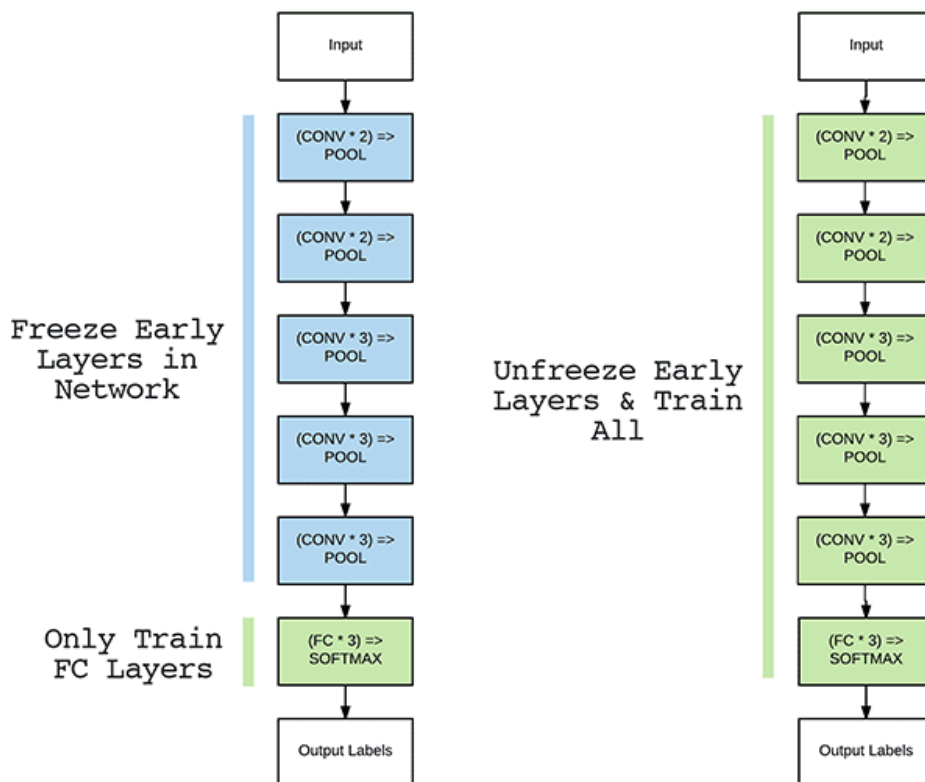
Τα συστήματα βαθιάς μάθησης κατασκευάζονται συνήθως με μια ακολουθία συνελκτικών στρωμάτων και στρωμάτων μέγιστης συγκέντρωσης (max-pooling) για να μάθουν ιεραρχικές αναπαραστάσεις των δεδομένων. Αυτά τα στρώματα λειτουργούν ως εξαγωγείς χαρακτηριστικών που συνδέονται με τα τελικά στρώματα, υπεύθυνα για την εκτέλεση μιας συγκεκριμένης εργασίας, όπως ταξινόμηση ή παλινδρόμηση. Συνήθως, τα τελικά στρώματα περιέχουν πλήρως συνδεδεμένους νευρώνες. Αυτή η αρχιτεκτονική μας επιτρέπει να χρησιμοποιήσουμε ένα προ-εκπαιδευμένο μοντέλο για να εκτελέσουμε μεταφορά μάθησης, αντικαθιστώντας τα τελευταία στρώματα και στη συνέχεια προσαρμόζοντας (fine-

tuning) το μοντέλο για μια εργασία-στόχο. Επίσης είναι δυνατό το «πάγωμα» (freezing) των πρώτων στρωμάτων, χρησιμοποιώντας ένα μέρος του μοντέλου της πηγής ως εξαγωγέα χαρακτηριστικών για τα τελευταία στρώματα στην εργασία στόχο. Τέλος, μπορεί να γίνει συνδυασμός και των δύο προσεγγίσεων επιλέγοντας ποια στρώματα θα «παγώσουν» ή θα επανεκπαιδευτούν, κάτι που μπορεί να είναι χρήσιμο ανάλογα με τους τομείς προέλευσης και στόχου.³⁵

Η πιο συνηθισμένη ροή εργασίας για τη μεταφορά γνώσης σε ένα μοντέλο βαθιάς μάθησης είναι η ακόλουθη³⁶:

1. Εισαγωγή των στρωμάτων ενός προεκπαιδευμένου μοντέλου.
2. Πάγωμα (freezing) αυτών των στρωμάτων, ούτως ώστε να μην καταστραφεί η πληροφορία η οποία περιέχουν στους επόμενους εκπαιδευτικούς κύκλους.
3. Προσθήκη νέων εκπαιδύσιμων (trainable) στρωμάτων, πάνω από τα παγωμένα στρώματα. Τα καινούρια στρώματα θα μάθουν να μετατρέπουν τα παλιά χαρακτηριστικά σε προβλέψεις στο νέο σύνολο δεδομένων.
4. Εκπαίδευση των καινούριων στρωμάτων στο σύνολο δεδομένων του στόχου.

Ένα τελευταίο, προαιρετικό βήμα, είναι η προσαρμογή (fine-tuning), η οποία αποτελείται από το ξεπάγωμα ολόκληρου του μοντέλου (ή μέρος του) και την εκ νέου εκπαίδευσή του στα νέα δεδομένα με πολύ χαμηλό ρυθμό εκμάθησης (**Εικόνα 2.2**). Αυτό μπορεί να επιτύχει δυναμικά σημαντικές βελτιώσεις, προσαρμόζοντας σταδιακά τα προεκπαιδευμένα χαρακτηριστικά στα νέα δεδομένα.



Εικόνα 2.2: Διαδικασία της προσαρμογής (fine-tuning)³⁷

Υπάρχουν πολλά δίκτυα βαθιάς μάθησης που έχουν πετύχει κορυφαίες επιδόσεις σε διάφορες εργασίες, όπως το VGG³⁸, το Inception^{39,40}, το ResNet⁴¹, το Xception⁴², το EfficientNet^{43,44} και πολλά άλλα. Τα περισσότερα από αυτά τα μοντέλα είναι διαθέσιμα για το κοινό. Σε πολλές περιπτώσεις, οι συγγραφείς μοιράζονται λεπτομέρειες σχετικά με την υλοποίηση και τον πηγαίο κώδικα ή ακόμη και δίνουν την δυνατότητα για λήψη των προεκπαιδευμένων μοντέλων στα πιο χρησιμοποιούμενα frameworks βαθιάς μάθησης. Το βασικό κίνητρο για τη διαθεσιμότητα αυτών των μοντέλων είναι ότι έχουν ήδη επιτύχει σταθερή απόδοση σε συγκεκριμένες εργασίες και μπορούν να χρησιμοποιηθούν με ασφάλεια για τη μεταφορά μάθησης σε σχετικά προβλήματα.³⁵

2.2 Επεξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI)

Η βαθιά μάθηση έχει συμβάλει σημαντικά στην πρόσφατη πρόοδο της τεχνητής νοημοσύνης. Σε σύγκριση με τις παραδοσιακές μεθόδους μηχανικής μάθησης, όπως τα δέντρα αποφάσεων και οι μηχανές διανυσμάτων υποστήριξης (support vector machines), οι μέθοδοι βαθιάς μάθησης έχουν επιτύχει ουσιαστική βελτίωση σε πληθώρα προβλημάτων πρόβλεψης. Ωστόσο, τα βαθιά νευρωνικά δίκτυα παρουσιάζουν αδυναμία στην επεξήγηση της διαδικασίας με την οποία λαμβάνουν αποφάσεις και συνήθως αντιμετωπίζονται ως «μαύρα κουτιά» τόσο από τους προγραμματιστές, όσο και από τους χρήστες.⁴⁵ Ένα σύστημα τεχνητής νοημοσύνης είναι επεξηγήσιμο, όταν παράγει πληροφορίες οι οποίες μπορούν να αιτιολογήσουν την λειτουργία του, κάνοντας την ξεκάθαρη και εύκολη στην κατανόηση.⁴⁶

Οι σημαντικότεροι λόγοι για την δημιουργία επεξηγήσιμων συστημάτων είναι οι ακόλουθοι⁴⁷:

- **Επαλήθευση του συστήματος:** Υπάρχουν αρκετές εφαρμογές στις οποίες δεν μπορεί να υπάρξει τυφλή εμπιστοσύνη σε ένα σύστημα τεχνητής νοημοσύνης που λειτουργεί σαν «μαύρο κουτί». Για παράδειγμα, στον ιατρικό τομέα η χρήση μοντέλων που μπορούν να ερμηνευθούν και να επαληθευτούν από ειδικούς ιατρούς είναι απόλυτη αναγκαιότητα. Επίσης η επεξηγησιμότητα αποτελεί βασική προϋπόθεση, ώστε οι ασθενείς να αποδεχτούν και να εμπιστευτούν συστήματα τεχνητής νοημοσύνης που υποβοηθούν στη λήψη ιατρικών αποφάσεων. Χωρίς την δυνατότητα της επεξήγησης της συλλογιστικής διαδικασίας που οδήγησε στη λήψη μιας απόφασης, θα είναι πολύ δύσκολο να εφαρμοστούν τέτοια συστήματα στην πράξη.
- **Βελτίωση του συστήματος:** Το πρώτο βήμα προς την βελτίωση ενός συστήματος τεχνητής νοημοσύνης είναι η κατανόηση των αδυναμιών του. Προφανώς, είναι πιο δύσκολο να πραγματοποιηθεί μια ανάλυση αδυναμίας σε ένα μοντέλο «μαύρου κουτιού» σε σχέση με ένα μοντέλο που είναι ερμηνεύσιμο. Επίσης, η ανίχνευση των προκαταλήψεων ενός μοντέλου ή ενός συνόλου δεδομένων είναι ευκολότερη, εάν υπάρχει επεξήγηση της λειτουργίας του μοντέλου και γιατί καταλήγει στις προβλέψεις του. Επιπλέον, η επεξηγησιμότητα του μοντέλου μπορεί να είναι χρήσιμη όταν συγκρίνονται διαφορετικά μοντέλα ή αρχιτεκτονικές.
- **Μάθηση από τη γνώση του συστήματος:** Τα σύγχρονα συστήματα τεχνητής νοημοσύνης εκπαιδεύονται με εκατομμύρια παραδείγματα, παρατηρώντας μοτίβα στα δεδομένα τα οποία δεν είναι διαθέσιμα στους ανθρώπους, οι οποίοι μπορούν να μαθαίνουν μόνο με ένα περιορισμένο αριθμό παραδειγμάτων. Χρησιμοποιώντας επεξηγήσιμα συστήματα τεχνητής νοημοσύνης, η γνώση που έχει αποκτήσει το σύστημα τεχνητής νοημοσύνης μπορεί να εξαχθεί με σκοπό την απόκτηση νέας γνώσης. Ένα παράδειγμα μεταφοράς γνώσης από το σύστημα τεχνητής νοημοσύνης στον άνθρωπο αφορά το AlphaGo, ένα σύστημα τεχνητής νοημοσύνης κατασκευασμένο από την DeepMind το οποίο είχε εκπαιδευτεί ώστε να παίζει αυτόνομα το παιχνίδι Go. Το AlphaGo δημιούργησε νέες στρατηγικές στο παιχνίδι, οι οποίες έχουν

ενσωματωθεί πλέον από τους επαγγελματίες παίκτες. Ένας άλλος τομέας στον οποίο η εξαγωγή πληροφορίας από το μοντέλα θα ήταν κρίσιμη είναι οι επιστήμες. Με απλά λόγια, οι φυσικοί, οι χημικοί και οι βιολόγοι ενδιαφέρονται περισσότερο να αναγνωρίσουν κρυμμένους νόμους της φύσης, παρά απλώς να προβλέψουν κάποια ποσότητα με μοντέλα «μαύρου κουτιού».

- **Συμμόρφωση με τη νομοθεσία:** Τα συστήματα τεχνητής νοημοσύνης επηρεάζουν ολοένα και περισσότερους τομείς της καθημερινής μας ζωής. Αυτό έχει ως συνέπεια, την αυξημένη προσοχή στις νομικές προεκτάσεις που έχει η εφαρμογή τέτοιων συστημάτων, π.χ. η ανάθεση ευθύνης όταν τα συστήματα λαμβάνουν μια λανθασμένη απόφαση. Δεδομένου ότι μπορεί να είναι αδύνατο να βρεθούν ικανοποιητικές απαντήσεις για αυτά τα νομικά ερωτήματα όταν βασιζόμαστε σε μοντέλα «μαύρου κουτιού», τα μελλοντικά συστήματα τεχνητής νοημοσύνης θα πρέπει αναγκαστικά να γίνουν πιο επεξηγήσιμα. Ένα άλλο παράδειγμα όπου οι κανονισμοί μπορεί να γίνουν κινητήρια δύναμη για περισσότερη επεξηγησιμότητα στην τεχνητή νοημοσύνη είναι τα προσωπικά δικαιώματα. Οι άμεσα επηρεαζόμενοι από τις αποφάσεις ενός συστήματος τεχνητής νοημοσύνης (π.χ. κάποιος που απορρίφθηκε η αίτηση του για δάνειο από την τράπεζα) μπορεί να ζητήσουν να μάθουν γιατί το σύστημα αποφάσισε με αυτόν τον τρόπο. Αυτές οι πληροφορίες μπορούν να παρέχονται μόνο από ερμηνεύσιμα συστήματα τεχνητής νοημοσύνης. Αυτές οι ανησυχίες οδήγησαν την Ευρωπαϊκή Ένωση να δημιουργήσει νέους κανονισμούς, οι οποίοι περιλαμβάνουν το «δικαίωμα στην επεξήγηση», σύμφωνα με το οποίο ο χρήστης έχει το δικαίωμα να ζητήσει την επεξήγηση μιας αλγοριθμικής απόφασης που έγινε σε βάρος του.

Αυτά τα παραδείγματα υποδεικνύουν ότι η επεξηγησιμότητα δεν αποτελεί μόνο σημαντικό και επίκαιρο ακαδημαϊκό ενδιαφέρον, αλλά θα έχει καθοριστικό ρόλο στα μελλοντικά συστήματα τεχνητής νοημοσύνης. Ειδικότερα στον τομέα της ιατρικής εμπεριέχονται οι έννοιες του ρίσκου και της ευθύνης, οι οποίες καθιστούν το ζήτημα της επεξηγησιμότητας ιδιαίτερα κρίσιμο. Όταν λαμβάνονται ιατρικές αποφάσεις συχνά διακυβεύονται ανθρώπινες ζωές, συνεπώς δεν είναι δυνατή η ανάθεση αυτής της ευθύνης σε ένα σύστημα το οποίο δεν μπορεί να αιτιολογήσει τις αποφάσεις που λαμβάνει. Εκτός από ζητήματα ηθικής, αυτό είναι ένα σοβαρό κενό ασφαλείας που θα μπορούσε να γίνει καταστροφικό σε περίπτωση που γίνει εκμετάλλευση του με κακόβουλη πρόθεση.⁴⁸

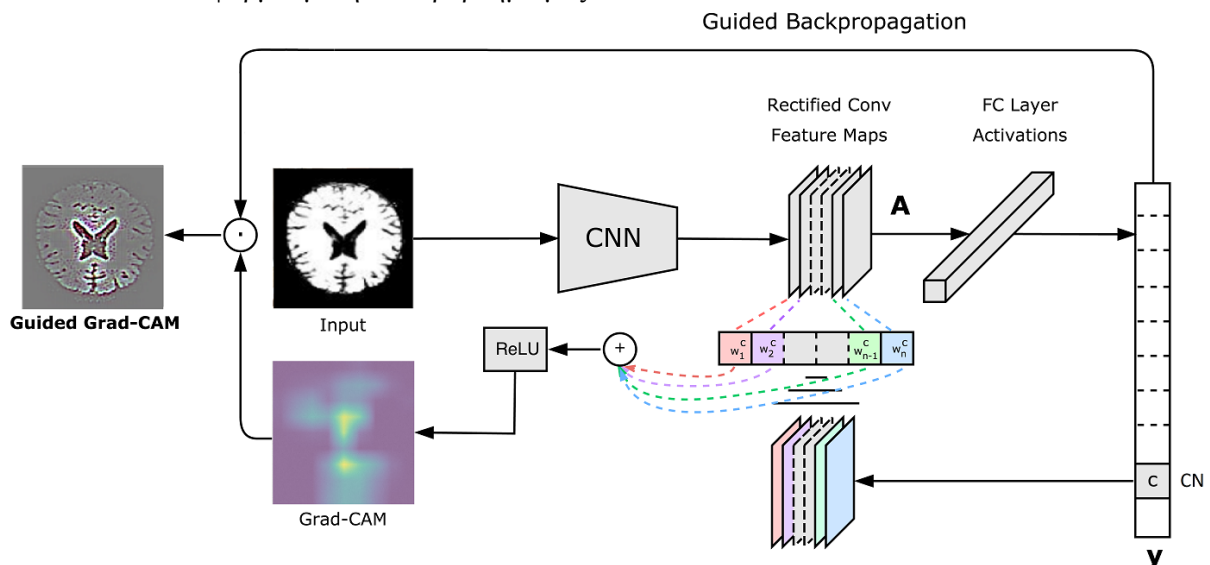
Στα μοντέλα όρασης υπολογιστών ο Διαχωριστικός Εντοπισμός (Discriminative Localization) είναι ο βασικός τρόπος με τον οποίο επιτυγχάνεται η ερμηνευσιμότητα. Με τον όρο αυτό περιγράφεται η υπόδειξη των περιοχών μιας εικόνας που οδήγησαν στο νευρωνικό δίκτυο στην πρόβλεψη μιας κλάσης έναντι κάποιας άλλης. Στα Συνελικτικά Νευρωνικά Δίκτυα το πιο χαρακτηριστικό παράδειγμα ερμηνευσιμότητας είναι οι Χάρτες Ενεργοποίησης Κλάσης (Class Activation Maps, CAM).⁴⁹ Οι χάρτες αυτοί απεικονίζουν τα σημεία της εικόνας τα οποία εστιάζει το νευρωνικό δίκτυο όταν πραγματοποιεί προβλέψεις. Στην **Εικόνα 2.3** φαίνονται οι περιοχές εστίασης που προκύπτουν από τους χάρτες ενεργοποίησης κλάσης. Αυτή η τεχνική επιτρέπει στο εκπαιδευμένο για ταξινόμηση CNN να ταξινομεί την εικόνα και ταυτόχρονα να εντοπίζει περιοχές της εικόνας σχετικές με την κλάση που αναγνωρίζει σε ένα πέρασμα προς τα εμπρός, π.χ. η οδοντόβουρτσα στο βούρτσισμα των δοντιών και το αλυσοπρίο στο κόψιμο των δέντρων.



Εικόνα 2.3: Χάρτες Ενεργοποίησης Κλάσης⁴⁹

Στην παρούσα εργασία θα χρησιμοποιηθεί το Grad-CAM⁵⁰, το οποίο αποτελεί μια αξιοσημείωτη επέκταση των Χαρτών Ενεργοποίησης Κλάσης. Το Grad-CAM χρησιμοποιεί την πληροφορία από τις κλίσεις (gradients) που καταλήγουν στο τελευταίο συνελκτικό επίπεδο ενός CNN, ώστε να εκχωρήσει τιμές σημασίας σε κάθε νευρώνα για μια συγκεκριμένη απόφαση ενδιαφέροντος. Ένα σημαντικό πλεονέκτημα αυτής της τεχνικής, αποτελεί η δυνατότητα να εφαρμοστεί σε ένα ευρύ φάσμα νευρωνικών δικτύων χωρίς την ανάγκη επανεκπαίδευσης ή προσθήκης επιπρόσθετων στρωμάτων. Αυτές οι απεικονίσεις είναι ιδιαίτερα χρήσιμες για την κατανόηση των λόγων που ένα μοντέλο ταξινομεί σωστά ή λανθασμένα.

Επίσης, στην παρούσα εργασία εφαρμόζεται και μια επέκταση του Grad-CAM, το Guided Grad-CAM, το οποίο στοχεύει να τονίσει τις σημαντικές περιοχές με μεγαλύτερη λεπτομέρεια. Αυτό επιτυγχάνεται, χρησιμοποιώντας την οπισθοδιάδοση (backpropagation), η οποία μηδενίζει στοιχεία στις κλίσεις που έχουν αρνητική επίδραση στην απόφαση. Αυτή η τεχνική είχε εφαρμοστεί αρχικά για την ερμηνεία των αποτελεσμάτων ενός δικτύου VGG16, ενισχύοντας τόσο την πιστότητα (faithfulness) όσο και την ερμηνευσιμότητα του μοντέλου. Σε αυτό το πλαίσιο, η πιστότητα είναι η ακρίβεια που απεικονίζεται η λειτουργία του μοντέλου και η ερμηνευσιμότητα έχει να κάνει με τον βαθμό στον οποίο τα αποτελέσματα μπορούν να γίνουν κατανοητά από τους ανθρώπους. Ενώ συνήθως παρατηρείται αντίθεση μεταξύ αυτών των δύο, αυτή η τεχνική σημείωσε υψηλή βαθμολογία και στις δύο μετρήσεις σε ένα μείγμα τεχνικών και ανθρώπινων πειραμάτων. Στην **Εικόνα 2.4** απεικονίζεται η αρχιτεκτονική του Grad-CAM εφαρμοσμένη στο πρόβλημα μας.



Εικόνα 2.4: Αρχιτεκτονική του Grad-CAM

3. Σχεδιασμός και Υλοποίηση

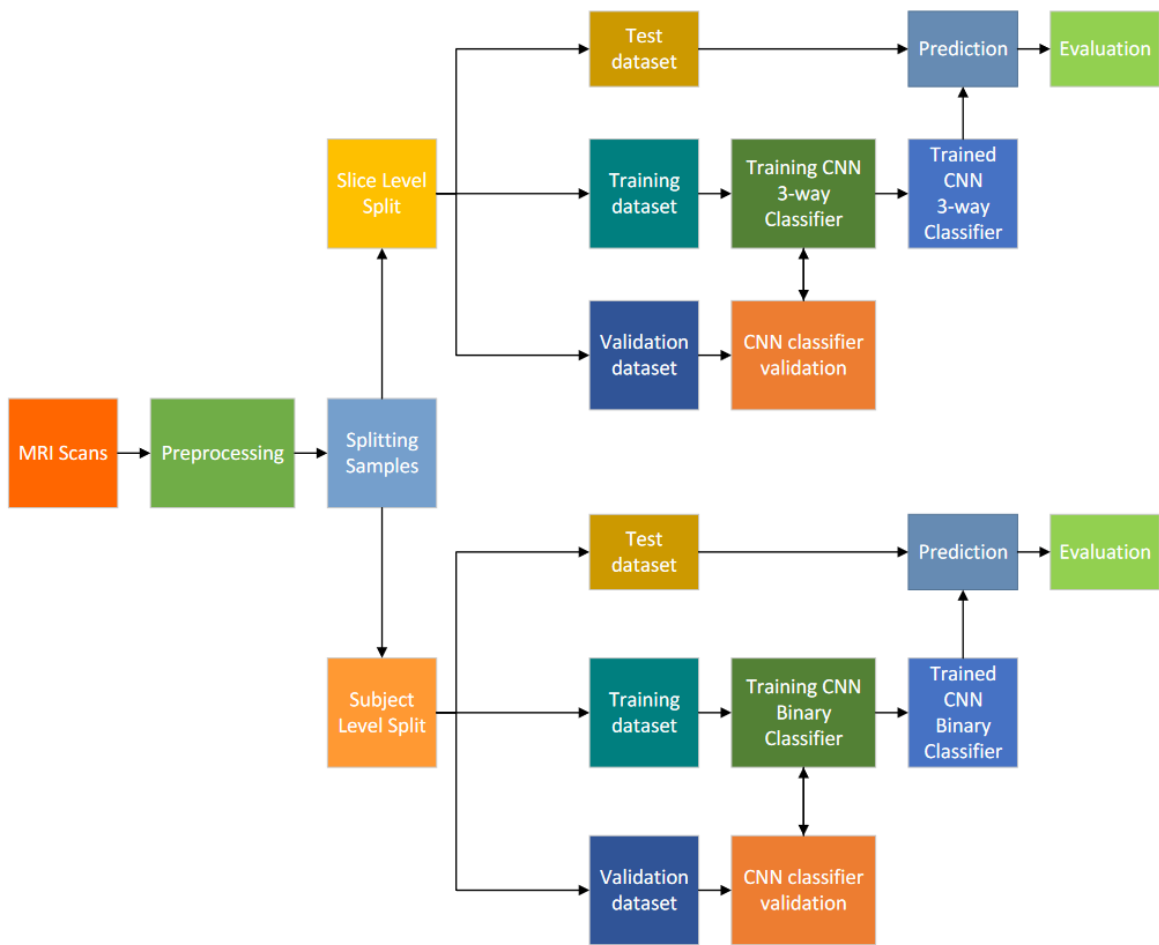
3.1 Δομή του συστήματος

3.1.1 Δομή του συστήματος διάγνωσης του Alzheimer

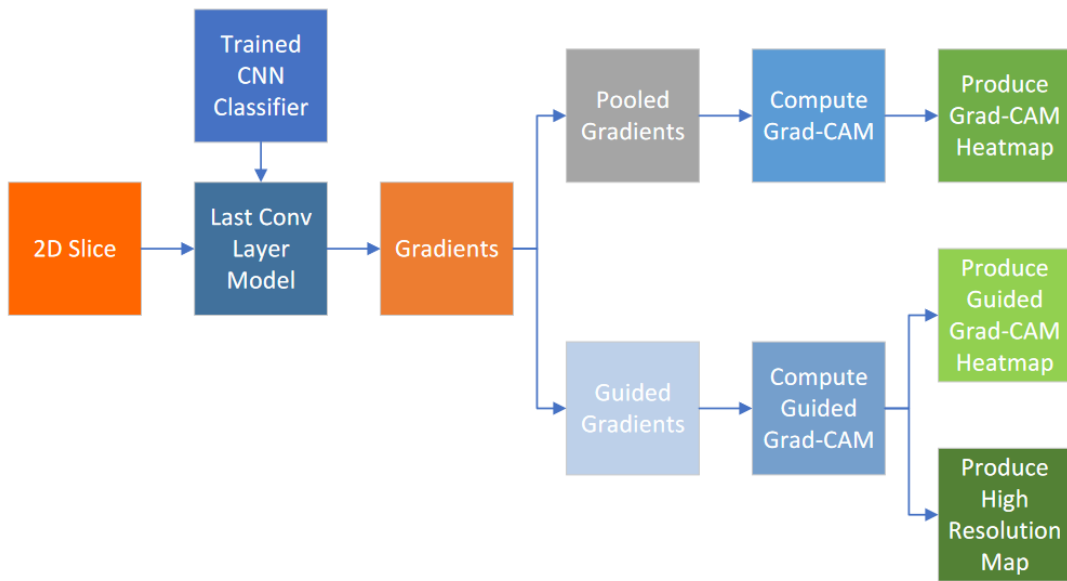
Η αναγνώριση της νόσου του Alzheimer είναι μια αρκετά πολύπλοκη διαδικασία καθώς καλούμαστε να παράγουμε διανύσματα πιθανοτήτων από raw εικόνες μαγνητική τομογραφίας εγκεφάλου. Οι εικόνες που προκύπτουν από τον μαγνητικό τομογράφο δεν είναι σε μορφή κατάλληλη για να τροφοδοτηθούν σε ένα νευρωνικό δίκτυο. Οι εικόνες MRI έχουν τρεις διαστάσεις, μια για κάθε επίπεδο στο χώρο, και μπορεί να παρουσιάζουν διαφορετική ανάλυση και επεξεργασία, ανάλογα από τον μαγνητικό τομογράφο από τον οποίο έχουν προέλθει. Συνεπώς κρίνεται απαραίτητος ο μετασχηματισμός τους σε ένα κοινό σύστημα συντεταγμένων σε συνδυασμό με ειδική επεξεργασία με σκοπό την απαλοιφή περιττών πληροφοριών για το πρόβλημα που μας ενδιαφέρει. (π.χ. η πληροφορία του κρανίου). Τέλος οι εικόνες θα πρέπει να μετασχηματιστούν στον δισδιάστατο χώρο, ούτως ώστε να μπορούν να αποτελέσουν είσοδο στις περισσότερες καθιερωμένες αρχιτεκτονικές δικτύων βαθιάς μάθησης. Όλες αυτές οι διαδικασίες που αναφέραμε πραγματοποιούνται στο βήμα της προεπεξεργασίας των δεδομένων. Έπειτα πριν την εκπαίδευση των δικτύων θα γίνει διαχωρισμός των δεδομένων με δύο τρόπους (α) σε επίπεδο εγκάρσιας τομής (slice-level) και (β) σε επίπεδο υποκειμένου (subject-level). Για κάθε μια από τις παραπάνω περιπτώσεις, τα δεδομένα θα διαχωριστούν περαιτέρω στα σύνολα εκπαίδευσης (training), επικύρωσης (validation) και ελέγχου (test). Έτσι μπορεί να ξεκινήσει η διαδικασία εκπαίδευσης των δικτύων βαθιάς μάθησης τα οποία εκπαιδεύονται στο training set, ελέγχουν την επίδοσή τους στο validation set και αξιολογούνται στο test set. Το pipeline της εφαρμογής ταξινόμησης των ιατρικών εικόνων στις αντίστοιχες κατηγορίες περιγράφεται γραφικά στην **Εικόνα 3.1**.

3.1.2 Δομή του συστήματος επεξήγησης των αποτελεσμάτων

Για την επεξήγηση των αποτελεσμάτων αρχικά φορτώνονται τα δίκτυα που εκπαιδεύτηκαν από το προηγούμενο βήμα και τους αφαιρείται το τελευταίο κομμάτι, ούτως ώστε η έξοδος τους να είναι η έξοδος του τελευταίου συνελκτικού επιπέδου. Η έξοδος αυτού του δικτύου είναι τα gradients στα οποία εφαρμόζεται μια διαδικασία pooling, ούτως ώστε να υπολογιστεί το Grad-CAM και το Guided Grad-CAM αντιστοίχως. Έτσι προκύπτουν αναπαραστάσεις με την μορφή χαρτών θερμότητας (heatmaps) οι οποίοι αναπαριστούν τις μέγιστες ενεργοποιήσεις του κάθε νευρωνικού κατά τη λήψη μιας απόφασης. Για την περίπτωση του Guided Grad-CAM γίνεται παραγωγή και ενός χάρτη υψηλής ανάλυσης ο οποίος αποτυπώνει τις λεπτομέρειες που οδήγησαν στο συγκεκριμένο αποτέλεσμα πρόβλεψης. Το pipeline του συστήματος επεξήγησης των αποτελεσμάτων περιγράφεται γραφικά στην **Εικόνα 3.2**.



Εικόνα 3.1: Δομή του συστήματος ταξινόμησης



Εικόνα 3.2: Δομή του συστήματος επεξήγησης των αποτελεσμάτων

3.2 Δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπόνηση αυτής της διπλωματικής εργασίας προήλθαν από το Alzheimer’s Disease Neuroimaging Initiative (ADNI). Κατά συνέπεια, οι ερευνητές εντός του ADNI συνέβαλαν στο σχεδιασμό και την υλοποίηση του ADNI ή/και παρείχαν δεδομένα, αλλά δεν συμμετείχαν στην ανάλυση ή τη σύνταξη αυτής της εργασίας. Η πλήρης λίστα των ερευνητών του ADNI είναι διαθέσιμη σε [αυτή τη διεύθυνση](#).

Το ADNI αποτελεί μια διαχρονική, πολυκεντρική μελέτη που έχει ως σκοπό την ανάπτυξη κλινικών, απεικονιστικών, γενετικών και βιοχημικών δεικτών για την έγκαιρη ανίχνευση και παρακολούθηση της ΝΑ⁵¹. Πρόκειται για μια σύμπραξη δημόσιου και ιδιωτικού τομέα η οποία από την ίδρυση της, το έτος 2004, έχει συνεισφέρει σημαντικά στην έρευνα για την ΝΑ, καθιστώντας δυνατή την ανταλλαγή δεδομένων μεταξύ ερευνητών σε όλο τον κόσμο. Οι τρεις κυριότεροι στόχοι του ADNI είναι:

1. Η ανίχνευση της ΝΑ σε όσο δυνατόν πιο πρώιμο στάδιο γίνεται (πριν την άνοια) και η αναγνώριση μεθόδων για την παρακολούθηση της νόσου μέσω βιοδεικτών.
2. Η υποστήριξη της προόδου στην παρέμβαση, πρόληψη και θεραπεία της ΝΑ μέσω εφαρμογής νέων διαγνωστικών μεθόδων στα κατά το δυνατόν πιο πρώιμα στάδια (όταν δηλαδή η παρέμβαση μπορεί να είναι και πιο αποτελεσματική).
3. Να συνεχίσει να διαχειρίζεται την καινοτόμο πολιτική πρόσβασης στα δεδομένα του ADNI, χωρίς εμπάργκο, για όλους τους επιστήμονες σε όλο τον κόσμο.

Από τα δεδομένα του ADNI λήφθηκαν πάνω από 2000 εικόνες μαγνητικής τομογραφίας (MRI) εγκεφάλου σε T1-weighted μορφή. Οι ασθενείς έχουν εύρος ηλικίας από 55 έως 92 ετών και είναι χωρισμένοι σε τρεις κατηγορίες: AD (ασθενείς με ΝΑ), MCI (ασθενείς με ήπια γνωστική εξασθένηση) και CN (υγιείς). Στον Πίνακα 3.1 παρακάτω φαίνεται η πλήρης κατανομή των δεδομένων.

Πίνακας 3.1: Κατανομή των ιατρικών εικόνων σε κλάσεις

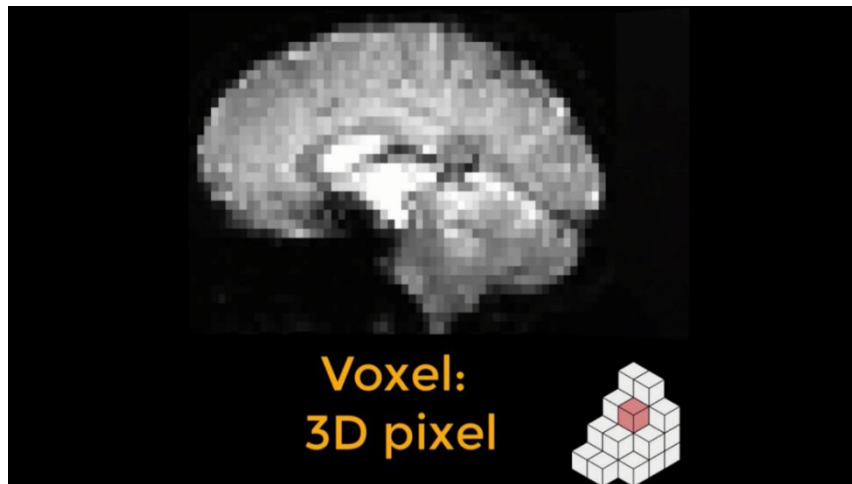
Κλάση	Πλήθος
AD	453
MCI	981
CN	746
Σύνολο	2180

Οι ασθενείς για τους οποίους υπάρχουν μαγνητικές τομογραφίες είναι μεταξύ 55 και 92 ετών. Το μέγεθος του συνόλου δεδομένων είναι αντίστοιχο με αυτό που έχει χρησιμοποιηθεί στις κυριότερες έρευνες που αναφέρθηκαν στην Ενότητα 2. Είναι επίσης σημαντικό να σημειωθεί ότι το σύνολο δεδομένων παρουσιάζει σημαντική ανισορροπία.

Τα δεδομένα των εικόνων χρησιμοποιήθηκαν χωρίς κάποια συμπληρωματική πληροφορία. Από το σύνολο των δεδομένων ένα 10% κρατήθηκε εκτός, για να χρησιμοποιηθεί ως σύνολο ελέγχου (test set). Το test set δεν χρησιμοποιείται στη φάση της εκπαίδευσης και είναι το βέλτιστο σύνολο για να ελέγξουμε το πόσο καλά αποδίδει το σύστημα μας σε νέα δεδομένα. Παράλληλα, από το υπόλοιπο σύνολο εκπαιδευτικών δεδομένων, δεσμεύουμε ακόμη 10% για να το χρησιμοποιήσουμε ως σύνολο επικύρωσης (validation set). Το validation set χρησιμοποιείται για να ελέγχεται η επίδοση το μοντέλου κατά την διαδικασία της εκπαίδευσης, ούτως ώστε να αποφευχθεί η υπερεκπαίδευση (overfitting) του μοντέλου.

3.2.1 Προ-επεξεργασία των Δεδομένων

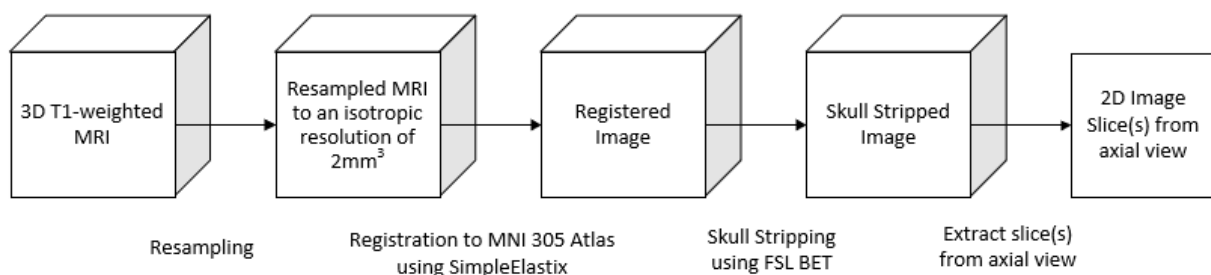
Οι εικόνες μαγνητικής τομογραφίας που λήφθηκαν από την βάση δεδομένων του ADNI παρουσιάζαν μεγάλη ποικιλία ως προς τις διαστάσεις τους. Επίσης ορισμένες είχαν ήδη υποστεί μια προεπεξεργασία με διάφορες τεχνικές. Συνεπώς, κρίθηκε απαραίτητο να γίνει καταχώρηση τους (registration) σε ένα κοινό άτλαντα (atlas), έτσι ώστε όλες να έχουν τις ίδιες διαστάσεις και η ανατομική πληροφορία που αποτυπώνεται σε κάθε voxel να είναι η ίδια. Με τον όρο voxel αναφερόμαστε στα στοιχεία που συνθέτουν μια εικόνα μαγνητικής τομογραφίας. Είναι το αντίστοιχο του pixel για τον τρισδιάστατο χώρο.



Εικόνα 3.3: Αναπαράσταση Voxel⁵²

Τα voxel που συνθέτουν μια εικόνα MRI μπορούν να αναπαρασταθούν σαν κύβους όπως φαίνεται στο κάτω δεξιά μέρος στην **Εικόνα 3.3** παραπάνω. Κάθε ένας από αυτούς τους κύβους περιέχει έναν μόνο αριθμό που αντιπροσωπεύει το σήμα που μετρήθηκε σε αυτό το voxel. Τα voxel μπορεί να είναι είτε ισότροπα (isotropic), με διαστάσεις ίσων μηκών, είτε ανισότροπα (anisotropic), με τουλάχιστον μία διάσταση είτε μεγαλύτερη είτε μικρότερη από τις άλλες διαστάσεις.

Η διαδικασία προεπεξεργασίας των εικόνων MRI που ακολουθήσαμε στην παρούσα εργασία συνοψίζεται στην **Εικόνα 3.4** που φαίνεται παρακάτω και βασίστηκε στην υλοποίηση του O. Plasencia⁵³.

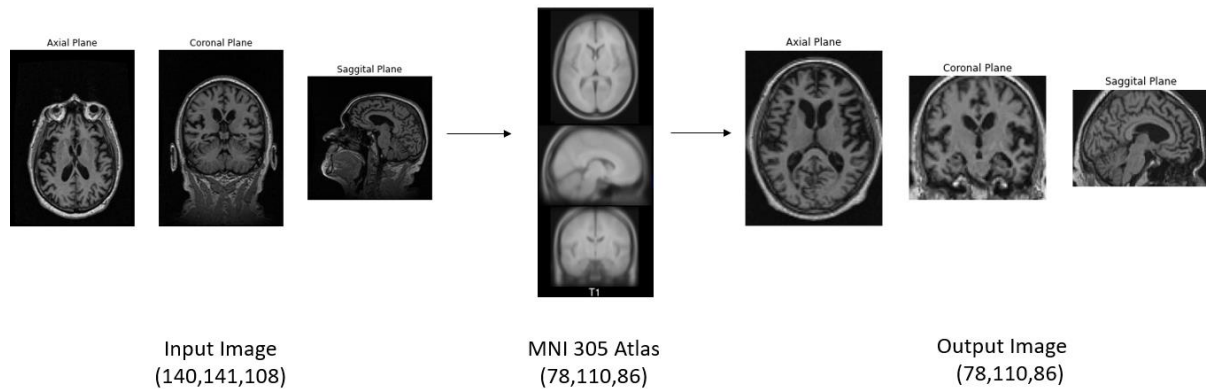


Εικόνα 3.4: Βήματα προεπεξεργασίας των δεδομένων

Το πρώτο βήμα για την προεπεξεργασία των εικόνων MRI, είναι η κανονικοποίηση ως προς τον χώρο (spatial normalization). Είναι συχνό, να παρατηρούνται διαφορές στις διαστάσεις των voxels, ακόμη και όταν οι εικόνες προέρχονται από το ίδιο μαγνητικό τομογράφο. Για να μεριμνήσουμε για αυτές τις περιπτώσεις εφαρμόζουμε επαναδειγματοληψία (resampling) σε μια ιστροπική ανάλυση. Με τον όρο resampling, εννοούμε την τροποποίηση μιας εικόνας όσον αφορά την ανάλυση, τις διαστάσεις ή και τα

δύο. Μια συχνή χρήση του resampling είναι η τροποποίηση των διαστάσεων των voxel, το οποίο με τη σειρά του δημιουργεί μεγαλύτερα ή μικρότερα voxel. Τα μεγαλύτερα voxel έχουν ως αποτέλεσμα μια εικόνα χαμηλότερης ανάλυσης, ενώ τα μικρότερα voxel οδηγούν σε μια εικόνα υψηλότερης ανάλυσης. Στην περίπτωση μας εφαρμόσαμε resampling, ώστε να κανονικοποιήσουμε τις εικόνες σε μια ισοτροπική ανάλυση 2mm^3 για κάθε voxel. Η διαδικασία αυτή είχε ως αποτέλεσμα να μειωθούν οι διαστάσεις όλων των εικόνων στο μισό, γεγονός το οποίο έκανε το πρόβλημα περισσότερο διαχειρίσιμο υπολογιστικά.

Ένα τυπικό βήμα στην ανάλυση εικόνων μαγνητικής τομογραφίας εγκεφάλου (π.χ. T1-weighted MRI) είναι η καταχώριση όλων των εικόνων του συνόλου εκπαίδευσης σε ένα πρότυπο αναφοράς, όπως ένας mean atlas. Ο atlas που χρησιμοποιήθηκε σε αυτή την εργασία είναι ο MNI 305 και έχει προκύψει από το μέσο όρο 305 T1-weighted εικόνων μαγνητικής τομογραφίας, μετασχηματισμένο γραμμικά στον χώρο Talairach⁵⁴⁻⁵⁶. Αντίστοιχα με το σύνολο των εικόνων, εφαρμόστηκε κανονικοποίηση και στον atlas μετασχηματίζοντας τον σε μια ισοτροπική ανάλυση 2mm^3 για κάθε voxel. Κατά συνέπεια η ανάλυση του atlas που προέκυψε είναι $78 \times 110 \times 86$, άρα και όλες οι τελικές εικόνες θα έχουν αυτή την ανάλυση.



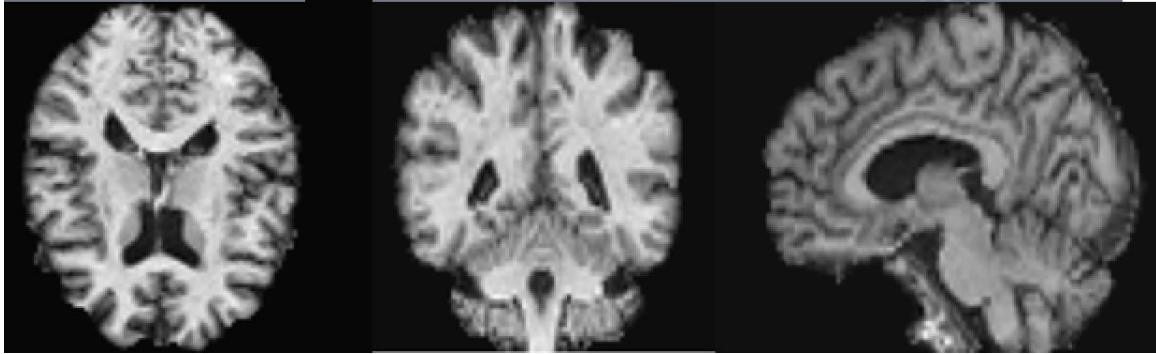
Εικόνα 3.5: Διαδικασία καταχώρισης (register) στον atlas

Η διαδικασία του resampling υλοποιήθηκε μέσω μιας συνάρτησης, η οποία βασίστηκε σε παραδείγματα κώδικα της βιβλιοθήκης DLTK⁵⁷. Από την άλλη, η διαδικασία του registration σε ένα κοινό atlas, υλοποιήθηκε κάνοντας χρήση του SimpleElastix⁵⁸, μια βιβλιοθήκη καταχώρισης ιατρικών εικόνων που βασίζεται στο SimpleITK^{59,60}.

Τέλος, υλοποιήθηκε μια διαδικασία για την εξαγωγή του κρανίου (skull stripping) από τις εικόνες. Αφαιρώντας την πληροφορία του κρανίου από τις εικόνες, ουσιαστικά μένει μόνο η πληροφορία του εγκεφαλικού ιστού, η οποία είναι και η σημαντική στο πρόβλημα μας. Αυτή η διαδικασία υλοποιήθηκε με τη χρήση του FSL BET⁶¹, ενός λογισμικού νευροαπεικόνισης που σχεδιάστηκε για αυτό τον σκοπό. Η βασική του παράμετρος είναι το κλασματικό κατώφλι της έντασης, το οποίο παίρνει τιμές μεταξύ 0 και 1 και έχει το 0.5 ως προκαθορισμένη τιμή. Όσο μικρότερη είναι η τιμή κατωφλίου, τόσο μεγαλύτερο εκτιμάται το περίγραμμα του εγκεφάλου. Ουσιαστικά, πρόκειται για ένα μέτρο επιθετικότητας του αλγορίθμου στην αφαίρεση στοιχείων της εικόνας που δεν αντιστοιχούν σε εγκεφαλικό ιστό. Μια πολύ μικρή τιμή θα άφηνε μεγάλο μέρος μη σχετικής πληροφορία στην εικόνα, ενώ μια πολύ μεγάλη τιμή θα μπορούσε να αφαιρέσει μέρος του εγκεφάλου, χάνοντας έτσι σημαντική πληροφορία. Μετά από δοκιμές, μια τιμή κατωφλίου ίση με 0.2 φάνηκε να διατηρεί την καλύτερη δυνατή ισορροπία στο σύνολο των εικόνων.

Σε αυτό το σημείο, αξίζει να αναφερθεί ότι το FSL BET (Brain Extraction Tool) είναι ένα εργαλείο το οποίο τρέχει από την γραμμή εντολών σε συστήματα UNIX. Για να γίνει ενσωμάτωση του σε κώδικα Python, έγινε χρήση της Nipyre⁶², μιας open source βιβλιοθήκης σχεδιασμένης για την διαχείριση νευροαπεικονιστικών δεδομένων.

Στην **Εικόνα 3.6** βλέπουμε τρεις τομές μιας εικόνας η οποία έχει υποστεί όλα τα βήματα της προεπεξεργασίας. Πιο συγκεκριμένα βλέπουμε το εγκάρσιο, στεφανιαίο και οβελιαίο επίπεδο μιας εικόνας στην οποία αρχικά εφαρμόσαμε resampling, στη συνέχεια την καταχωρήσαμε στον atlas και τέλος αφαιρέσαμε την πληροφορία του κρανίου κάνοντας skull stripping. Σε αυτήν την περίπτωση, παρατηρούμε ότι δεν έχει μείνει κάποιο υπόλειμμα από το κρανίο και παράλληλα δεν έχουμε απώλεια πληροφορίας ιστού του εγκεφάλου. Παρόλα αυτά, σε άλλες εικόνες μπορεί να έχει συμβεί κάτι από τα δύο, δεδομένου ότι η τιμή του κατωφλιού του skull stripping δεν είχε ιδανικά αποτελέσματα σε όλες τις εικόνες.



Εικόνα 3.6 Εγκάρσιο, στεφανιαίο και οβελιαίο επίπεδο, αντιστοίχως, μιας καταχωρημένης εικόνας στην οποία έχει αφαιρεθεί το κρανίο

Οι τεχνικές προεπεξεργασίας ιατρικών εικόνων μπορούν να γίνουν εξαιρετικά πολύπλοκες, αλλά στα πλαίσια της παρούσας εργασίας χρησιμοποιήθηκε μια πολύ απλή προσέγγιση. Έγινε χρήση βιβλιοθηκών ανοιχτού κώδικα και εργαλείων με σκοπό το καλύτερο δυνατό αποτέλεσμα, χωρίς να διαχειριζόμαστε κάθε εικόνα ξεχωριστά. Με αυτόν τον τρόπο μπορούμε να ελέγξουμε το εύρος των δυνατοτήτων που μας παρέχουν τα διαθέσιμα εργαλεία στην ανάπτυξη τέτοιων συστημάτων, καθώς η επεξεργασία κάθε εικόνας μεμονωμένα μπορεί να μην είναι δυνατή για ερευνητές και επιστήμονες που δεν έχουν γνώση στον τομέα της ραδιολογίας.

3.2.2 Μετατροπή σε 2D και διαχωρισμός των δεδομένων (dataset split)

Στην παρούσα εργασία, θα χρησιμοποιηθούν μοντέλα αναγνώρισης εικόνας τα οποία έχουν ήδη προ-εκπαιδευτεί στο ImageNet και τα οποία θα κάνουμε fine-tune στο πρόβλημα μας. Αυτά τα μοντέλα είναι φτιαγμένα να δέχονται ως είσοδο RGB εικόνες, δηλαδή εικόνες οι οποίες έχουν δύο διαστάσεις για την δομή και μια για τα τρία κανάλια του χρώματος. Συνεπώς, δεν μπορούν να δεχτούν ως είσοδο τις εικόνες του συνόλου δεδομένων μας ως έχουν. Θα χρειαστεί να υλοποιήσουμε μια μέθοδο ώστε να μετασχηματίσουμε τις εικόνες του συνόλου δεδομένων μας στις δύο διαστάσεις. Για τον μετασχηματισμό 3D σε 2D, έχει εφαρμοστεί ένας απλός αλγόριθμος, εμπνευσμένος από την εργασία των Ding et al. και Plasencia^{53,63}.

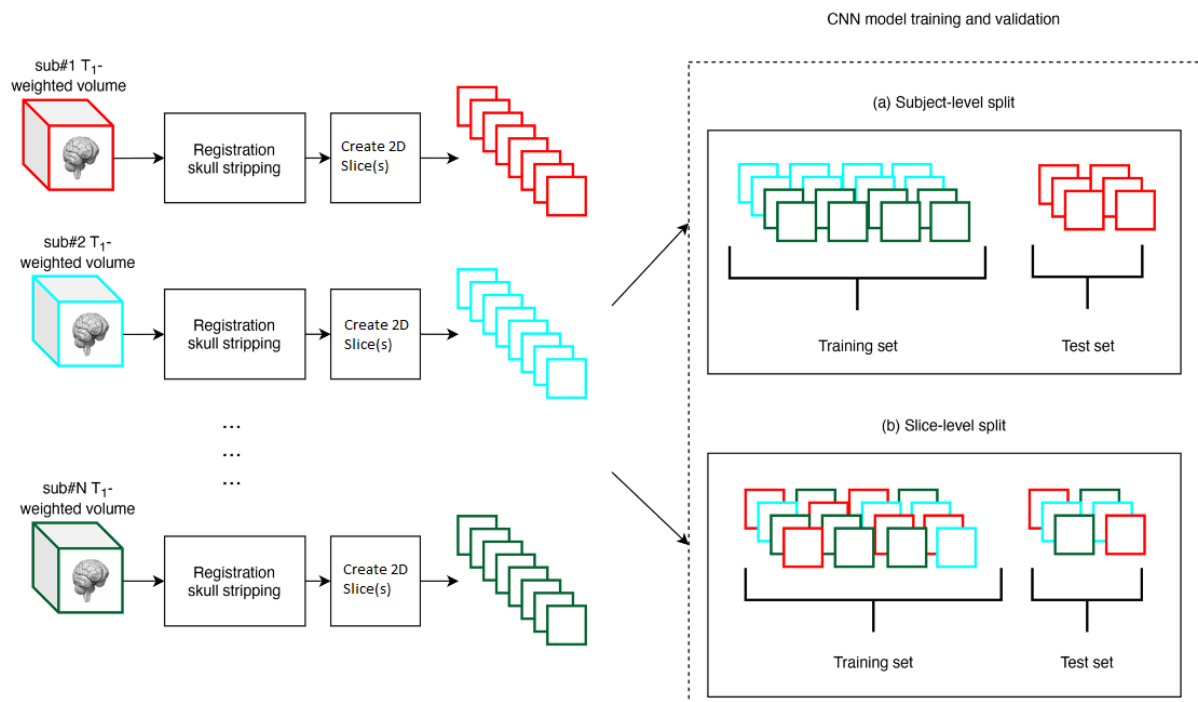
Το σύνολο δεδομένων ADNI αποτελεί ένα longitudinal dataset, δηλαδή πρόκειται για ένα σύνολο δεδομένων στο οποίο υπάρχουν πολλαπλές σαρώσεις MRI για τα ίδια υποκείμενα, σε βάθος χρόνου. Αυτό επιτρέπει τον διαχωρισμό των δεδομένων με δύο τρόπους: (α) διαχωρισμός σε επίπεδο υποκειμένου (subject-level split), (β) διαχωρισμός σε επίπεδο εγκάρσιων τομών (slice-level split).

Η αμερόληπτη αξιολόγηση των αλγορίθμων ταξινόμησης είναι καθοριστική ώστε να αξιολογηθεί η πιθανή κλινική τους αξία. Μια σημαντική πηγή προκατάληψης (bias) είναι η διαρροή δεδομένων (data leakage), δηλαδή η χρήση δεδομένων από το σύνολο ελέγχου (test data) σε οποιοδήποτε μέρος της εκπαιδευτικής διαδικασίας. Σε περίπτωση που ο διαχωρισμός των δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου δεν γίνει σε επίπεδο υποκειμένου, τότε δεδομένα από το ίδιο υποκείμενο

μπορούν να εμφανιστούν σε περισσότερα από ένα σύνολα. Αυτό το πρόβλημα μπορεί να εμφανιστεί όταν πολλαπλά slices έχουν εξαχθεί από μια 3D εικόνα ή όταν εικόνες από το ίδιο υποκείμενο είναι διαθέσιμες σε διαφορετικές χρονικές στιγμές. Επίσης το σύνολο ελέγχου είναι απαραίτητο να χρησιμοποιείται μόνο για την αξιολόγηση της τελικής επίδοσης του μοντέλου και να μην εμφανίζεται καθόλου κατά την διάρκεια της εκπαιδευτικής διαδικασίας.⁶⁴

Οι μελετητές Backstrom et al.⁶⁵ παρατηρήσαν, ότι σε ένα longitudinal dataset, η χρήση διαχωρισμού στο επίπεδο της εικόνας μπορεί να οδηγήσει σε αύξηση της τιμής της ορθότητας (accuracy) κατά 8 ποσοστιαίες μονάδες σε σύγκριση με αυτή που θα προέκυπτε αν ο διαχωρισμός γινόταν σε subject-level. Πιο συγκεκριμένα χρησιμοποίησαν ένα υποσύνολο του ADNI με 1198 MRI από 340 υποκείμενα που ανήκαν σε δύο κατηγορίες (AD/CN). Συνεπώς, η επίδραση που παρατηρήθηκε αφορά στο πρόβλημα της δυαδικής ταξινόμησης και μπορεί να γίνει εντονότερη στο διαχωρισμό πολλαπλών κλάσεων (AD/MCI/CN). Επίσης, περισσότερες από τις μισές δημοσιεύσεις που αφορούν ταξινόμηση της νόσου του Alzheimer μέσω MRI με συνελκτικά νευρωνικά δίκτυα (CNN), φαίνεται να παρουσιάζουν κάποια μορφή διαρροής δεδομένων.⁶⁴

Στην παρούσα εργασία θα γίνει μελέτη και των δύο τρόπων διαχωρισμού των δεδομένων, ούτως ώστε να μελετηθεί η επίδραση που έχει στην απόδοση των μοντέλων.

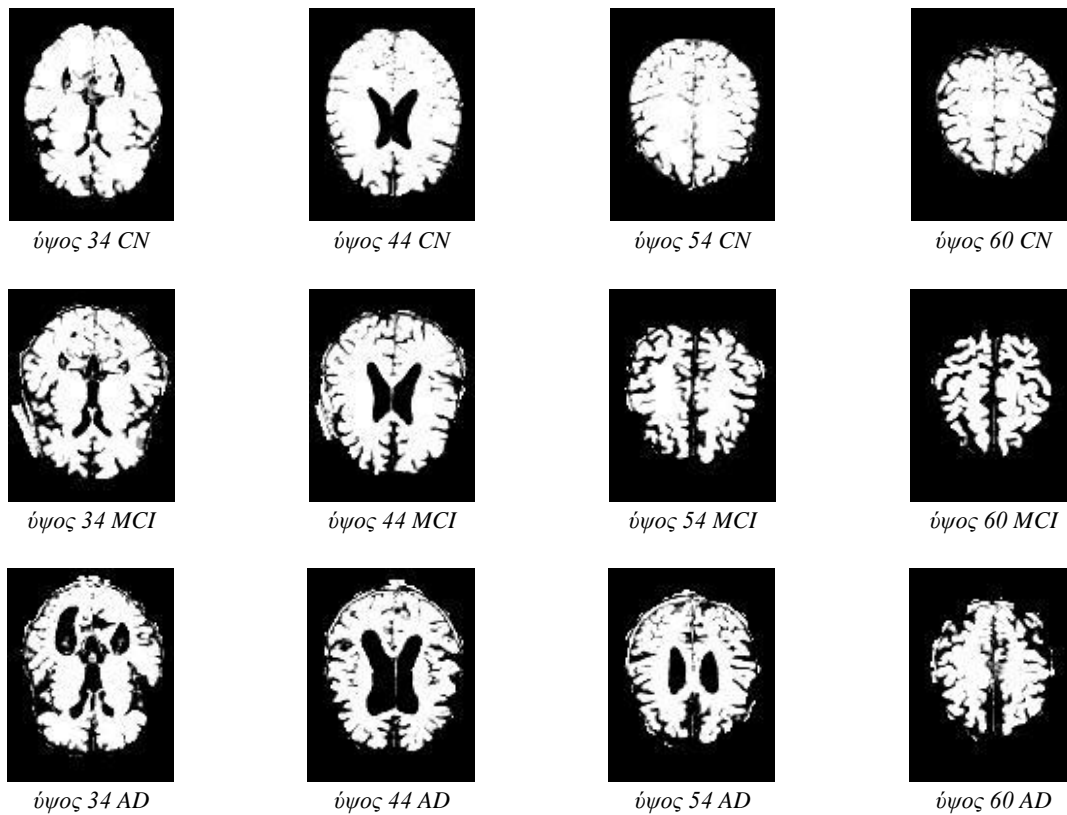


Εικόνα 3.7: Διάγραμμα της επεξεργασίας των εικόνων και του διαχωρισμού σε δεδομένα εκπαίδευσης και ελέγχου⁶⁶

Πριν τη διαδικασία εξαγωγής των slices οι τρισδιάστατες εικόνες κανονικοποιούνται, ούτως ώστε να έχουν μηδενική μέση τιμή και τυπική απόκλιση ίση με τη μονάδα. Με άλλα λόγια, εφαρμόζεται μια διαδικασία whitening. Ο αλγόριθμος μετατροπής των δεδομένων σε 2D επιλέγει ένα ή περισσότερα slices ανάλογα με τον τύπο του διαχωρισμού. Κάθε slice αναπαράγεται τρεις φορές, ώστε να ταιριάζει με τις διαστάσεις του RGB.

Πιο συγκεκριμένα όσον αφορά το slice-level split, έγινε επιλογή μόνο του μεσαίου slice από κάθε 3D T1-weighted MRI, ούτως ώστε να περιοριστεί το data leakage. Έτσι προκύπτει μόνο ένα slice για κάθε επίσκεψη ενός ατόμου με διαστάσεις 110x86x3, το οποίο δεν μπορεί να βρίσκεται ταυτόχρονα στο training set και στο test set. Παρόλα αυτά, ένας ασθενής μπορεί να έχει πολλαπλές επισκέψεις και έτσι, κάποιο slice από προγενέστερη ή μεταγενέστερη επίσκεψη είναι δυνατό να βρεθεί στο test set.

Για το subject level διαχωρισμό δεν υπήρχε περιορισμός ως προς τον αριθμό των slices που μπορεί να χρησιμοποιηθεί από κάθε MRI, δεδομένου ότι εξασφαλίζεται ο διαχωρισμός των δεδομένων ανά υποκείμενο στα σύνολα εκπαίδευσης και ελέγχου. Παρόλα αυτά, έγινε επιλογή 4 slices από κάθε MRI έτσι ώστε ο αριθμός των εκπαιδευτικών παραδειγμάτων να είναι διαχειρίσιμος από τους υπολογιστικούς πόρους του συστήματος που χρησιμοποιήθηκε για τα πειράματα. Πιο συγκεκριμένα έγινε επιλογή 4 slices σε τιμές εγκάρσιου ύψους μεταξύ 28 και 60, οι οποίες μετά από πολλαπλές δοκιμές κρίθηκε ότι αναπαριστούν επαρκή ποσότητα πληροφορίας. Τα slices που είχαν ύψος μεγαλύτερο από 60 έδειχναν το άνω μέρος του εγκεφάλου από πολύ μακριά, ενώ τα slices που είχαν ύψος μικρότερο από 28 αποτύπωναν κυρίως περιοχές του εγκεφαλικού στελέχους. Τα slices που επιλέχθηκαν είχαν ύψος 34, 44, 54 και 60 ούτως ώστε να καλύπτουν ένα μεγάλο εύρος από τον ιστό του εγκεφάλου και να παρέχουν διαφορετικές οπτικές από τις οποίες θα μπορούσαν να προκύψουν ενδιαφέρουσες πληροφορίες. Στην **Εικόνα 3.8** παρουσιάζονται ενδεικτικά παραδείγματα.



Εικόνα 3.8: Αποτελέσματα μετατροπής των MRI σε 2D εικόνες

Πίνακας 3.2: Κατανομή των *subject splitted* εικόνων σε κλάσεις

Κλάση	Πλήθος
AD	1584
MCI	2368
CN	2160
Σύνολο	6112

Έτσι προκύπτει ένα καινούριο σύνολο δεδομένων το οποίο αποτελείται συνολικά από 6112 εικόνες. Η επίδοση των μοντέλων στην ταξινόμηση για 3 κλάσεις στο *subject level split* δεν ήταν ικανοποιητική με τιμές αντίστοιχες της μελέτης των Valliani και Soni⁶⁷ και έτσι επιλέξαμε να αντιμετωπίσουμε το πρόβλημα ως δυαδική ταξινόμηση (*binary classification*), δηλαδή CN vs AD, CN vs MCI και MCI vs AD, όπως συνηθίζεται και στην βιβλιογραφία^{65,67-72}.

Ο διαχωρισμός των υγείων ατόμων (CN) από τους ασθενείς με νόσο του Alzheimer (AD), είναι η πιο συχνά εμφανιζόμενη στην βιβλιογραφία, παρότι η κλινική σημασία αυτού του προβλήματος είναι περιορισμένη. Πιο συγκεκριμένα 25 από τις 32 εργασίες που εξετάστηκαν από τους ερευνητές Wen et al.⁶⁴, πραγματεύονταν αυτόν τον διαχωρισμό.

Επίσης ο διαχωρισμός των υγείων ατόμων (CN) από αυτούς με ήπια γνωστική διαταραχή αποτελεί ένα πολύ ενδιαφέρον πρόβλημα, αφού η έγκαιρη διάγνωση μιας τέτοιας κατάστασης είναι ιδιαίτερα σημαντική για την έγκαιρη φαρμακευτική παρέμβαση και κατά συνέπεια την πρόοδο και εξέλιξη της νόσου. Συνολικά 9 μελέτες εξέτασαν αυτόν τον διαχωρισμό με βάση την έρευνα των Wen et al.

Αντίστοιχα και ο διαχωρισμός των ασθενών με MCI και AD παρουσιάζει ενδιαφέρον για την παρακολούθηση και κατανόηση της εξέλιξης της νόσου. Βάσει την μελέτη των Wen et al. αυτού του είδους ο διαχωρισμός εξετάστηκε σε 7 εργασίες.

Στον **Πίνακα 3.3** βλέπουμε την κατανομή του νέου συνόλου δεδομένων σε δεδομένα εκπαίδευσης (*training set*), επικύρωσης (*validation set*) και ελέγχου (*test set*).

Πίνακας 3.3: Κατανομή των εικόνων σε δεδομένα εκπαίδευση, επικύρωσης και ελέγχου.

	CN	MCI	AD
<i>Training Set</i>	1728	1892	1264
<i>Validation Set</i>	216	236	156
<i>Test Set</i>	216	240	164

3.3 Μοντέλα

Σε αυτήν την ενότητα θα γίνει μια σύντομη περιγραφή των μοντέλων που χρησιμοποιήθηκαν στην εργασία. Πιο συγκεκριμένα επιλέχθηκαν 4 μοντέλα από διαφορετικές οικογένειες δικτύων:

Πίνακας 3.4: Πλήθος παραμέτρων των μοντέλων που χρησιμοποιήθηκαν

Μοντέλο	Παράμετροι
VGG16	138.4 εκατομμύρια
ResNet50V2	25.6 εκατομμύρια
Xception	22.9 εκατομμύρια
EfficientNetV2-M	54.4 εκατομμύρια

3.3.1 VGG16

Το VGG16³⁸ είναι ένα βαθύ συνελκτικό νευρωνικό δίκτυο 16 στρωμάτων το οποίο δημιουργήθηκε από το Visual Geometry Group (VGG), των Karen Simonyan και Andrew Zisserman. Με αυτό το μοντέλο συμμετείχαν στον ετήσιο διαγωνισμό ImageNet Challenge 2014 (ILSVRC), όπου και κέρδισαν την 1^η και 2^η θέση στις κατηγορίες ταξινόμησης (classification) και εντοπισμού (localization) αντιστοίχως. Βασίστηκε στην αρχιτεκτονική AlexNet⁷³, με κύρια ιδέα την εμβάθυνση του δικτύου.

Τα κυριότερα χαρακτηριστικά της αρχιτεκτονικής του μοντέλου είναι ότι :

- Χρησιμοποιεί συνελκτικά στρώματα που έχουν μέγεθος πυρήνα (kernel size) ίσο με 3x3, με βήμα (stride) ίσο με 1 ενώ το γέμισμα (padding) είναι τέτοιο ώστε να διατηρούνται σταθερές οι διαστάσεις του ύψους και πλάτους κάθε εικόνας ενώ αυξάνεται ο αριθμός των φίλτρων.
- Χρησιμοποιεί Στρώματα Μέγιστης Συγκέντρωσης (Max Pooling Layers) μεγέθους 2x2 με βήμα = 2. Αυτό έχει ως αποτέλεσμα να μειώνονται στο μισό οι διαστάσεις του ύψους και πλάτους της εικόνας, ενώ παραμένει σταθερός ο αριθμός των φίλτρων.

Το VGG αποτελεί σημαντικό μοντέλο γιατί εισήγαγε την έννοια της εμβάθυνσης του δικτύου για την εξαγωγή περισσότερων χαρακτηριστικών. Επίσης, το μέγεθος των φίλτρων στα συνελκτικά στρώματα είναι αρκετά μικρότερο συγκριτικά με τα μέχρι τότε χρησιμοποιούμενα δίκτυα, π.χ. το AlexNet χρησιμοποιούσε φίλτρα μεγέθους 11x11.

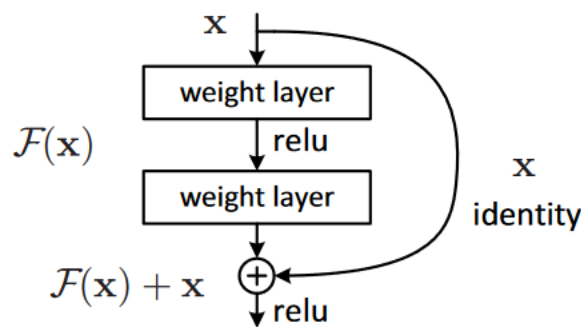


Εικόνα 3.9: Αρχιτεκτονική VGG16⁷⁴

3.3.2 ResNet50

Οι αρχιτεκτονικές των ResNet⁴¹ δημιουργήθηκαν από τους He et al., μέλη της ερευνητικής ομάδας της Microsoft και δημοσιεύτηκαν στο άρθρο “Deep Residual Learning for Image Recognition”, το 2015. Βασίζόμενοι σε αυτές τις αρχιτεκτονικές κατόρθωσαν να κερδίσουν την πρώτη θέση στους διαγωνισμούς ImageNet Challenge (ILSVRC) και COCO το 2015, στις κατηγορίες Imagenet detection, Imagenet localization, COCO detection, και COCO segmentation.

Μετά την υλοποίηση του VGG του 2014, είχε ξεκινήσει η τάση για τοποθέτηση πολλών διαδοχικών συνελκτικών επιπέδων στις αρχιτεκτονικές των δικτύων βαθιάς μάθησης. Αυτό οδήγησε στην παρουσίαση του προβλήματος του vanishing gradient, το γεγονός δηλαδή ότι κατά την οπισθοδιάδοση (backpropagation) η παράγωγος λόγω της τιμής των βαρών γίνεται τόσο μικρή με αποτέλεσμα να αδυνατεί να επηρεάσει τα υπόλοιπα βάρη με αποτέλεσμα την επιβράδυνση ή ακόμα το σταμάτημα της εκπαίδευσης του δικτύου. Η απάντηση σε αυτό το πρόβλημα δόθηκε από τους He et al. μέσω των Residual Networks ή ResNets. Το βασικό χαρακτηριστικό των ResNet είναι τα residual blocks τα οποία στην ουσία αποτελούν συνδέσεις που επιτρέπουν την παρακάμψη ενός ή περισσότερων επιπέδων.

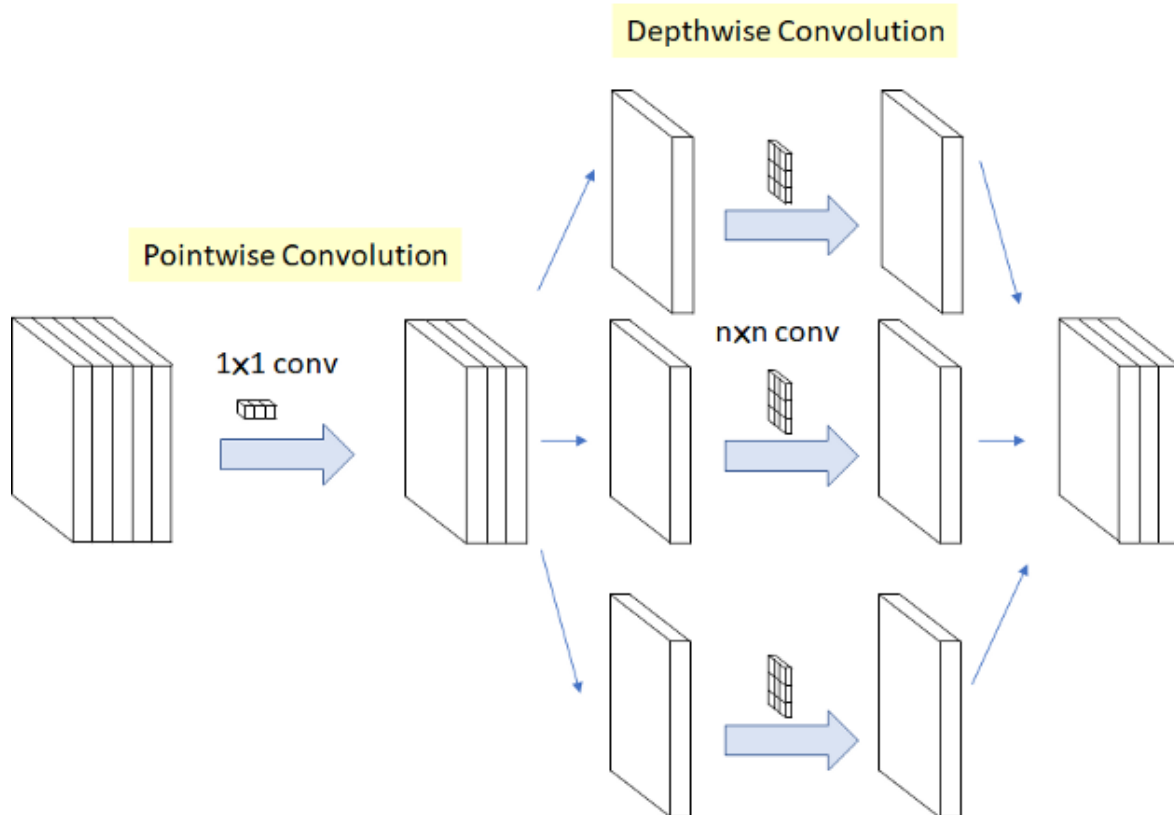


Εικόνα 3.10: Η δομή ενός Residual block⁴¹

Η χρήση των παρακάμψεων μέσω των residual blocks, επέτρεψε στα ResNets να έχουν πολλά περισσότερα επίπεδα σε σχέση με τα έως τότε προηγούμενα μοντέλα. Για παράδειγμα το VGG είχε μέχρι 19 επίπεδα και το GoogleNet³⁹ (ή αλλιώς InceptionV1) μέχρι 22. Αντιθέτως τα ResNets κατάφεραν να φτάσουν τα 50, 101 και 152 επίπεδα με τις αρχιτεκτονικές ResNet50, ResNet101 και ResNet152 αντιστοίχως. Στην παρούσα εργασία χρησιμοποιήθηκε το ResNet50, δηλαδή το ResNet των 50 επιπέδων.

3.3.3 Xception

Το δίκτυο Xception⁴² δημιουργήθηκε από τον Francois Chollet, μέλος της ερευνητικής ομάδας της Google και δημοσιεύτηκε στο άρθρο “Xception: Deep Learning with Depthwise Separable Convolutions”, το 2016. Πρόκειται για ένα βαθύ συνελκτικό δίκτυο το οποίο βασίζεται στο InceptionV3⁴⁰, αλλά χρησιμοποιεί μια παραλλαγή των συνελίξεων του Inception, την οποία ονομάζει depthwise separable convolution, ενώ επίσης χρησιμοποιεί και συνδέσεις παράκαμψης (skip connections), όπως τα ResNets. Η τεχνική depthwise separable convolution αποτελείται από το συνδυασμό της συνέλιξης κατά βάθος (depthwise convolution) και κατά σημείο (pointwise convolution). Πιο συγκεκριμένα, για κάθε εικόνα εισόδου υπολογίζονται ξεχωριστές αναπαραστάσεις για κάθε κανάλι της εικόνας, εφαρμόζοντας ένα φίλτρο σε ένα κανάλι κάθε φορά. Έτσι, προκύπτει μια αναπαράσταση για κάθε κανάλι και όλες μαζί αποτελούν μια αναπαράσταση με βάθος ίδιο με της αρχικής εικόνας. Οι αναπαραστάσεις αυτές συνδυάζονται μέσω της σημειακής συνέλιξης, χρησιμοποιώντας φίλτρα 1x1, δηλαδή χρησιμοποιώντας ένα pixel-σημείο από κάθε αναπαράσταση. Στο τέλος προκύπτει η έξοδος η οποία περιλαμβάνει πληροφορίες από όλα τα κανάλια της αρχικής εικόνας, όπως συμβαίνει και στην περίπτωση της κλασικής συνέλιξης, αλλά σε αυτή την περίπτωση το υπολογιστικό κόστος είναι μικρότερο. Στο Xception, χρησιμοποιείται η αντίστροφη διαδικασία, εφαρμόζοντας δηλαδή αρχικά σημειακή συνέλιξη με σκοπό τον συνδυασμό των πληροφοριών από όλα τα κανάλια και στη συνέχεια συνέλιξη κατά βάθος, ούτως ώστε να προκύψουν αναπαραστάσεις για όλα τα διαφορετικά κανάλια που προέκυψαν στο πρώτο βήμα. Στην **Εικόνα 3.11** παρουσιάζεται η τροποποιημένη εκδοχή του depthwise separable convolution που αντικαθιστά τα Inception Modules στο Xception, γνωστή και ως η «ακραία» (extreme) έκδοση του Inception, εξ' ου και το όνομα Xception (Xtreme Inception). Για μια εικόνα RGB ως είσοδο έχουμε $n=3$.



Εικόνα 3.11: Depthwise separable convolution στο Xception⁷⁵

3.3.4 EfficientNetV2M

Οι αρχιτεκτονικές EfficientNet δημιουργήθηκαν από τους Mingxing Tan και Quoc V. Le, μέλη της ομάδας της Google Brain και δημοσιεύθηκαν στο άρθρο “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”⁴⁴, το 2019. Βασικός στόχος της μελέτης τους ήταν η δημιουργία μιας αποτελεσματικής στρατηγικής για την κλιμάκωση των νευρωνικών δικτύων, παράλληλα όμως εισήγαγαν και μια νέα οικογένεια νευρωνικών δικτύων τα EfficientNets.

Η κλιμάκωση των νευρωνικών δικτύων αφορά την αύξηση των διαστάσεων του δικτύου, δηλαδή του πλάτους και του βάθους του δικτύου καθώς και την αύξηση της ανάλυσης των εικόνων εισόδου στο δίκτυο. Η ρύθμιση αυτών των τριών παραμέτρων με σκοπό την βελτίωση της επίδοσης είναι μια περίπλοκη διαδικασία, η οποία συνήθως απαιτεί ιδιαίτερη προσοχή και χρόνο. Για αυτό τον σκοπό οι μελετητές Tan et al. ανέπτυξαν ένα συστηματικό τρόπο για την ομοιόμορφη κλιμάκωση (compound scaling) του δικτύου μέσω ενός μόνο συντελεστή (compound coefficient). Το compound scaling βασίζεται στην ιδέα ότι η βέλτιστη εξισορρόπηση του πλάτους, του βάθους και της ανάλυσης ανάλογα με τους διαθέσιμους πόρους θα παράγει και τα βέλτιστα αποτελέσματα. Πιο συγκεκριμένα οι διαστάσεις προσαρμόζονται, μέσω ενός συντελεστή ϕ και σύμφωνα με έναν ορισμένο κανόνα, ως εξής:

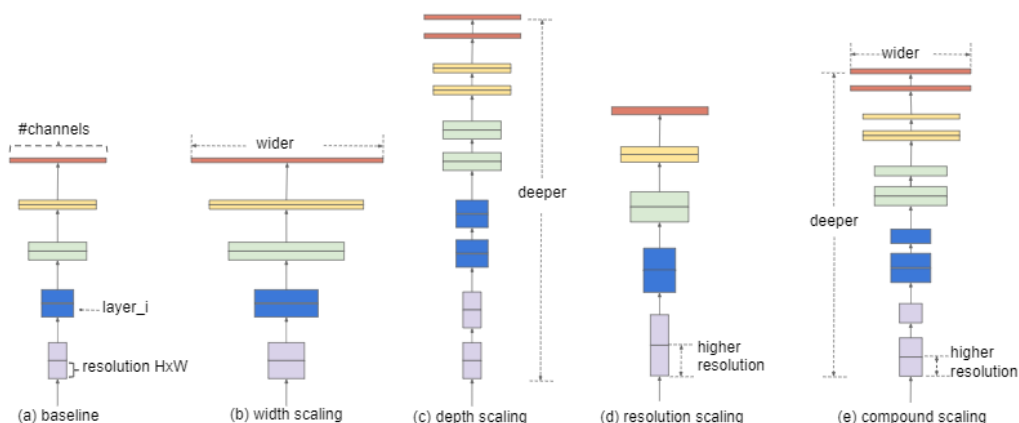
$$depth = \alpha^\phi, width = \beta^\phi, resolution = \gamma^\phi$$

έτσι ώστε,

$$\alpha\beta^2\gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

Η τιμή του συντελεστή ϕ επιλέγεται από τον χρήστη, ο οποίος αφορά τους πόρους σε FLOPs (Floating Point Operations) που μπορούν να διατεθούν για την μεταβολή των τριών διαστάσεων, ενώ οι σταθερές α, β, γ προσδιορίζουν την κατανομή των πόρων στις τρεις διαστάσεις (βάθος, πλάτος και ανάλυση αντιστοίχως). Με άλλα λόγια, πρόκειται για μία μέθοδο η οποία στοχεύει στην ισορροπία μεταξύ των τριών αλληλοεξαρτώμενων διαστάσεων. Τα δίκτυα που προτάθηκαν είναι γνωστά ως EfficientNets, ενώ έχουν διαμορφωθεί ποικίλες παραλλαγές για διαφορετικές τιμές του συντελεστή ϕ , όπως: EfficientNet-B0 έως Efficient-B7, με το βασικό μοντέλο να είναι το Efficient-B0. Οι αρχιτεκτονικές EfficientNetV2⁴³ αποτελούν επέκταση των EfficientNets και βελτιστοποιούν όχι μόνο τις παραμέτρους αλλά και την ταχύτητα της εκπαίδευσης. Στην παρούσα εργασία χρησιμοποιήθηκε το μοντέλο EfficientNetV2-M το οποίο πετυχαίνει καλύτερες τιμές accuracy σε σύγκριση με το EfficientNet-B7, σε συνδυασμό με σημαντικά μικρότερο χρόνο εκπαίδευσης (13 ώρες αντί για 139).



Εικόνα 3.12: Κλιμάκωση πλάτους, βάθους και ομοιόμορφη κλιμάκωση⁴⁴

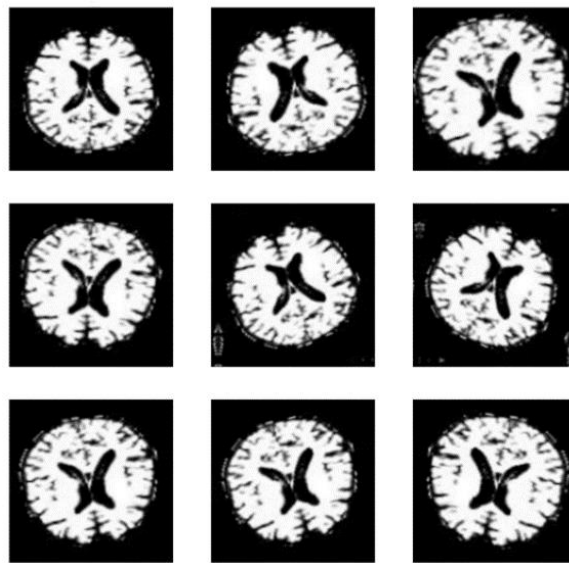
3.4 Εκπαίδευση των μοντέλων

Η διαδικασία που ακολουθήθηκε για την εκπαίδευση των μοντέλων ήταν ίδια για όλα τα πειράματα ανεξάρτητα από τον διαχωρισμό των δεδομένων ούτως ώστε η αναπαραγωγή των πειραμάτων και η σύγκριση των αποτελεσμάτων να γίνεται ευκολότερα.

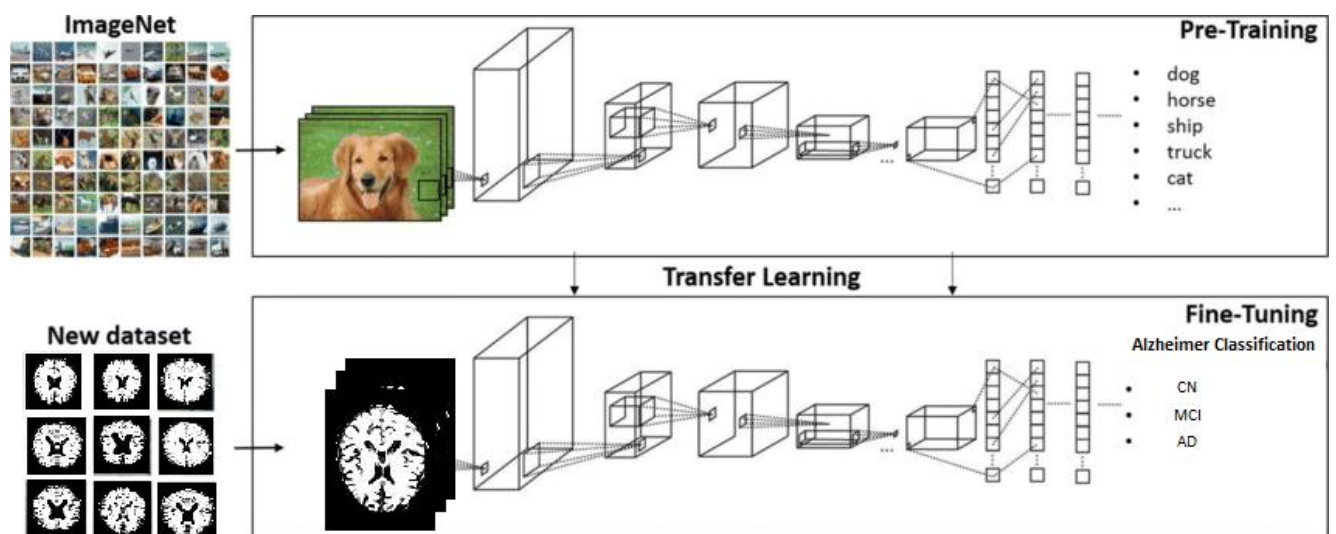
- Έγινε αλλαγή του μεγέθους όλων των εικόνων του συνόλου δεδομένων στην ενιαία ανάλυση των 150x150 pixel. Αυτό έγινε με σκοπό την μείωση της πολυπλοκότητας των υπολογισμών
- Πραγματοποιήθηκε κανονικοποίηση των τιμών των pixel των εικόνων από το σύνολο [0,255] στο σύνολο [0,1] (και [-1, 1] για το Xception). Αυτό έγινε προσθέτοντας ένα Rescaling⁷⁶ layer στην δομή των δικτύων.
- Έγινε χρήση τυχαίου data augmentation (επαύξηση δεδομένων) στα δεδομένα εκπαίδευσης με σκοπό την μείωση του overfitting. Το data augmentation αποτελεί μια καλή πρακτική εισαγωγής ποικιλομορφίας στα δεδομένα με τεχνητό τρόπο, όταν το σύνολο των δεδομένων δεν είναι μεγάλο. Οι τυχαίοι, αλλά ρεαλιστικοί μετασχηματισμοί στις εικόνες εκπαίδευσης, βοηθούν στην έκθεση του μοντέλου σε διαφορετικές πτυχές των δεδομένων, ενώ παράλληλα επιβραδύνει το overfitting.⁷⁷ Πιο συγκεκριμένα πραγματοποιήθηκαν διάφοροι τυχαίοι μετασχηματισμοί στα δεδομένα εισόδου όπως horizontal και vertical flip, rotation με τιμή 0.1 και zoom με τιμή 0.1. Ένα παράδειγμα των μετασχηματισμών που εφαρμόστηκαν φαίνεται στην **Εικόνα 3.13**. Αυτό πραγματοποιήθηκε και σε αυτή την περίπτωση προσθέτοντας αντίστοιχα layers στην δομή των δικτύων.
- Επίσης χρησιμοποιήσαμε ένα Dropout⁷⁸ layer με τιμή 0.2 πριν το στρώμα εξόδου με σκοπό την αποφυγή του overfitting και την καλύτερη γενίκευση του μοντέλου. Αυτό επιτυγχάνεται με την «απενεργοποίηση» κόμβων του νευρωνικού δικτύου κατά την διάρκεια της εκπαιδευτικής διαδικασίας και έχει ως αποτέλεσμα να χρησιμοποιούνται διαφορετικοί συνδυασμοί κόμβων σε κάθε εποχή. Έτσι αποφεύγεται η προσαρμογή του συνόλου των βαρών στα δεδομένα εκπαίδευσης και το μοντέλο γενικεύει καλύτερα.
- Για κάθε αρχιτεκτονική φορτώσαμε ως base model το προεκπαιδευμένο μοντέλο μέσω του Keras Applications⁷⁹ χωρίς να περιλάβουμε το top layer (δηλαδή το στρώμα εξόδου) και πάνω σε αυτό προσθέσαμε ένα Global Average Pooling⁸⁰ layer, το Dropout layer και το Dense layer της εξόδου, οπού ανάλογα με το είδος της ταξινόμησης είχε ένα ή τρεις νευρώνες (binary ή three-way αντιστοίχως).
- Σε κάθε περίπτωση «παγώσαμε» το base model έτσι ώστε να διατηρήσουμε τα βάρη από το ImageNet αέραια και ταυτόχρονα να βρίσκεται σε inference mode ώστε να μείνουν παγωμένα τα Batch Normalization layers των δικτύων όταν θα το «ξεπαγώσουμε» ξανά για να κάνουμε fine-tuning. Η διαδικασία του fine-tuning παρουσιάζεται στην **Εικόνα 3.14**.
- Αρχικά εκπαιδεύσαμε μόνο το top layer χρησιμοποιώντας ως optimizer τον Adam⁸¹ με learning rate = 0.001 και ως loss το Binary Cross Entropy⁸² για την περίπτωση της δυαδικής ταξινόμησης και το Sparse Categorical Crossentropy⁸³ στην περίπτωση του three-way classification για 20 epochs.

- Έπειτα «ξεπαγώσαμε» το base model και πραγματοποιήσαμε επανεκπαίδευση (fine-tuning) όλου του δικτύου για 200 epochs χρησιμοποιώντας ένα μικρό learning rate = $1e-5$.
- Κατά την διάρκεια του fine-tuning έγινε χρήση της callback μεθόδου ModelCheckpoint⁸⁴ ώστε να αποθηκεύουμε τα καλύτερα ενδιάμεσα μοντέλα, με βάση την τιμή του validation accuracy.

Η διεξαγωγή όλων των πειραμάτων πραγματοποιήθηκε στον προσωπικό μου υπολογιστή με μια NVIDIA Geforce RTX 2060 GPU μεγέθους 6 GB με 1920 CUDA cores και 240 tensor cores τα οποία βελτιώνουν την ταχύτητα των πολλαπλασιασμών πινάκων που χρησιμοποιούνται κατά κόρον σε εφαρμογές βαθιάς μηχανικής μάθησης.



Εικόνα 3.13: Τυχαίοι μετασχηματισμοί data augmentation



Εικόνα 3.14: Μεταφορά μάθησης και fine-tuning στο πρόβλημα του Alzheimer Classification

3.5 Μετρικές αξιολόγησης των αποτελεσμάτων

3.5.1 Confusion Matrix (Πίνακας Σύγχυσης)

Ο πίνακας σύγχυσης χρησιμοποιείται για να ορίσει την απόδοση ενός αλγορίθμου ταξινόμησης. Ένας πίνακας σύγχυσης απεικονίζει και συνοψίζει την απόδοση ενός αλγορίθμου ταξινόμησης. Για την δυαδική ταξινόμηση (binary classification) ο πίνακας διαμορφώνεται όπως φαίνεται στην **Εικόνα 3.15**.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Εικόνα 3.15: Confusion Matrix⁸⁵

Όπου έχουμε:

- **True Positive** (Αληθώς θετικά): ο αριθμός των δειγμάτων που ταξινομήθηκαν ως θετικά και είναι πράγματι θετικά.
- **False Positive** (Ψευδώς θετικά): ο αριθμός των δειγμάτων που ταξινομήθηκαν ως θετικά ενώ στην πραγματικότητα ήταν αρνητικά.
- **True Negative** (Αληθώς αρνητικά): ο αριθμός των δειγμάτων που ταξινομήθηκαν ως αρνητικά και είναι πράγματι αρνητικά.
- **False Negative** (Ψευδώς αρνητικά): ο αριθμός των δειγμάτων που ταξινομήθηκαν ως αρνητικά ενώ στην πραγματικότητα ήταν θετικά.

Με βάση τους παραπάνω ορισμούς προκύπτουν οι παρακάτω μετρικές:

- **Accuracy** (Ορθότητα): Πρόκειται για την πιο συνηθισμένη μετρική. Η ορθότητα είναι το ποσοστό των αληθών αποτελεσμάτων στον συνολικό αριθμό παραδειγμάτων που εξετάστηκαν.

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

- **Precision** (Ακρίβεια): Η ακρίβεια είναι το ποσοστό των παραδειγμάτων που ορθώς ταξινομήθηκαν θετικά από το συνολικό αριθμό παραδειγμάτων που ταξινομήθηκαν ως θετικά.

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

- **Recall / Sensitivity** (Ανάκληση / Ευαισθησία): Η ανάκληση ή ευαισθησία είναι το ποσοστό των παραδειγμάτων που ορθώς ταξινομήθηκαν θετικά από το σύνολο των θετικών παραδειγμάτων. Η ευαισθησία είναι επίσης γνωστή και ως True Positive Rate (TPR).

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

- **Specificity** (Ειδικότητα): Η ειδικότητα είναι το ποσοστό των παραδειγμάτων που ορθώς ταξινομήθηκαν αρνητικά από το σύνολο των αρνητικών παραδειγμάτων.

$$Specificity = \frac{TN}{TN + FP}$$

- **F1-Score:** Το F1-Score είναι ο αρμονικός μέσος precision και recall.

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

3.5.2 Καμπύλη ROC - AUC

Η καμπύλη ROC προκύπτει από τον σχεδιασμό της ευαισθησίας (true positive rate) στον άξονα y, έναντι του 1 – ειδικότητα (false positive rate) στον άξονα x. Μια καμπύλη ROC, η οποία είναι πάνω στην διαγώνιο γραμμή $y = x$, παράγει ψευδώς θετικά αποτελέσματα με το ίδιο ποσοστό με το οποίο παράγει αληθώς θετικά. Συνεπώς, αναμένουμε από ένα ικανό μοντέλο να έχει μια καμπύλη ROC στο αριστερό τρίγωνο πάνω από την γραμμή $y = x$, η οποία εμφανίζεται ως διακεκομμένη στα διαγράμματα. Η περιοχή κάτω από την καμπύλη ROC (area under curve, AUC) είναι ένα συνολικό μέτρο της ικανότητας ενός διαχωριστή (classifier) να διακρίνει εάν υπάρχει ή όχι μια συγκεκριμένη κατάσταση. Μια AUC ίση με 0.5 αντιπροσωπεύει έναν διαχωριστή χωρίς διακριτική ικανότητα (δηλαδή, όχι καλύτερο από την τύχη), ενώ μια AUC ίση με 1.0 αντιπροσωπεύει έναν διαχωριστή με τέλεια ικανότητα διάκρισης.⁸⁶ Στην **Εικόνα 3.16** φαίνεται μια γραφική επεξήγηση της καμπύλη ROC.

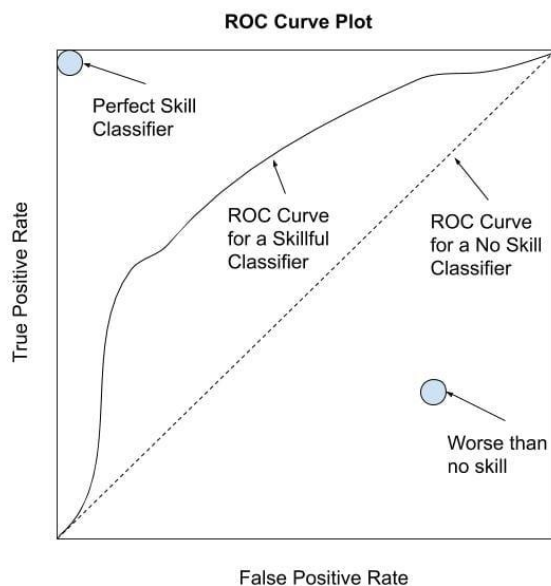
Οι καμπύλες ROC χρησιμοποιούνται συνήθως στη δυαδική ταξινόμηση για τη μελέτη της εξόδου ενός ταξινομητή. Προκειμένου να επεκταθεί η καμπύλη ROC και η AUC σε ταξινόμηση πολλαπλών ετικετών, είναι απαραίτητη η δυαδική μετατροπή της εξόδου. Μια καμπύλη ROC μπορεί να σχεδιαστεί ανά ετικέτα, αλλά μπορεί επίσης να σχεδιαστεί θεωρώντας κάθε στοιχείο του πίνακα ετικετών ως δυαδική πρόβλεψη (micro-averaging). Ένα άλλο μέτρο αξιολόγησης για την ταξινόμηση πολλαπλών ετικετών είναι το macro-averaging, το οποίο δίνει την ίδια βαρύτητα στην ταξινόμηση κάθε ετικέτας και συνιστάται σε περιπτώσεις που παρουσιάζεται ανισορροπία στο σύνολο των δεδομένων και όλες οι κλάσεις είναι εξίσου σημαντικές.⁸⁷

3.5.3 Precision – Recall Curve (Καμπύλη Ακρίβειας – Ανάκλησης)

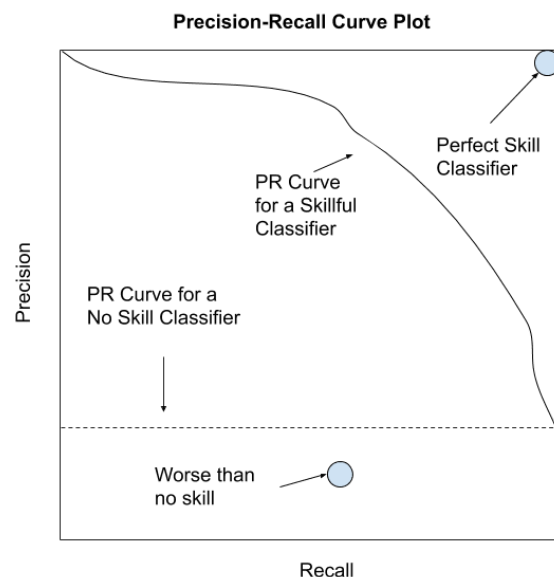
Η καμπύλη ακρίβειας-ανάκλησης απεικονίζει το αντάλλαγμα μεταξύ ακρίβειας και ανάκλησης για διάφορα κατώφλια. Μια υψηλή τιμή στην περιοχή κάτω από την καμπύλη αναπαριστά τόσο υψηλή ανάκληση όσο και υψηλή ακρίβεια, όπου η υψηλή ακρίβεια σχετίζεται με χαμηλό ποσοστό ψευδώς θετικών και η υψηλή ανάκληση σχετίζεται με χαμηλό ποσοστό ψευδώς αρνητικών.

Ένα σύστημα με υψηλή ανάκληση αλλά χαμηλή ακρίβεια επιστρέφει πολλά αποτελέσματα, αλλά οι περισσότερες από τις προβλεπόμενες ετικέτες του είναι εσφαλμένες σε σύγκριση με τις επιθυμητές. Ένα σύστημα με υψηλή ακρίβεια αλλά χαμηλή ανάκληση είναι ακριβώς το αντίθετο, επιστρέφοντας πολύ λίγα αποτελέσματα, αλλά οι περισσότερες από τις προβλεπόμενες ετικέτες του είναι σωστές σε σύγκριση με τις επιθυμητές. Ένα ιδανικό σύστημα με υψηλή ακρίβεια και υψηλή ανάκληση θα επιστρέφει πολλά αποτελέσματα, με όλα τα αποτελέσματα να επισημαίνονται σωστά.⁸⁸ Στην **Εικόνα 3.17** φαίνεται μια γραφική επεξήγηση της καμπύλης Precision-Recall.

Αξίζει να αναφερθεί ότι οι καμπύλες ROC και Precision-Recall σχεδιάστηκαν με την βοήθεια της βιβλιοθήκης Scikit-plot⁸⁹.



Εικόνα 3.16: Επεξήγηση της καμπύλης ROC⁹⁰



Εικόνα 3.17: Επεξήγηση της καμπύλης Precision-Recall⁹⁰

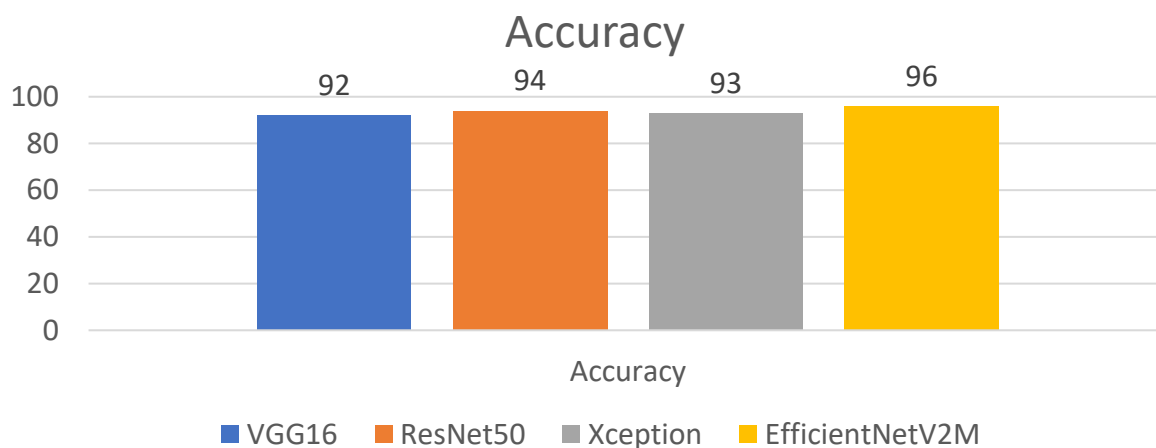
4. Πειραματικά Αποτελέσματα

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα των πειραμάτων μαζί με τις παρατηρήσεις και τα συμπεράσματα που προκύπτουν από αυτά. Πραγματοποιείται μια συγκριτική αξιολόγηση των αποτελεσμάτων που προκύπτουν από τα μοντέλα που χρησιμοποιήθηκαν σε συνδυασμό με τις παρατηρήσεις και τα συμπεράσματα του συγγραφέα.

4.1 Slice Level Split

4.1.1 Μετρικές Αξιολόγησης

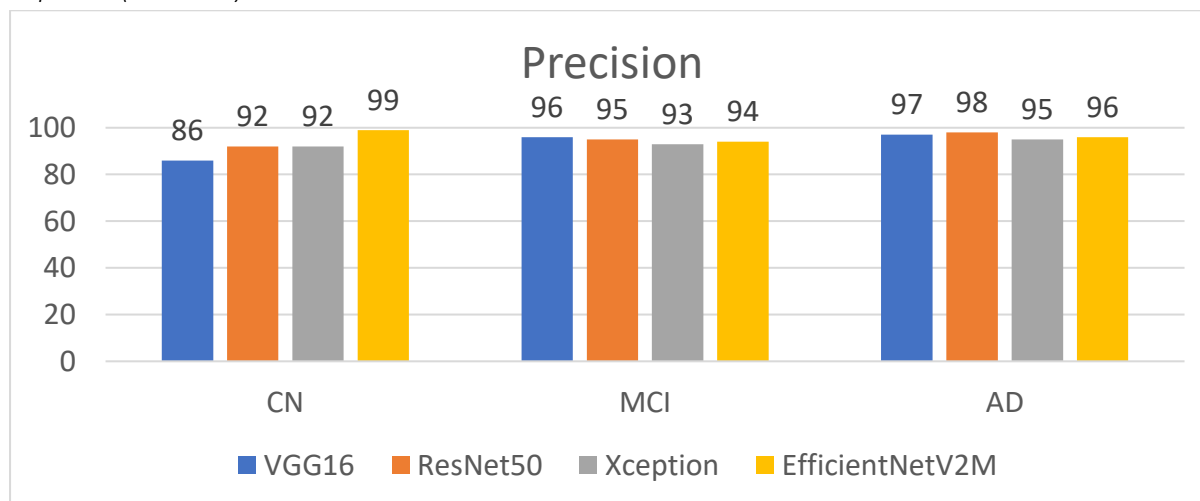
Ορθότητα (Accuracy)



Παρατηρούμε ότι η μέγιστη τιμή ορθότητας (accuracy) επιτεύχθηκε με το μοντέλο EfficientNetV2M και είναι ίση με 96%. Παρόλα αυτά και στα 4 μοντέλα η επίδοση είναι πολύ καλή με τιμές ορθότητας που ξεπερνούν το 90%.

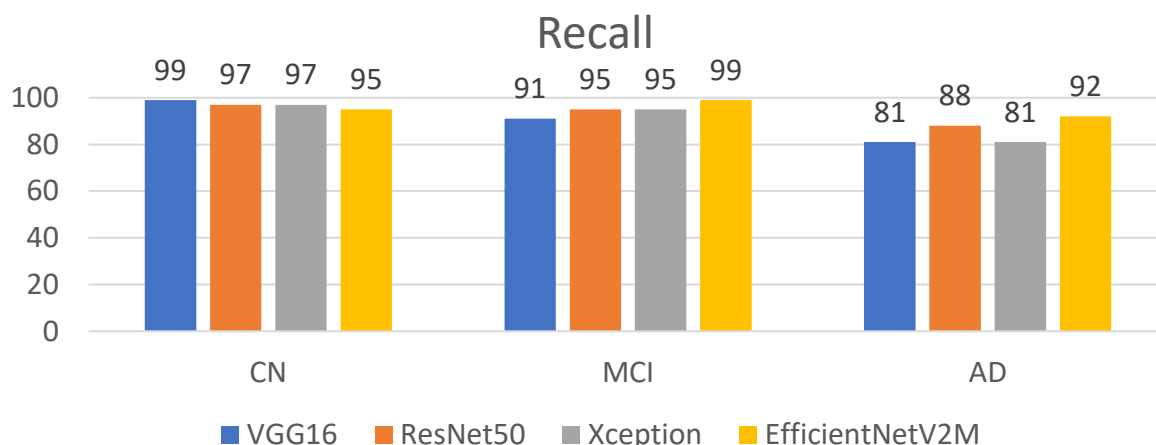
Ο υπολογισμός των μετρικών που ακολουθούν γίνεται για κάθε κλάση ταξινόμησης ξεχωριστά.

Ακρίβεια (Precision)



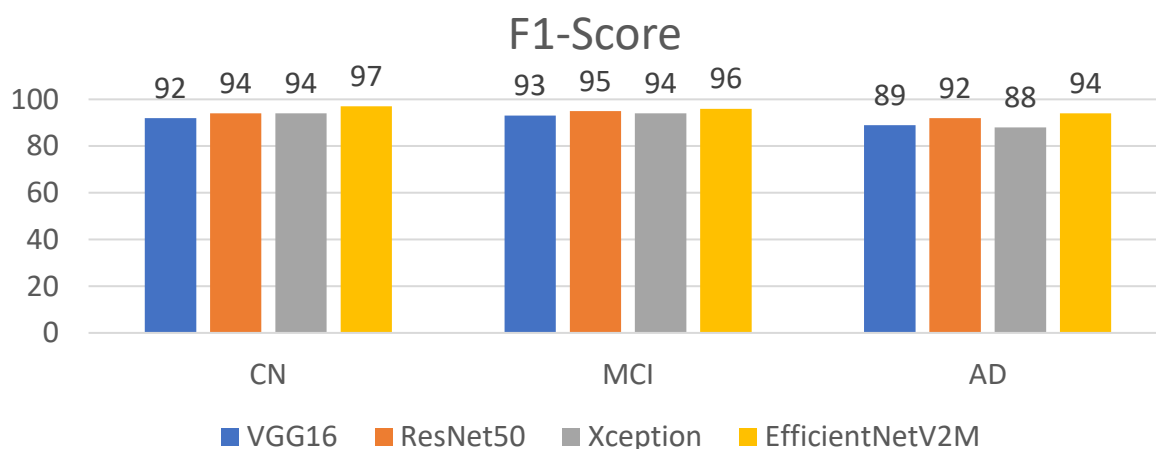
Όσον αφορά το precision παρατηρούμε ότι για την κλάση CN η μέγιστη τιμή προκύπτει από το EfficientNetV2M με τιμή 99%. Για την κλάση MCI η μέγιστη τιμή είναι 96% και προκύπτει από το μοντέλο VGG16. Τέλος για την κλάση AD η μέγιστη τιμή που παρατηρείται είναι 98% και προήλθε από το μοντέλο ResNet50.

Ανάκληση (Recall)



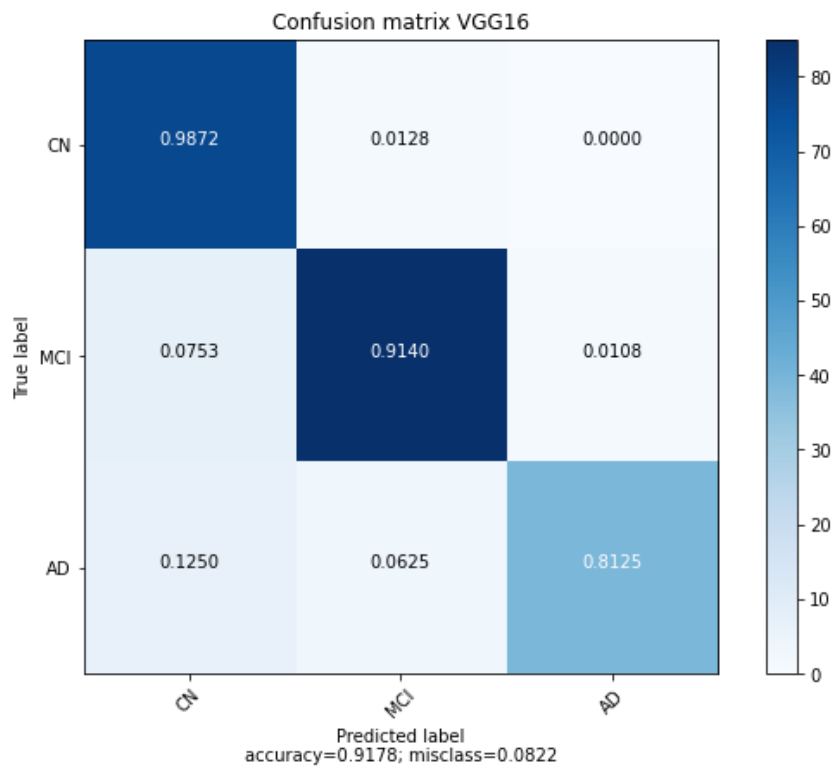
Όσον αφορά το recall, παρατηρούμε ότι για την κλάση CN η μέγιστη τιμή προκύπτει από το VGG16 με τιμή 99%. Για την κλάση MCI η μέγιστη τιμή είναι 99% και προκύπτει από το μοντέλο EfficientNetV2M. Τέλος για την κλάση AD η μέγιστη τιμή που παρατηρείται είναι 92% και προήλθε από το μοντέλο EfficientNetV2M. Συνεπώς παρατηρούμε ότι το μοντέλο EfficientNetV2M παρουσιάζει μεγαλύτερη ευαισθησία στις κλάσεις MCI και AD, συνεπώς είναι πιο ευαίσθητο στην αναγνώριση των χαρακτηριστικών της ήπιας γνωστικής διαταραχής και της νόσου του Alzheimer.

F1-Score

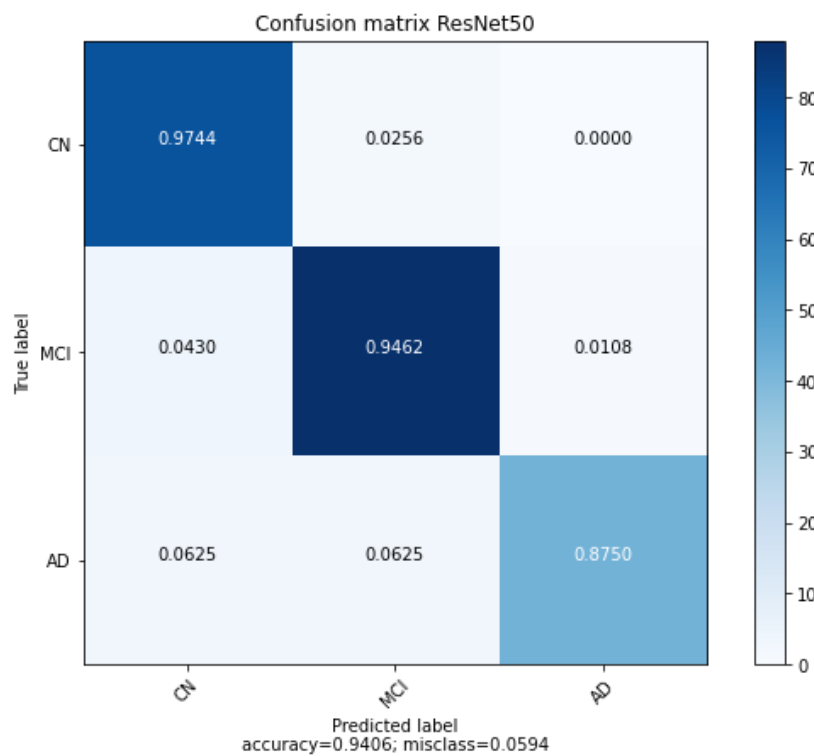


Όσον αφορά το F1-Score παρατηρούμε ότι για την κλάση CN η μέγιστη τιμή προκύπτει από το EfficientNetV2M με τιμή 97%. Για την κλάση MCI η μέγιστη τιμή είναι 96% και προκύπτει πάλι από το μοντέλο EfficientNetV2M. Τέλος για την κλάση AD η μέγιστη τιμή που παρατηρείται είναι 94% και προήλθε και αυτή από το μοντέλο EfficientNetV2M. Άρα συμπεραίνουμε ότι με βάση την μετρική F1-Score το μοντέλο EfficientNetV2M πετυχαίνει την καλύτερη επίδοση και για τις 3 κλάσεις.

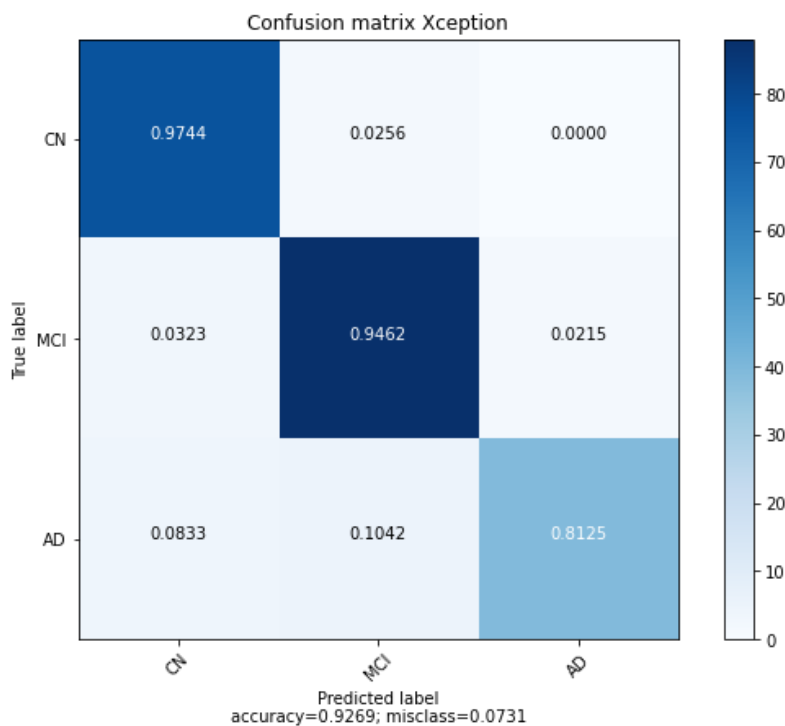
4.1.2 Πίνακες σύγχυσης (Confusion Matrix)



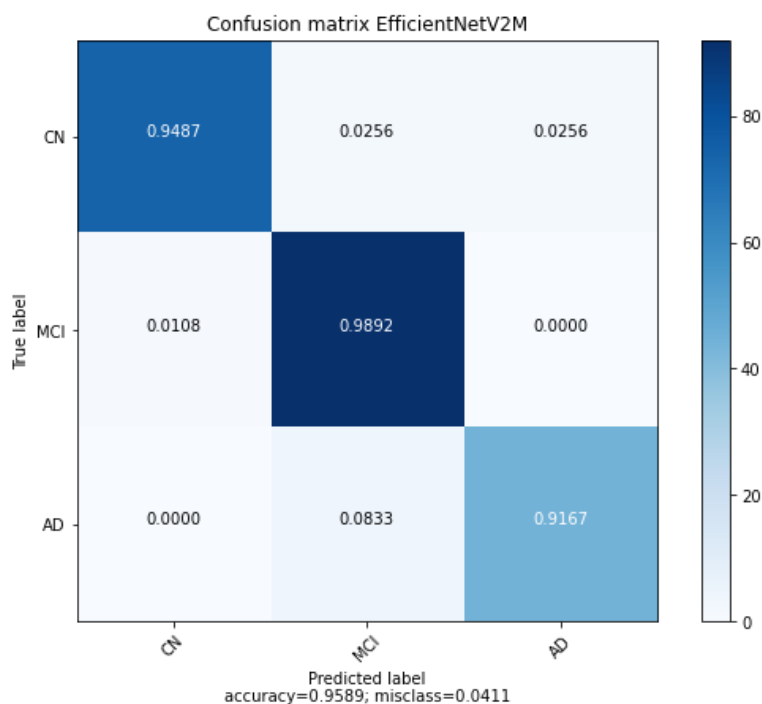
Εικόνα 4.1: Πίνακας Σύγχυσης VGG16



Εικόνα 4.2: Πίνακας Σύγχυσης ResNet50



Εικόνα 4.3: Πίνακας Σύγκρισης Xception



Εικόνα 4.4: Πίνακας Σύγκρισης EfficientNetV2M

Όπως είδαμε και προηγουμένως η μέγιστη ευαισθησία για τις κλάσεις MCI και AD επιτυγχάνεται με το μοντέλο EfficientNetV2M, ενώ για την κλάση CN από το μοντέλο VGG16.

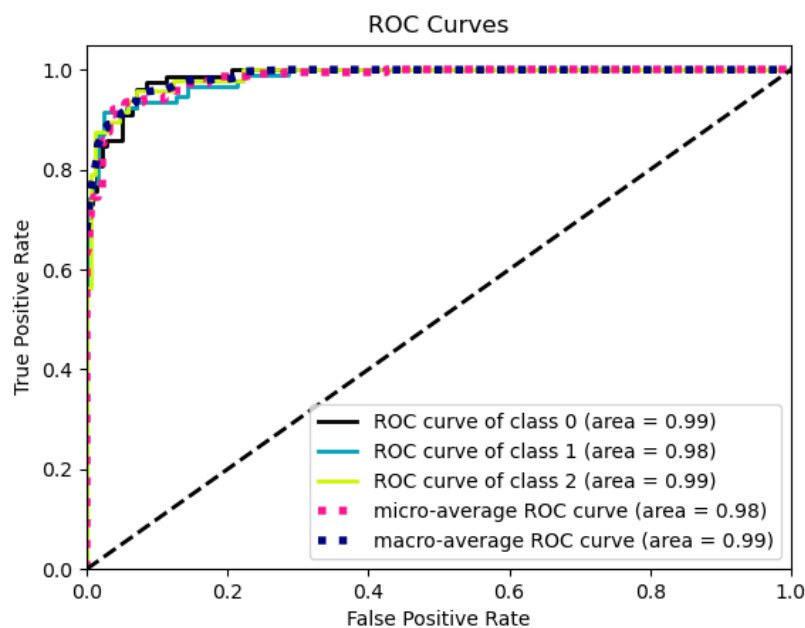
Πιο συγκεκριμένα, παρατηρούμε ότι το μοντέλο EfficientNetV2M παρουσιάζει το μέγιστο ποσοστό ορθώς θετικών (true positive) διαγνώσεων για τις κλάσεις MCI και AD με 99% και 92% αντιστοίχως.

Επίσης παρατηρούμε ότι όλες οι εσφαλμένες ταξινομήσεις της κλάσης AD τοποθετούνται στην κλάση MCI. Το γεγονός αυτό, ενισχύει την αξιοπιστία του μοντέλου καθώς ένας ασθενής που πάσχει από AD δεν υπάρχει πιθανότητα να ταξινομηθεί στην CN κλάση, δηλαδή να θεωρηθεί από το μοντέλο ως υγιής. Αντιθέτως, στα άλλα μοντέλα παρατηρούμε ότι υπάρχει πιθανότητα 12%, 6% και 8% κάποιος ασθενής με AD να ταξινομηθεί ως υγιής, γεγονός το οποίο μπορεί να αποβεί μοιραίο. Αντιστοίχως, για την κλάση MCI παρατηρούμε ότι η ταξινόμηση θα γίνει σωστά με πιθανότητα 99%, ενώ το 1%, που αφορά τις εσφαλμένες ταξινομήσεις, θα ταξινομηθεί στην κλάση CN. Οι εσφαλμένες αυτές ταξινομήσεις μπορεί να προκύπτουν, επειδή οι συγκεκριμένοι ασθενείς έχουν MCI σε πρώιμο στάδιο το οποίο να μην μπορεί να γίνει εύκολα αντιληπτό από τις εικόνες μαγνητικής τομογραφίας.

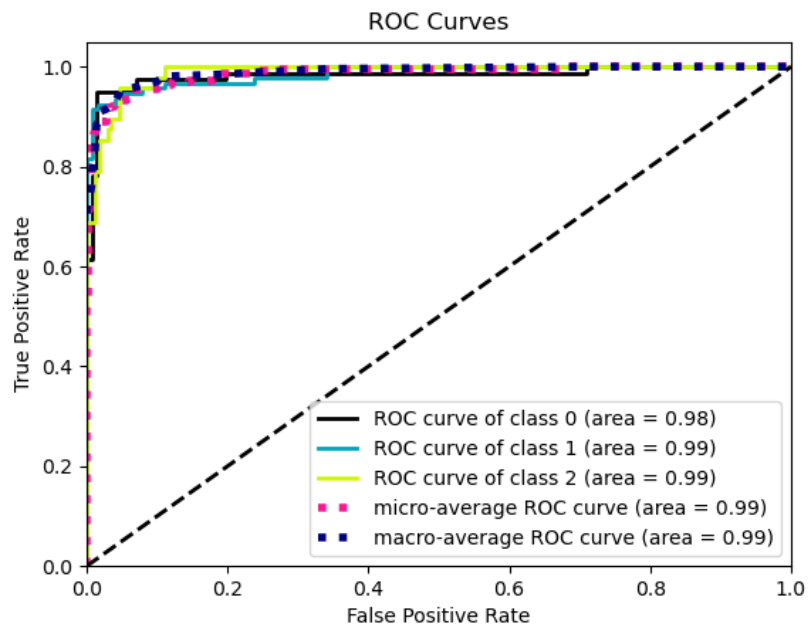
Τέλος, το μοντέλο VGG16 πετυχαίνει το μέγιστο αριθμό ορθών θετικών διαγνώσεων για την κλάση CN. Πιο συγκεκριμένα, η ταξινόμηση θα γίνει σωστά με πιθανότητα 99%, ενώ για το υπόλοιπο 1% των λανθασμένων ταξινομήσεων, αυτές θα γίνουν στην κλάση MCI. Αυτό πιθανώς οφείλεται στον λόγο που επεξηγήθηκε παραπάνω αλλά και στο γεγονός ότι οι ετικέτες των εικόνων προέκυψαν λαμβάνοντας υπόψιν και την κλινική εξέταση των ασθενών. Συνεπώς για τους ασθενείς οι οποίοι βρίσκονται σε πρώιμο στάδιο μιας κατηγορίας MCI ή AD, μπορεί οι αλλαγές που παρουσιάζουν στον εγκέφαλο να μην γίνονται ευκολά αντιληπτές από τα μέσα νευροαπεικόνισης.

4.1.3 Καμπύλη ROC (Receiver Operating Characteristic Curve)

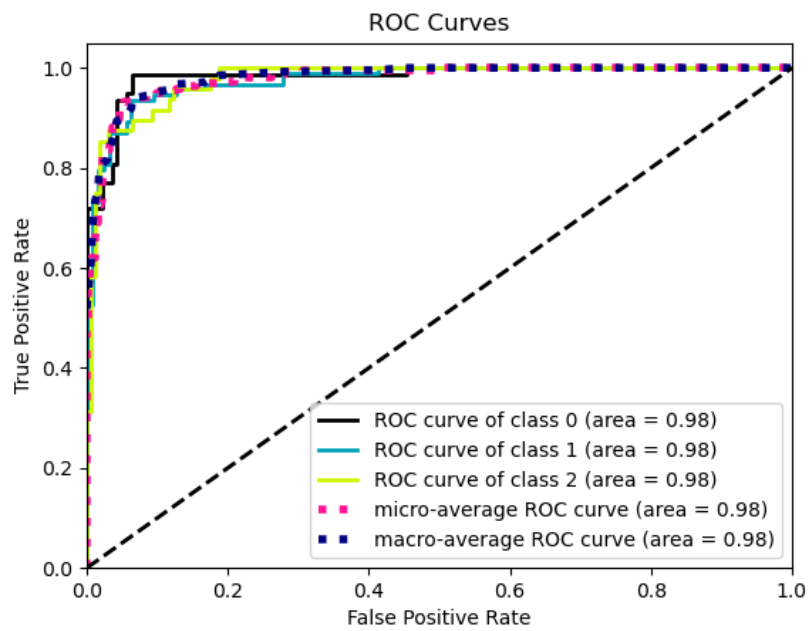
Παρακάτω θα δούμε τις καμπύλες ROC που προκύπτουν από τα 4 μοντέλα μας.



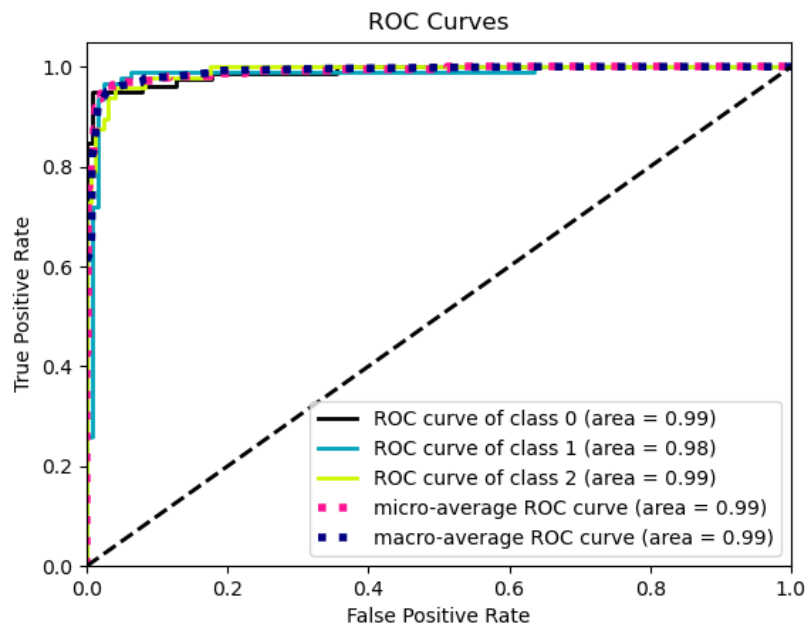
Εικόνα 4.5: Καμπύλη ROC για το μοντέλο VGG16



Εικόνα 4.6: Καμπύλη ROC για το μοντέλο ResNet50



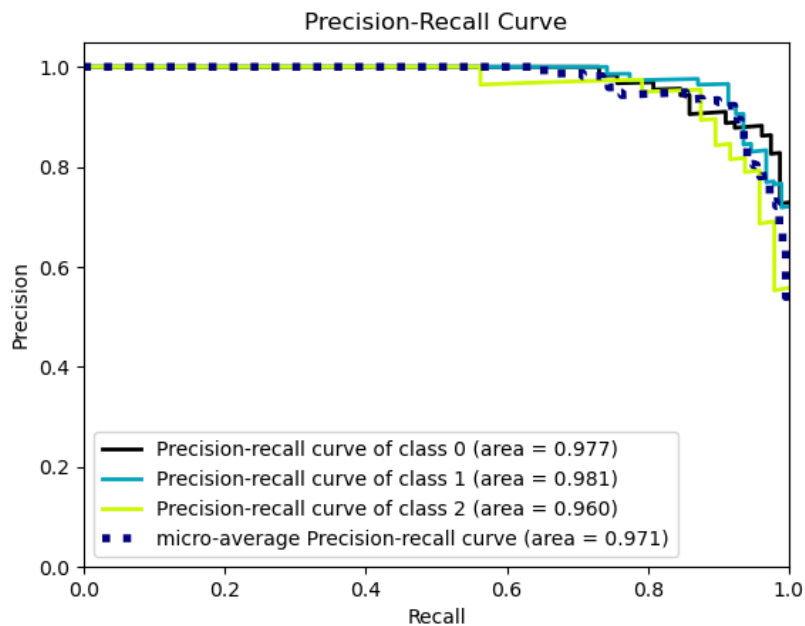
Εικόνα 4.7: Καμπύλη ROC για το μοντέλο Xception



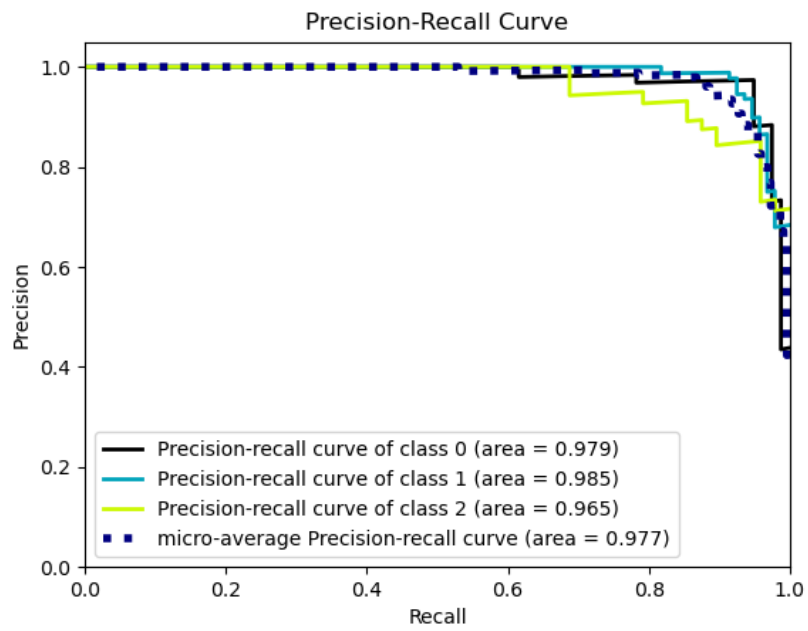
Εικόνα 4.8: Καμπύλη ROC για το μοντέλο EfficientNetV2M

Παρατηρούμε ότι και τα 4 μοντέλα ταξινόμησης καταφέρνουν σχεδόν τέλεια αποτελέσματα με τιμές AUC μεταξύ 0.98 και 0.99. Το Xception φαίνεται να παρουσιάζει χειρότερα αποτελέσματα σε σχέση με τα άλλα μοντέλα με τις τιμές AUC να ισούνται με 0.98 και για τις τρεις κλάσεις.

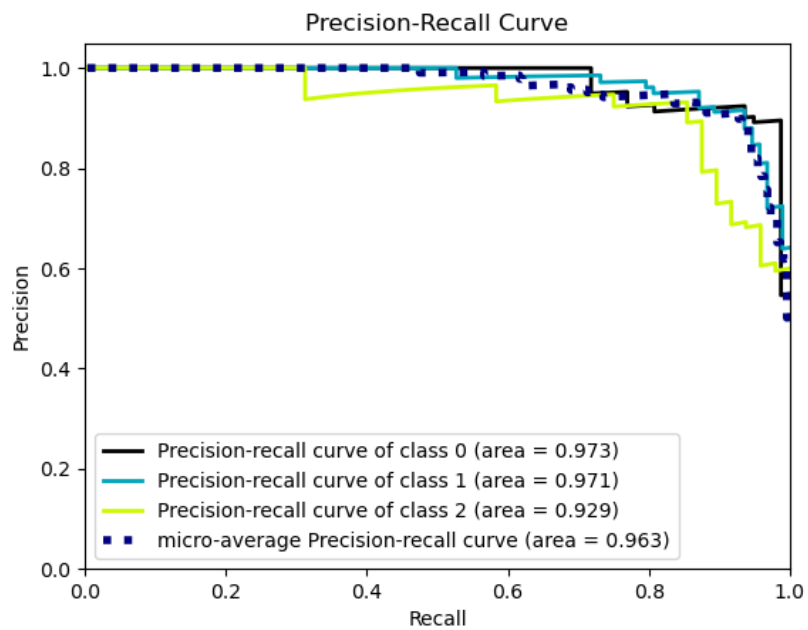
4.1.4 Καμπύλη Ακρίβειας-Ανάκλησης (Precision-Recall Curve)



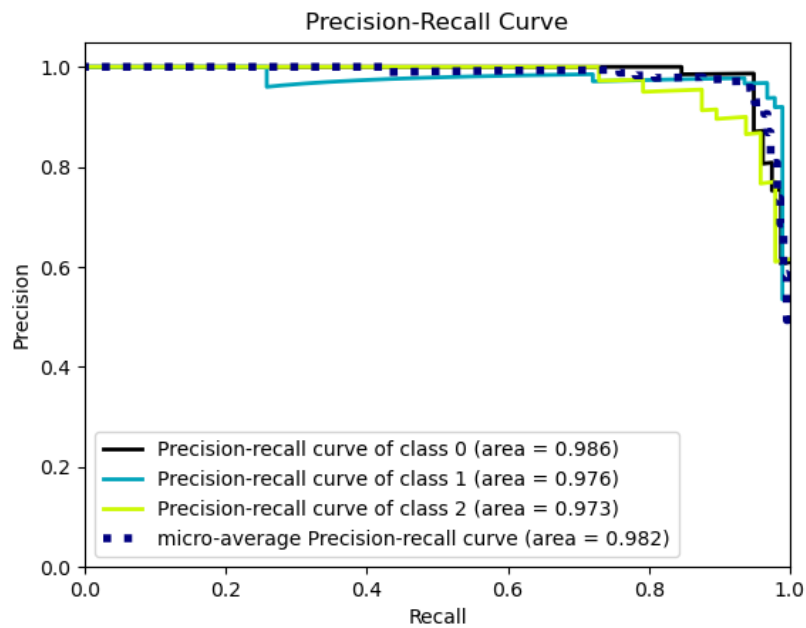
Εικόνα 4.9: Καμπύλη Precision-Recall για το μοντέλο VGG16



Εικόνα 4.10: Καμπύλη Precision-Recall για το μοντέλο ResNet50



Εικόνα 4.11: Καμπύλη Precision-Recall για το μοντέλο Xception



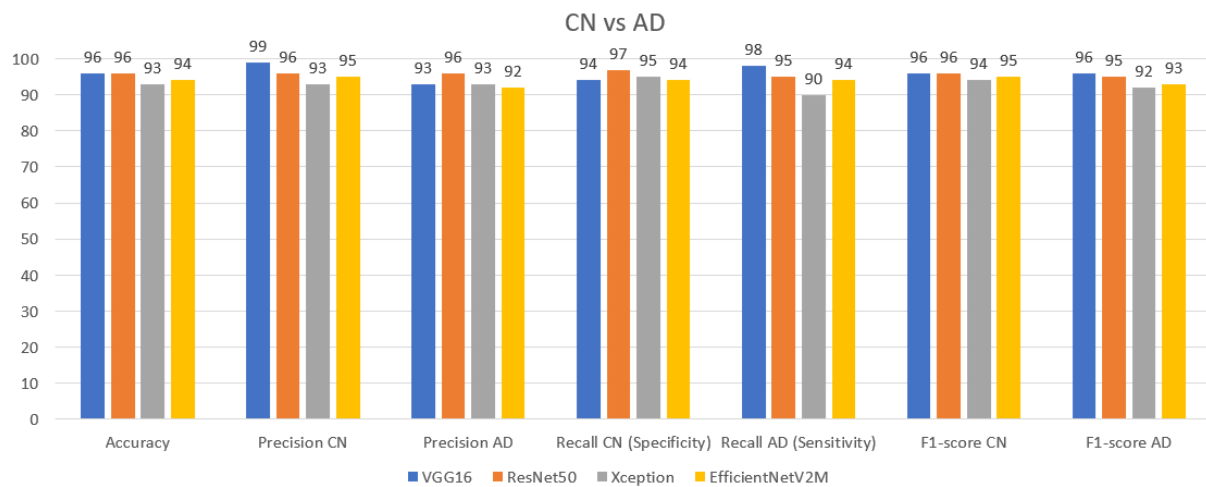
Εικόνα 4.12: Καμπύλη Precision-Recall για το μοντέλο EfficientNetV2M

Παρατηρούμε ότι το μοντέλο EfficientNetV2M έχει την υψηλότερη τιμή περιοχής κάτω από την καμπύλη ακρίβειας-ανάκλησης για τις κλάσεις CN και AD με τιμές 0.986 και 0.973 αντιστοίχως. Από την άλλη το μοντέλο ResNet50 παρουσιάζει την υψηλότερη τιμή περιοχής για την κλάση MCI που ισούται με 0.985. Τέλος το EfficientNetV2M παρουσιάζει επίσης την υψηλότερη τιμή περιοχής micro-average, ίση με 0.982.

4.2 Διαχωρισμός σε επίπεδο υποκειμένου (Subject level split)

4.2.1 Υγιείς έναντι νοσούντων (CN vs AD)

Μετρικές



Εικόνα 4.13: Συγκριτική αξιολόγηση μετρικών για CN vs AD

Αρχικά αυτό που παρατηρούμε είναι ότι οι τιμές για όλες τις μετρικές απόδοσης των μοντέλων μας είναι μεγαλύτερες από 90%. Αυτό υποδεικνύει ότι τα μοντέλα μας αποδίδουν καλά στο πρόβλημα της ταξινόμησης και είναι σε θέση να διαχωρίσουν τους υγιείς (CN) από αυτούς που έχουν νόσο του Alzheimer (AD).

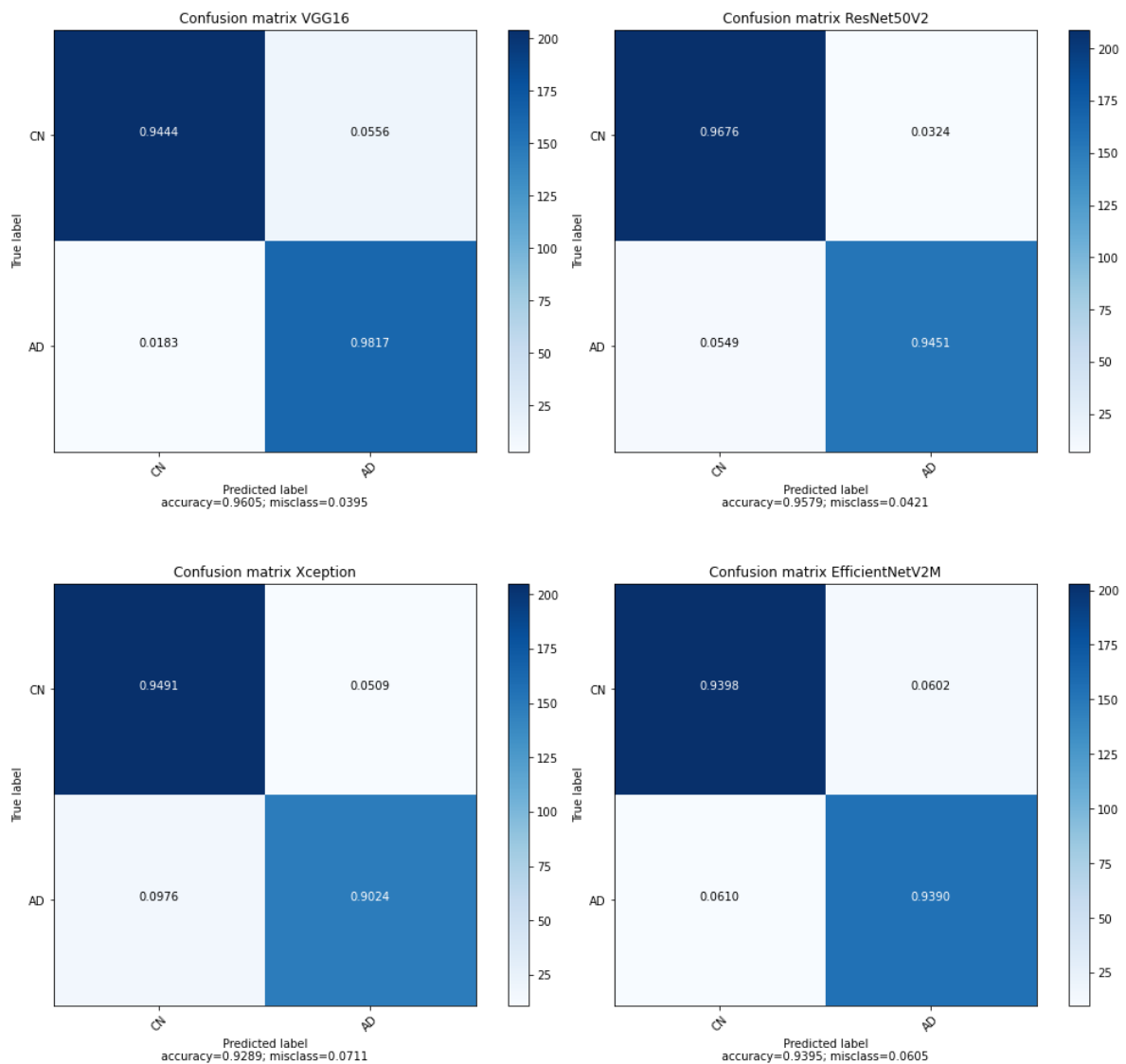
Όσον αφορά το accuracy παρατηρούμε ότι τα καλύτερα αποτελέσματα προκύπτουν από τα μοντέλα VGG16 και ResNet50 με τιμή 96% και για τα δύο. Όσον αφορά το precision παρατηρούμε ότι το μοντέλο VGG16 πετυχαίνει την υψηλότερη τιμή για την κλάση CN, ενώ το ResNet50 για την κλάση AD, με τιμές 99% και 96% αντιστοίχως.

Στην δυαδική ταξινόμηση η ανάκληση (recall) της θετικής κλάσης ονομάζεται και ευαισθησία (sensitivity), ενώ η ανάκληση της αρνητικής κλάσης ονομάζεται ειδικότητα (specificity). Παρατηρούμε ότι το μοντέλο VGG16 παρουσιάζει την μεγαλύτερη ευαισθησία με τιμή 98%, ενώ το μοντέλο ResNet50 παρουσιάζει την μέγιστη ειδικότητα με τιμή 97%.

Τέλος όσον αφορά το F1-score η μέγιστη τιμή για την κλάση CN είναι 96% και προκύπτει από το VGG16 και ResNet50, ενώ για την κλάση AD είναι πάλι 96% αλλά προκύπτει μόνο από το VGG16.

Συμπεραίνουμε ότι τα καλύτερα μοντέλα για την δυαδική ταξινόμηση CN και AD εικόνων είναι το VGG16 και το ResNet. Το VGG16 παρουσιάζει μεγαλύτερη ευαισθησία, δηλαδή είναι πιο αποτελεσματικό στην αναγνώριση των ασθενών με νόσο του Alzheimer, ενώ το ResNet50 παρουσιάζει μεγαλύτερη ειδικότητα, δηλαδή είναι πιο αποτελεσματικό στην αναγνώριση των ατόμων που είναι υγιείς.

Πίνακες Σύγκρισης



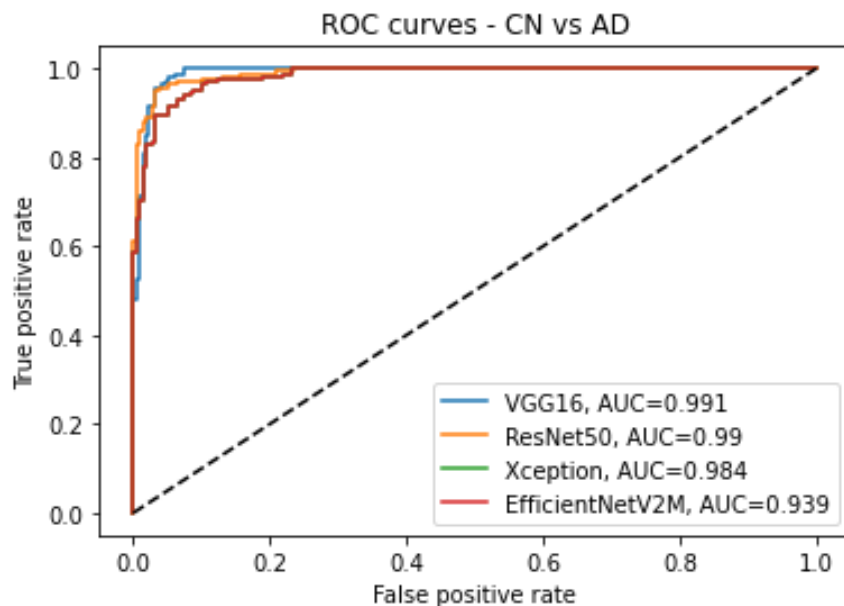
Εικόνα 4.14: Πίνακες σύγκρισης για CN vs AD

Όπως είδαμε και προηγουμένως η μέγιστη ευαισθησία προκύπτει από το μοντέλο VGG16 ενώ η μέγιστη ειδικότητα προκύπτει από το μοντέλο ResNet50.

Πιο συγκεκριμένα παρατηρούμε ότι το μοντέλο VGG16 παρουσιάζει το μέγιστο ποσοστό ορθώς θετικών (true positive) διαγνώσεων με τιμή 98.17%, ενώ το μοντέλο ResNet50 πετυχαίνει το μέγιστο ποσοστό ορθώς αρνητικών (true negatives) ταξινομήσεων με τιμή 96.76%.

Συνεπώς η πιθανότητα ταξινόμησης ενός ασθενή με AD ως CN, χρησιμοποιώντας το μοντέλο VGG16, είναι μικρότερη του 2%. Αντίστοιχα η πιθανότητα ενός υγιούς ατόμου να αναγνωριστεί ως ασθενής με νόσο του Alzheimer, χρησιμοποιώντας το μοντέλο ResNet50, είναι ίση με περίπου 3%.

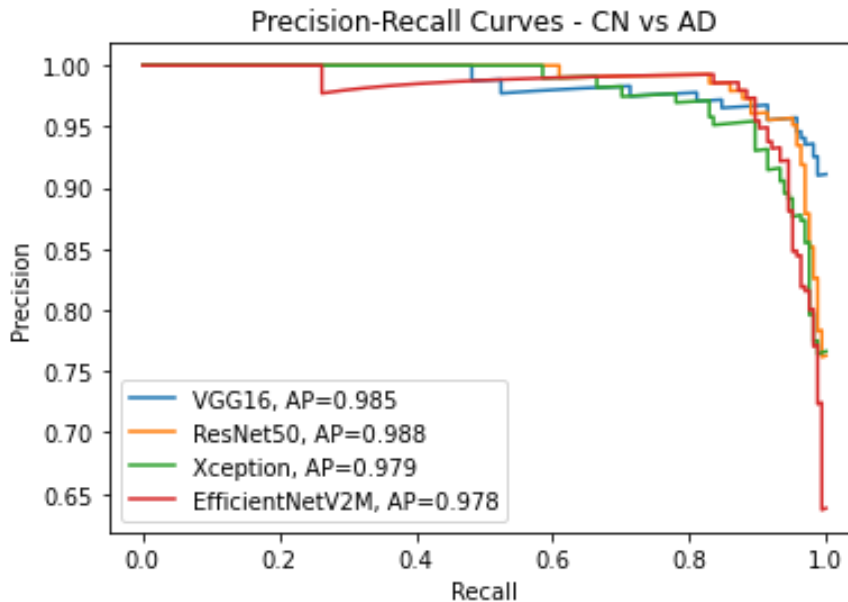
ROC Curve



Εικόνα 4.15: Συγκριτική αξιολόγηση καμπυλών ROC για CN vs AD

Παρατηρούμε ότι η μέγιστη τιμή περιοχής κάτω από την καμπύλη ROC είναι 0.991 και προκύπτει για το μοντέλο VGG16. Το μοντέλο ResNet50 πετυχαίνει την αμέσως μικρότερη τιμή AUC με διαφορά 0.001. Το EfficientNetV2M φαίνεται να πετυχαίνει τα χειρότερα αποτελέσματα με τιμή 0.939.

Precision-Recall Curves

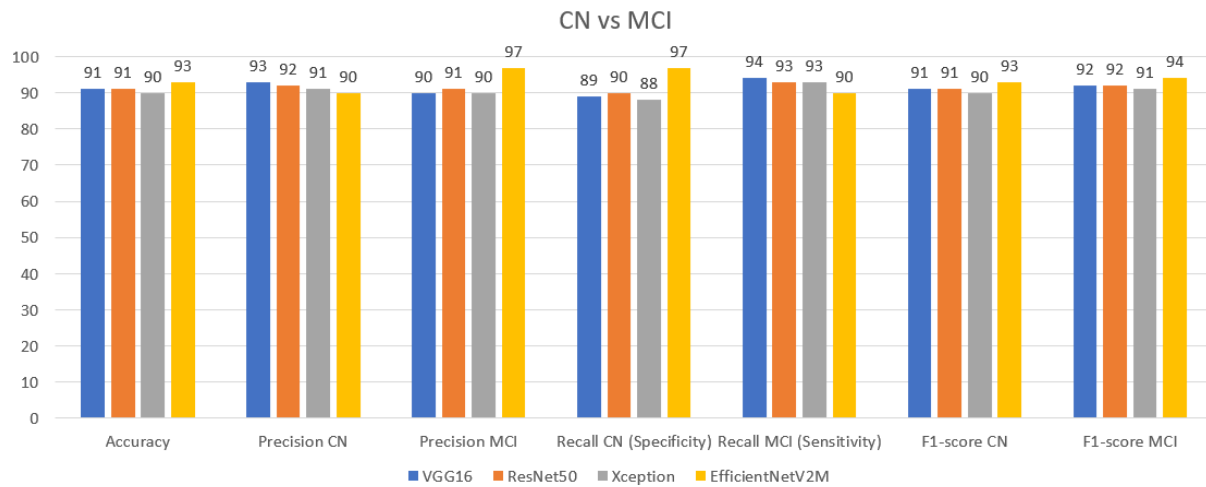


Εικόνα 4.16: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για CN vs AD

Αρχικά παρατηρούμε ότι όλες οι τιμές περιοχής κάτω από την καμπύλη ακρίβειας-ανάκλησης είναι αρκετά κοντά αφού η απόκλιση της μέγιστης από την ελάχιστη τιμή είναι ίση με 0.01. Πιο συγκεκριμένα το μοντέλο ResNet50 παρουσιάζει την υψηλότερη τιμή περιοχής που ισούται με 0.988. Το μοντέλο VGG16 παρουσιάζει την αμέσως μικρότερη τιμή περιοχής με διαφορά 0.003. Τέλος το EfficientNetV2M παρουσιάζει την μικρότερη τιμή περιοχής ίση με 0.978 η οποία είναι μικρότερη από αυτή του Xception κατά 0.001.

4.2.2 Υγιείς έναντι νοσούντων με ήπια γνωστική διαταραχή (CN vs MCI)

Μετρικές



Εικόνα 4.17: Συγκριτική αξιολόγηση μετρικών για CN vs MCI

Αρχικά αυτό που παρατηρούμε είναι ότι οι τιμές για όλες τις μετρικές απόδοσης των μοντέλων μας είναι μεγαλύτερες από 88%, ελαφρώς μειωμένες σε σχέση με το πρόβλημα CN vs AD. Αυτό πιθανώς οφείλεται στην μεγαλύτερη δυσκολία αυτού του προβλήματος σε σχέση με το προηγούμενο, καθώς το MCI είναι μια ήπια μορφή της νόσου η οποία μπορεί να είναι δύσκολα αντιληπτή από τις ιατρικές εικόνες. Παρόλα αυτά τα μοντέλα μας φαίνεται να αποδίδουν καλά και είναι σε θέση να διαχωρίσουν τους υγιείς (CN) από αυτούς που έχουν ήπια γνωστική διαταραχή (MCI).

Όσον αφορά το accuracy παρατηρούμε ότι τα καλύτερα αποτελέσματα προκύπτουν από το μοντέλο EfficientNetV2M με τιμή 93%.

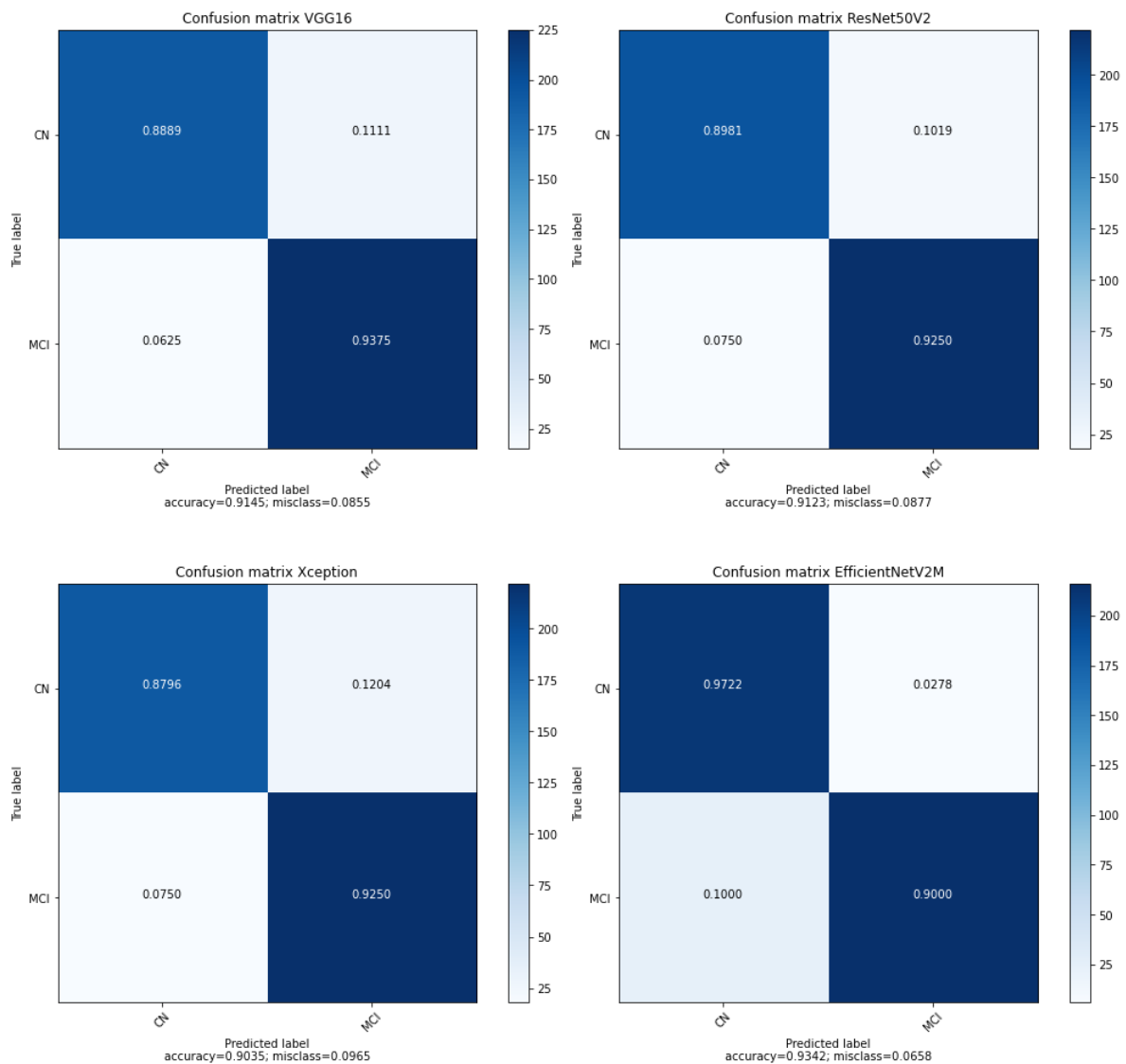
Όσον αφορά το precision παρατηρούμε ότι το μοντέλο VGG16 πετυχαίνει την υψηλότερη τιμή για την κλάση CN, ενώ το EfficientNetV2M για την κλάση MCI, με τιμές 93% και 97% αντιστοίχως.

Επίσης παρατηρούμε ότι το μοντέλο VGG16 παρουσιάζει την μεγαλύτερη ευαισθησία με τιμή 94%, ενώ το μοντέλο EfficientNetV2M παρουσιάζει την μέγιστη ειδικότητα με τιμή 97%.

Τέλος όσον αφορά το F1-score η μέγιστη επίδοση προκύπτει από το EfficientNetV2M με τιμές 93% και 94% για τις κλάσεις CN και MCI αντιστοίχως.

Συμπεραίνουμε ότι τα καλύτερα μοντέλα για την δυαδική ταξινόμηση CN και MCI εικόνων είναι το EfficientNetV2M και το VGG16. Το VGG16 παρουσιάζει μεγαλύτερη ευαισθησία, δηλαδή είναι πιο αποτελεσματικό στην αναγνώριση των ασθενών με νόσο του Alzheimer, ενώ το EfficientNetV2M παρουσιάζει μεγαλύτερη ειδικότητα, δηλαδή είναι πιο αποτελεσματικό στην αναγνώριση των ατόμων που είναι υγιείς.

Πίνακες Σύγκρισης



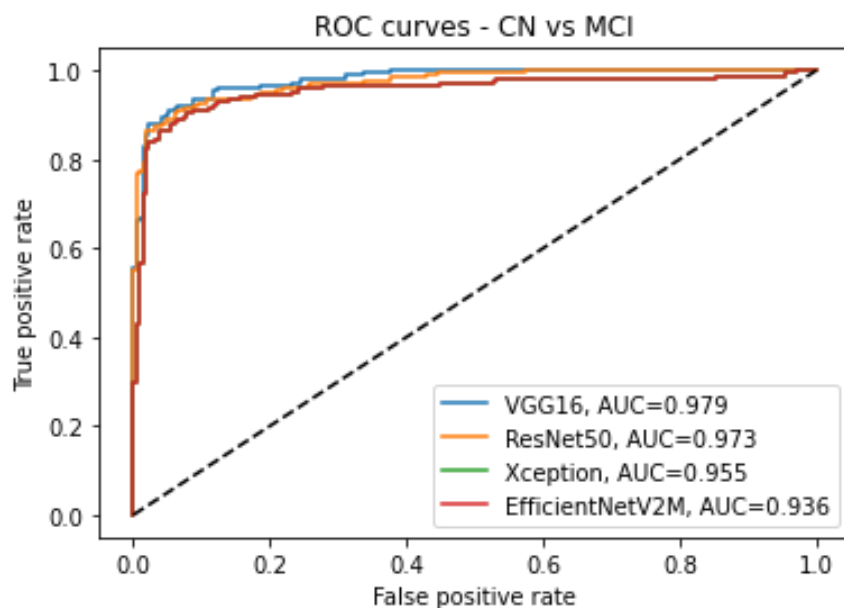
Εικόνα 4.18: Πίνακες σύγκρισης για CN vs MCI

Όπως είδαμε και προηγουμένως η μέγιστη ευαισθησία προκύπτει από το μοντέλο VGG16 ενώ η μέγιστη ειδικότητα προκύπτει από το μοντέλο EfficientNetV2M.

Πιο συγκεκριμένα παρατηρούμε ότι το μοντέλο VGG16 παρουσιάζει το μέγιστο ποσοστό ορθώς θετικών (true positive) διαγνώσεων με τιμή 93.75%, ενώ το μοντέλο EfficientNetV2M πετυχαίνει το μέγιστο ποσοστό ορθώς αρνητικών (true negatives) ταξινομήσεων με τιμή 97.22%.

Συνεπώς η πιθανότητα ταξινόμησης ενός ασθενή με MCI ως CN χρησιμοποιώντας το μοντέλο VGG16 είναι περίπου ίση με 6%. Αντίστοιχα η πιθανότητα ενός υγιούς ατόμου να αναγνωριστεί ως ασθενής που παρουσιάζει ήπια γνωστική διαταραχή, χρησιμοποιώντας το μοντέλο EfficientNetV2M, είναι μικρότερη από 3%.

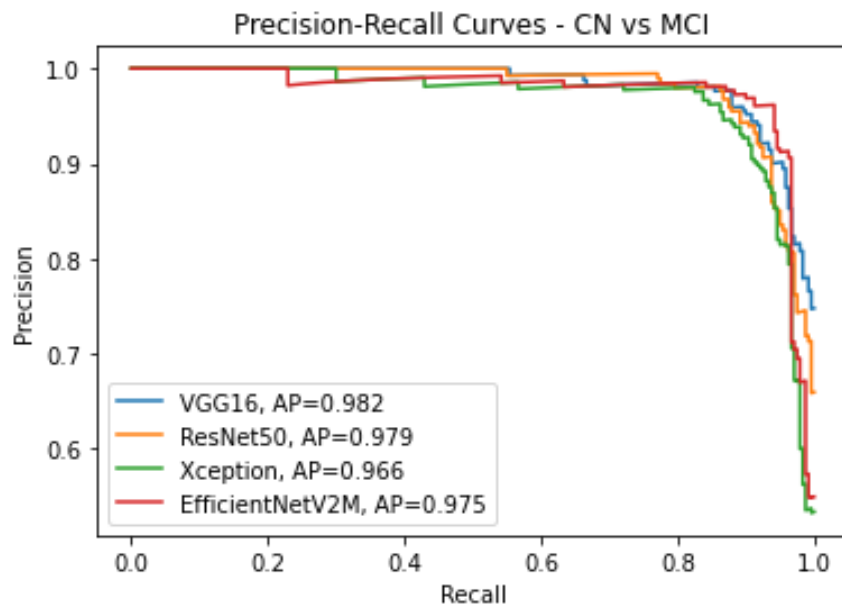
ROC Curves



Εικόνα 4.19: Συγκριτική αξιολόγηση καμπυλών ROC για CN vs MCI

Παρατηρούμε ότι η μέγιστη τιμή περιοχής κάτω από την καμπύλη ROC είναι 0.979 και προκύπτει από το μοντέλο VGG16. Το μοντέλο ResNet50 πετυχαίνει την αμέσως μικρότερη τιμή AUC με διαφορά 0.06. Το EfficientNetV2M φαίνεται να πετυχαίνει τα χειρότερα αποτελέσματα με τιμή 0.936.

Precision-Recall Curves

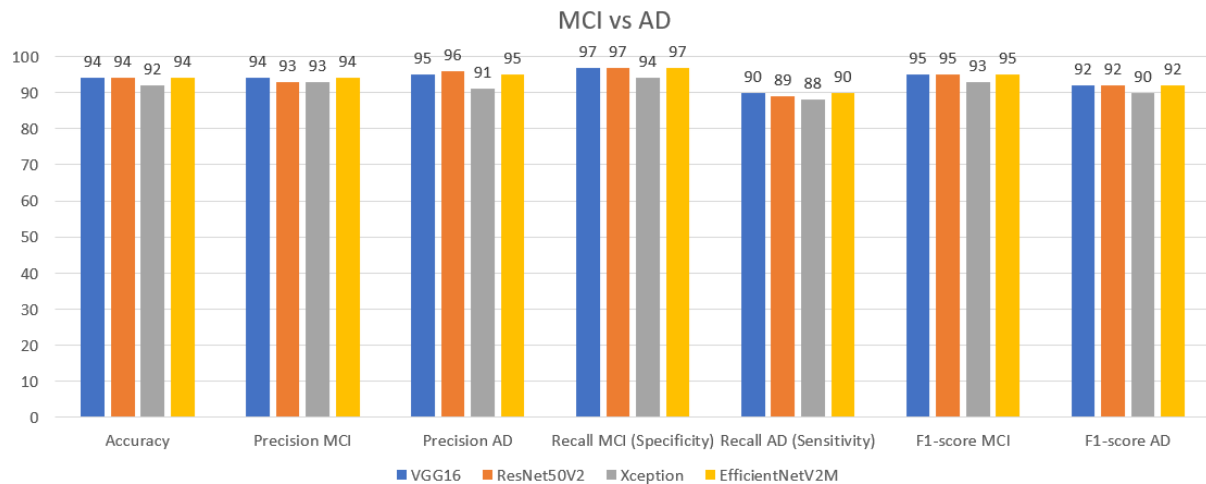


Εικόνα 4.20: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για CN vs MCI

Αρχικά παρατηρούμε ότι όλες οι τιμές περιοχής κάτω από την καμπύλη ακρίβειας-ανάκλησης είναι αρκετά κοντά αφού η απόκλιση της μέγιστης από την ελάχιστη τιμή είναι μικρότερη από 0.02. Πιο συγκεκριμένα το μοντέλο VGG16 παρουσιάζει την υψηλότερη τιμή περιοχής που ισούται με 0.982. Το μοντέλο ResNet50 πετυχαίνει την αμέσως μικρότερη τιμή περιοχής με διαφορά 0.003. Τέλος το Xception παρουσιάζει την μικρότερη τιμή περιοχής ίση με 0.966.

4.2.3 Νοσούντες με ήπια γνωστική διαταραχή έναντι νοσούντων με Alzheimer (MCI vs AD)

Μετρικές



Εικόνα 4.21: Συγκριτική αξιολόγηση μετρικών για MCI vs AD

Αρχικά αυτό που παρατηρούμε είναι ότι οι τιμές για όλες τις μετρικές απόδοσης των μοντέλων μας είναι μεγαλύτερες από 88%, δηλαδή και σε αυτή την περίπτωση είναι ελαφρώς μειωμένες σε σχέση με το πρόβλημα CN vs AD. Ο διαχωρισμός MCI vs AD παρουσιάζει πιο δυσδιάκριτες διαφορές σε σχέση με το διαχωρισμό CN και AD, όπου εκεί οι διαφορές είναι πιο εμφανείς. Παρόλα αυτά τα μοντέλα μας φαίνεται να αποδίδουν καλά και είναι σε θέση να διαχωρίσουν τα άτομα με ήπια γνωστική διαταραχή (MCI) από αυτούς που έχουν νόσο του Alzheimer (AD).

Όσον αφορά το accuracy παρατηρούμε ότι τα καλύτερα αποτελέσματα προκύπτουν από τρία μοντέλα, πιο συγκεκριμένα το VGG16, το ResNet50 και το EfficientNetV2M με τιμή 94%.

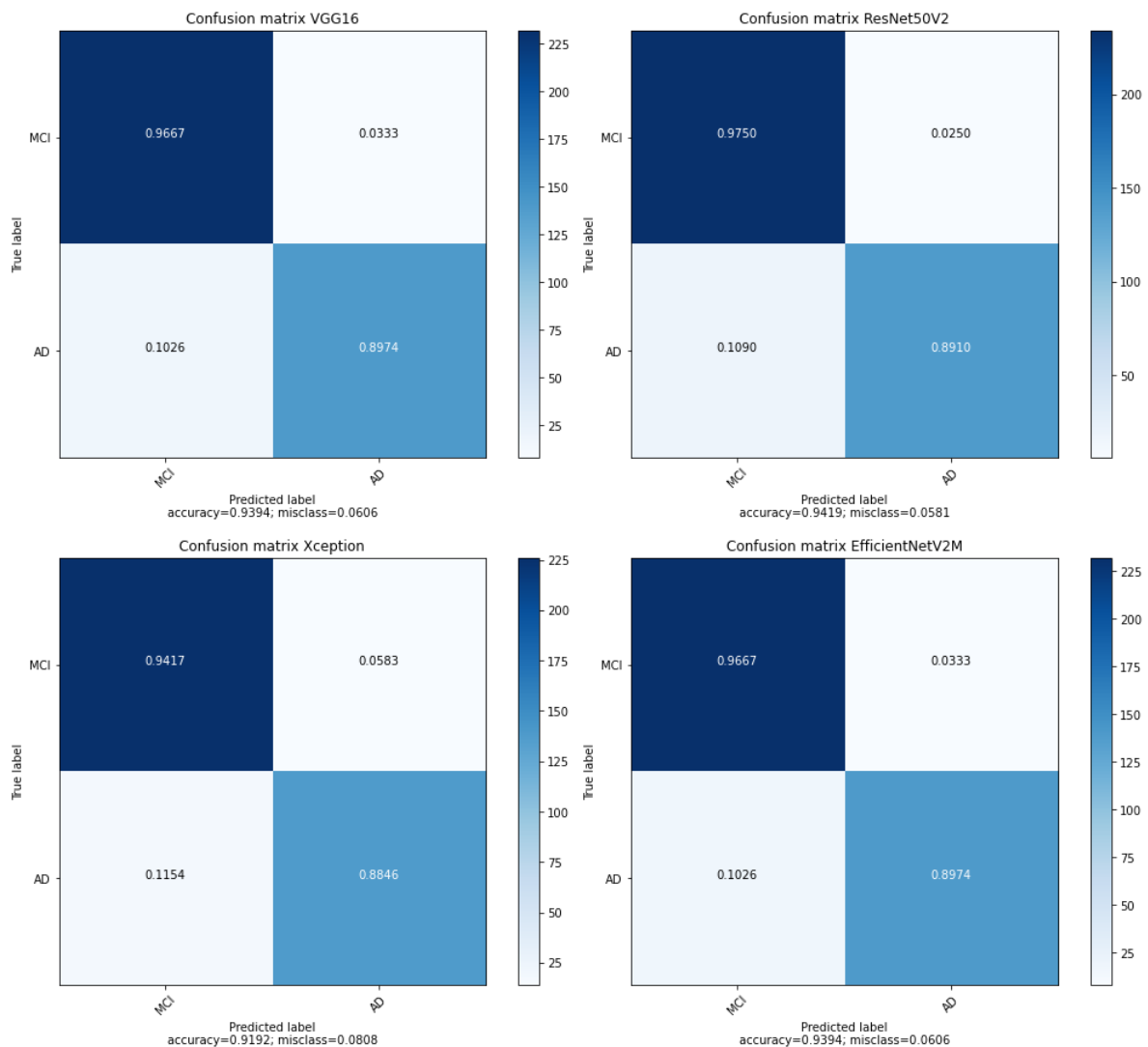
Όσον αφορά το precision παρατηρούμε ότι τα μοντέλα VGG16 και EfficientNetV2M πετυχαίνουν την υψηλότερη τιμή για την κλάση MCI, ενώ το ResNet50 για την κλάση AD, με τιμές 94% και 96% αντιστοίχως.

Επίσης παρατηρούμε ότι τα μοντέλα VGG16, ResNet50 και EfficientNetV2M παρουσιάζουν την μεγαλύτερη ειδικότητα με τιμή 97%, ενώ τα μοντέλα VGG16 και EfficientNetV2M παρουσιάζουν την μέγιστη ευαισθησία με τιμή 90%.

Τέλος όσον αφορά το F1-score η μέγιστη επίδοση για την κλάση MCI προκύπτει από τα VGG16, ResNet50 και EfficientNetV2M με τιμή 95%, ενώ για την κλάση AD τα ίδια μοντέλα πετυχαίνουν 92%. Συμπεραίνουμε ότι τα καλύτερα μοντέλα για την δυαδική ταξινόμηση MCI και AD εικόνων είναι το VGG16, το ResNet50 και το EfficientNetV2M, καθώς και τα τρία μοντέλα πετυχαίνουν εξίσου καλές τιμές σχεδόν σε όλες τις μετρικές αξιολόγησης. Το ResNet50 πετυχαίνει μεγαλύτερη τιμή precision για την κλάση AD, αλλά έχει μικρότερη ευαισθησία από το VGG16 και το EfficientNet. Παρόλα αυτά οι αποκλίσεις είναι της τάξης του 1%, συνεπώς μπορούν να θεωρηθούν αμελητέες.

Τέλος αξίζει να αναφερθεί και ο παράγοντας της πολυπλοκότητας των υπολογισμών καθώς το EfficientNet για μια εποχή εκπαίδευσης απαιτεί περίπου 70 δευτερόλεπτα, το VGG16 περίπου 22 δευτερόλεπτα, ενώ το Resnet50 περίπου 19 δευτερόλεπτα. Συνεπώς, το VGG16 και το ResNet πετυχαίνουν εξίσου καλές επιδόσεις απαιτώντας το 1/3 του χρόνου που χρειάζεται το EfficientNetV2M για εκπαίδευση.

Πίνακες σύγχυσης



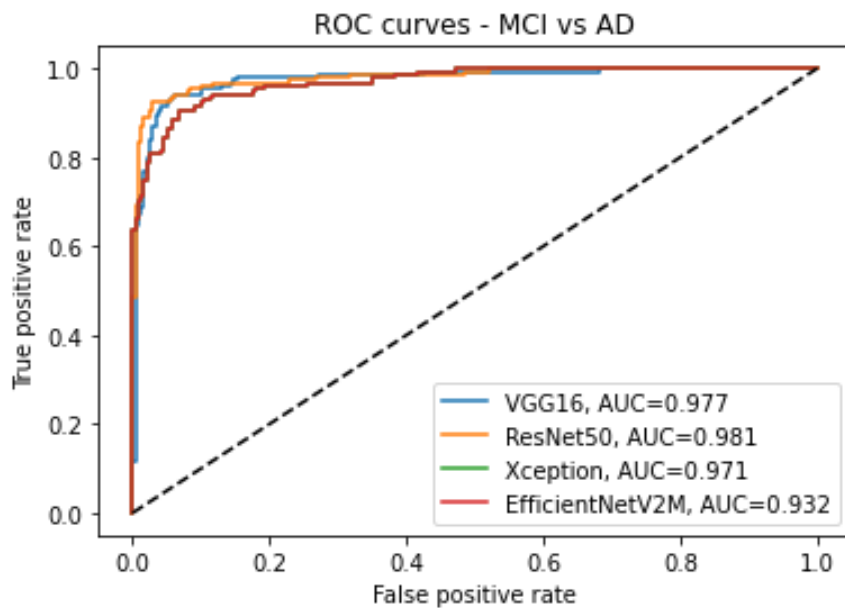
Εικόνα 4.22 Πίνακες σύγχυσης για MCI vs AD

Όπως είδαμε και προηγουμένως η μέγιστη ευαισθησία προκύπτει από τα μοντέλα VGG16 και EfficientNetV2M ενώ η μέγιστη ειδικότητα προκύπτει από το μοντέλο ResNet50.

Πιο συγκεκριμένα παρατηρούμε ότι τα μοντέλα VGG16 και EfficientNetV2M παρουσιάζουν το μέγιστο ποσοστό ορθώς θετικών (true positive) διαγνώσεων με τιμή 89,74% ενώ το μοντέλο ResNet50 πετυχαίνει το μέγιστο ποσοστό ορθώς αρνητικών (true negatives) ταξινομήσεων με τιμή 97.5%.

Συνεπώς η πιθανότητα ταξινόμησης ενός ασθενή με νόσο του Alzheimer ως MCI χρησιμοποιώντας τα μοντέλα VGG16 και EfficientNetV2M είναι περίπου ίση με 10%. Αντίστοιχα η πιθανότητα ενός ατόμου με ήπια γνωστική διαταραχή να αναγνωριστεί ασθενής με νόσο του Alzheimer, χρησιμοποιώντας το μοντέλο ResNet50, είναι ίση με 2.5%.

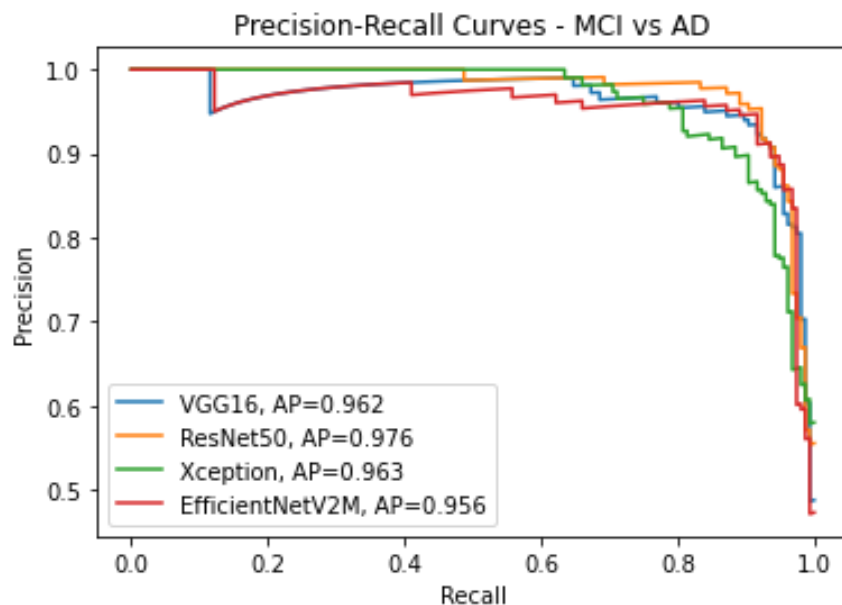
ROC Curves



Εικόνα 4.23: Συγκριτική αξιολόγηση καμπυλών ROC για MCI vs AD

Παρατηρούμε ότι η μέγιστη τιμή περιοχής κάτω από την καμπύλη ROC είναι 0.981 και προκύπτει για το μοντέλο ResNet50. Το μοντέλο VGG16 πετυχαίνει την αμέσως μικρότερη τιμή AUC με διαφορά 0.04. Το EfficientNetV2M φαίνεται να πετυχαίνει τα χειρότερα αποτελέσματα με τιμή 0.932.

Precision-Recall Curves



Εικόνα 4.24: Συγκριτική αξιολόγηση καμπυλών Precision-Recall για MCI vs AD

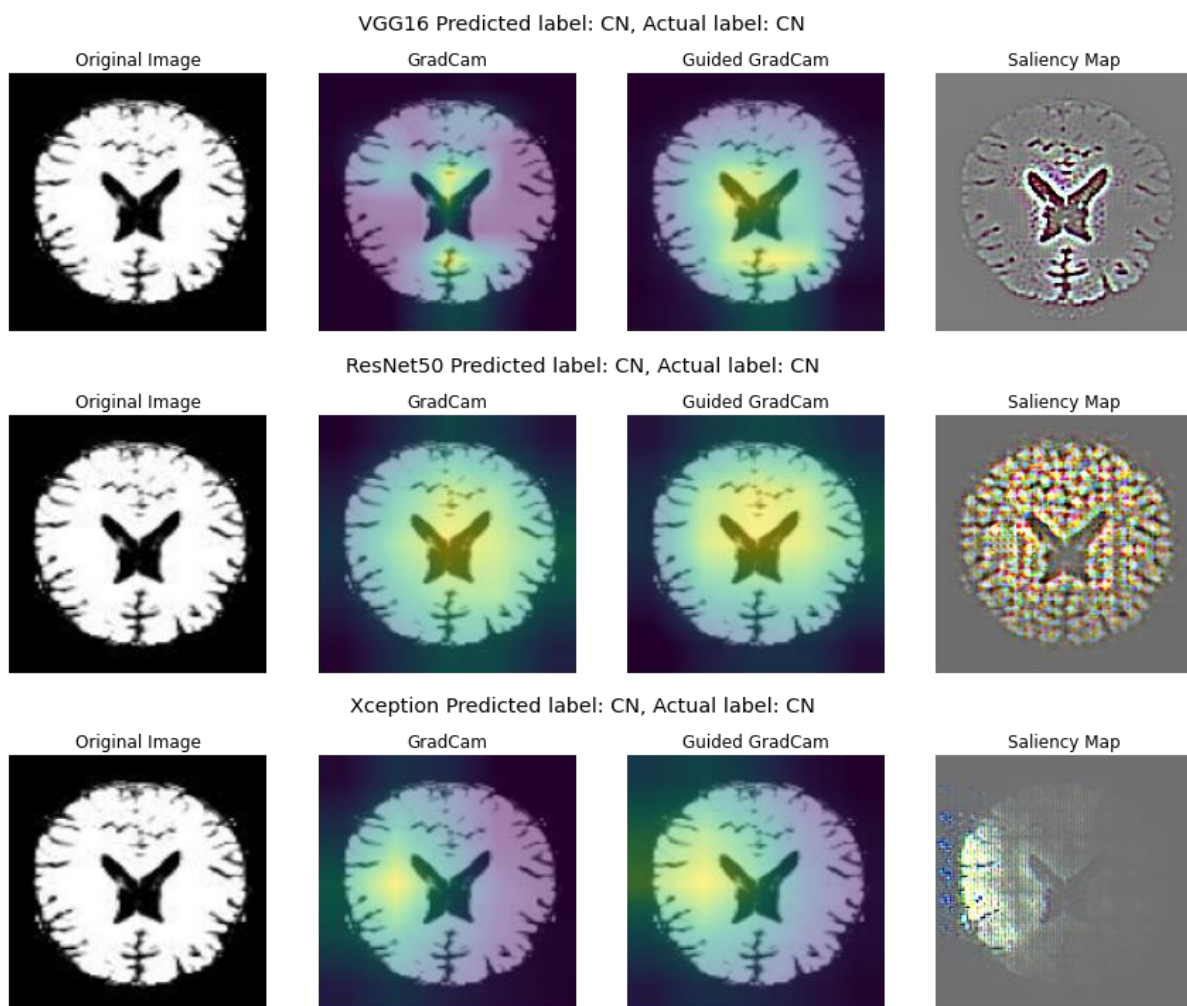
Αρχικά παρατηρούμε ότι όλες οι τιμές περιοχής κάτω από την καμπύλη ακρίβειας-ανάκλησης είναι αρκετά κοντά αφού η απόκλιση της μέγιστης από την ελάχιστη τιμή είναι ίση με 0.02. Το μοντέλο ResNet50 παρουσιάζει την υψηλότερη τιμή περιοχής που ισούται με 0.976. Το μοντέλο Xception πετυχαίνει την αμέσως μικρότερη τιμή περιοχής με διαφορά 0.963, ενώ το VGG16 ακολουθεί με 0.962. Τέλος το EfficientNetV2M παρουσιάζει την μικρότερη τιμή περιοχής ίση με 0.956.

5. Αποτελέσματα Επεξηγησιμότητας (Slice Level Split)

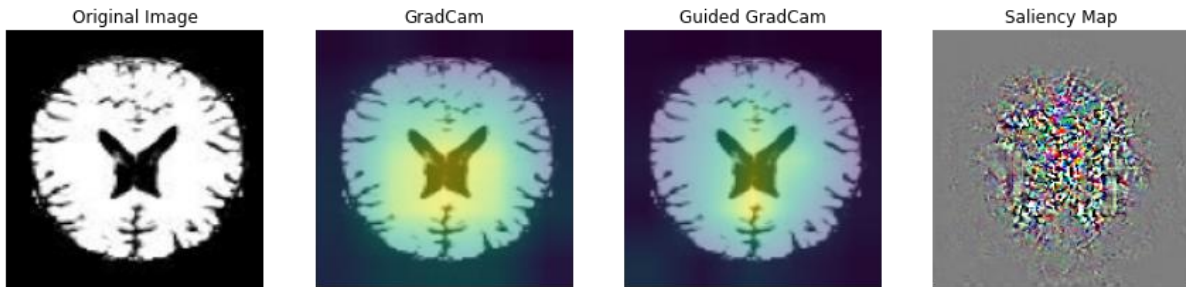
Σε αυτή την ενότητα παρουσιάζονται οι πιο σημαντικές περιοχές για την λήψη αποφάσεων κατά την διαδικασία της ταξινόμησης. Πιο συγκεκριμένα, θα δούμε τα αποτελέσματα που προκύπτουν χρησιμοποιώντας το Grad-CAM, το Guided Grad-CAM και τον Χάρτη Εστίασης (Saliency Map). Το Guided Grad-CAM αποτελεί μια βελτίωση του Grad-Cam στην οποία συνδυάζεται η οπισθοδιάδοση (backpropagation), η οποία μηδενίζει στοιχεία στις κλίσεις (gradients) που δρουν αρνητικά στην απόφαση. Τέλος, ο χάρτης εστίασης χρησιμοποιεί την προσέγγιση του Guided Grad-CAM ώστε να δημιουργήσει έναν χάρτη υψηλής ανάλυσης, ο οποίος έχει την ίδια ανάλυση με την αρχική εικόνα και εστιάζει σε λεπτομέρειες οι οποίες οδήγησαν στο αποτέλεσμα της πρόβλεψης. Οι υπολογισμοί και οι απεικονίσεις έγιναν χρησιμοποιώντας την υλοποίηση του I. Udin.⁹¹

5.1. Κλάση CN

5.1.1. Σωστές ταξινομήσεις



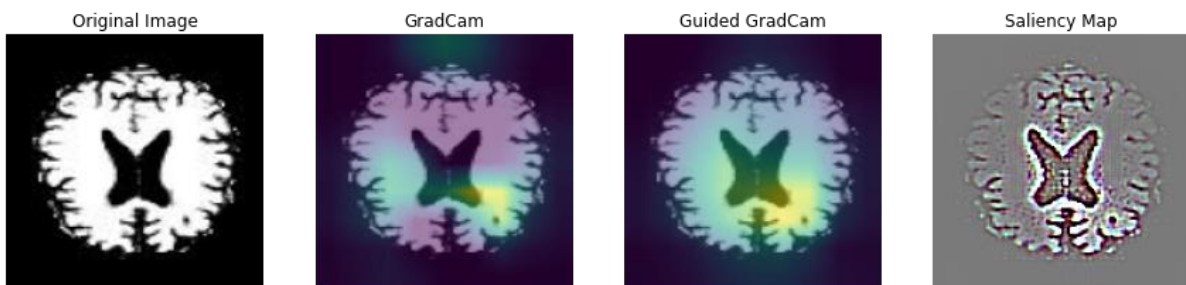
EfficientNetV2M Predicted label: CN, Actual label: CN



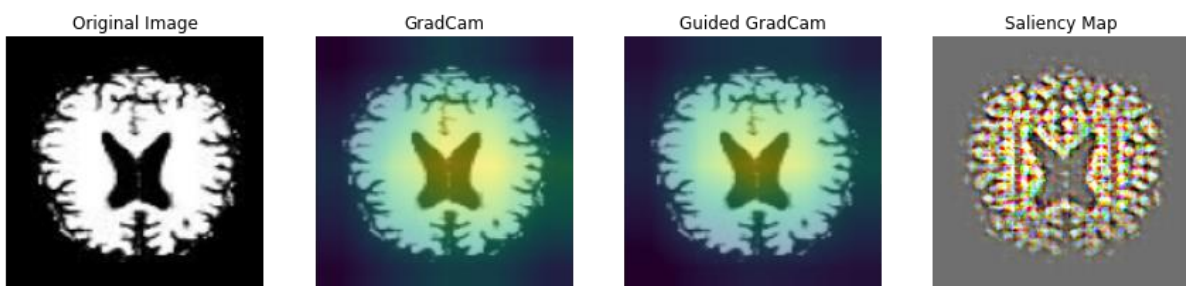
Παρατηρούμε ότι η περιοχή με το μεγαλύτερο ενδιαφέρον φαίνεται να βρίσκεται στο κέντρο της εικόνας, γύρω από την περιοχή των κοιλιών του εγκεφάλου. Πιο συγκεκριμένα βλέπουμε ότι η προσοχή των μοντέλων VGG16 και EfficientNetV2M βρίσκεται στις κοιλίες του εγκεφάλου και στις περιοχές πάνω και κάτω από αυτές. Το Resnet50 φαίνεται να ενεργοποιείται περισσότερο στις περιοχές πάνω από τις κοιλίες, ενώ το Xception ενεργοποιείται περισσότερο στην περιοχή αριστερά από την αριστερή κοιλία. Στο saliency map του VGG16 μπορούμε να δούμε πιο ξεκάθαρα την ενεργοποίηση στις κοιλίες, καθώς και στις περιοχές ακριβώς πάνω και κάτω από αυτές, όπου απεικονίζονται και πιο έντονα. Από την άλλη στο saliency map του Xception φαίνεται ακόμα πιο ξεκάθαρα η προσοχή του μοντέλου στο αριστερό ημισφαίριο του εγκεφαλικού φλοιού. Τέλος παρατηρούμε ότι στα saliency maps των ResNet50 και EfficientNetV2M η αναπαράσταση είναι λιγότερο ξεκάθαρη, παρόλα αυτά υπάρχει έντονη ενεργοποίηση στην κεντρική περιοχή του εγκεφάλου και όχι σε κάποιο άλλο μέρος της εικόνας. Γνωρίζοντας την παθολογία της νόσου του Alzheimer, μπορούμε να πούμε ότι οι περιοχές που φαίνεται να ήταν σημαντικές για τα μοντέλα ταυτίζονται με περιοχές που παρουσιάζουν αλλοιώσεις όταν κάποιος νοσεί (διευρυμένες κοιλίες, συρρίκνωση του φλοιού).

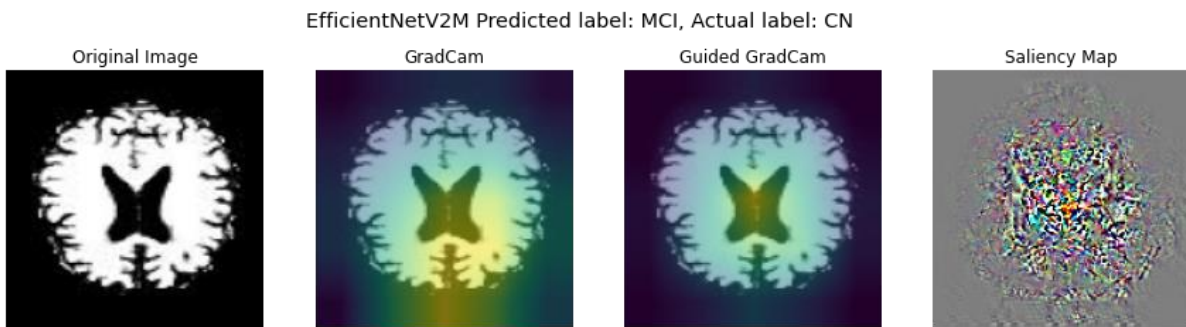
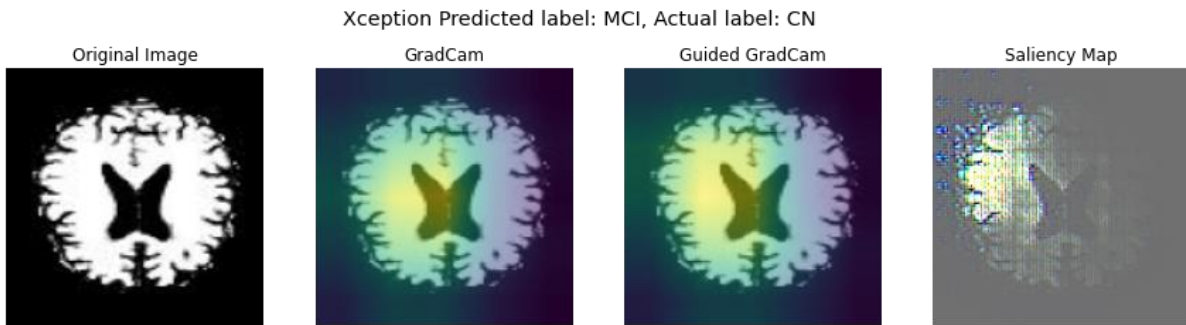
5.1.2. Λανθασμένες ταξινομήσεις

VGG16 Predicted label: MCI, Actual label: CN



ResNet50 Predicted label: MCI, Actual label: CN

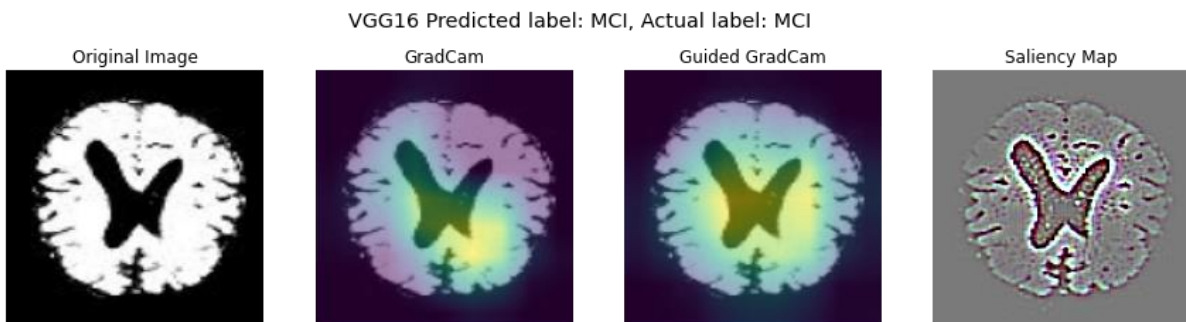


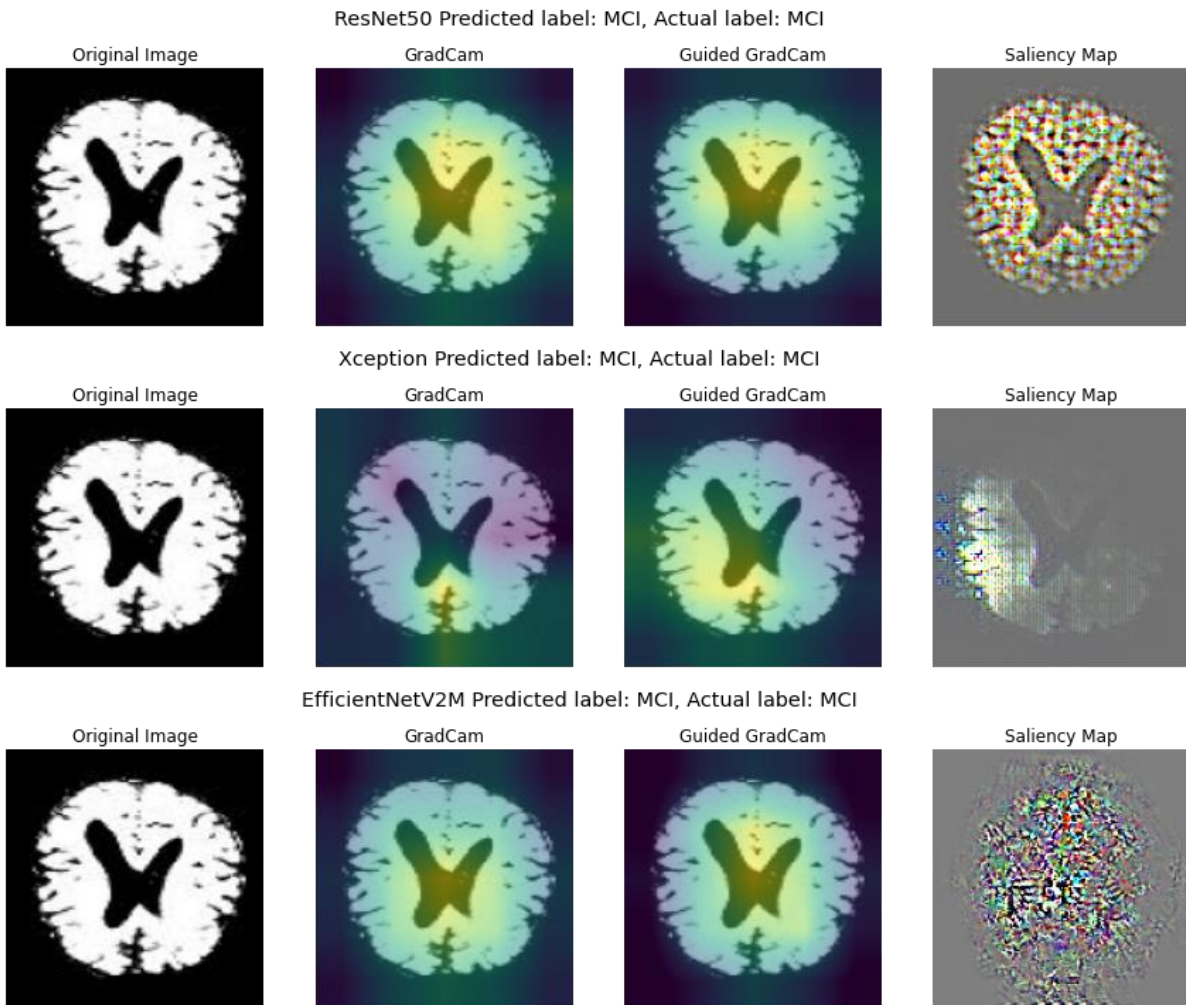


Εδώ παρουσιάζεται ένα παράδειγμα από τα δεδομένα ελέγχου, το οποίο κανένα από τα μοντέλα δεν κατάφερε να αναγνωρίσει και ταξινομήσει σωστά. Παρατηρούμε ότι και σε αυτή την περίπτωση οι περιοχές προσοχής των δικτύων, βρίσκονται γύρω από τις κοιλίες του εγκεφάλου. Πιο συγκεκριμένα, παρατηρούμε ότι το VGG16 φαίνεται να εστιάζει στο κάτω μέρος της δεξιάς κοιλίας, το Xception στην περιοχή που βρίσκεται πάνω αριστερά από την αριστερή κοιλία, ενώ το ResNet και το EfficientNet φαίνεται να εστιάζουν κυρίως εντός της περιοχής των κοιλιών, με το Resnet να δείχνει μεγαλύτερη προσοχή προς τα πάνω. Η επιθυμητή έξοδος αυτού του παραδείγματος είναι CN, ενώ αυτή που πρόβλεψαν και τα 4 μοντέλα είναι MCI. Παρατηρώντας την εικόνα μπορούμε να πούμε ότι οι κοιλίες φαίνονται πιο διευρυμένες και ο φλοιός του εγκεφάλου πιο συρρικνωμένος σε σχέση με το παράδειγμα που είδαμε προηγουμένως και ταξινομήθηκε σωστά. Συνεπώς, ίσως πρόκειται για κάποιο έκτοπο (outlier) στα δεδομένα.

5.2. Κλάση MCI

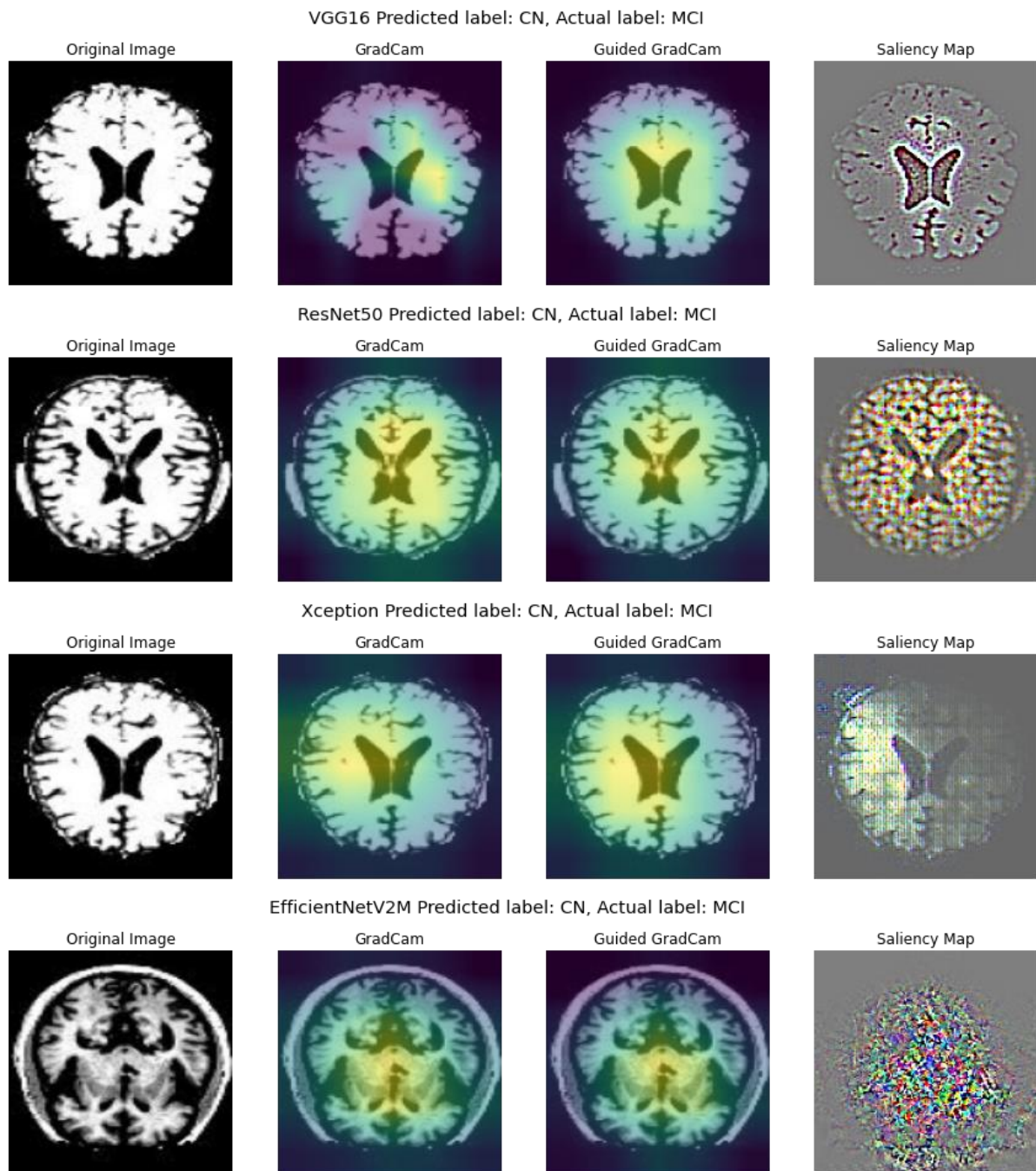
5.2.1. Σωστές Ταξινομήσεις





Παρατηρούμε ότι και σε αυτό το παράδειγμα, η περιοχή με το μεγαλύτερο ενδιαφέρον φαίνεται να βρίσκεται γύρω από την περιοχή των κοιλιών του εγκεφάλου. Το VGG16 φαίνεται να εστιάζει στις διευρυμένες κοιλίες και ιδιαίτερα στην κάτω δεξιά κοιλία, όπως βλέπουμε στην απεικόνιση του Grad-CAM. Το ResNet50 φαίνεται ιδιαίτερα ενεργό εντός των κοιλιών, καθώς και στην περιοχή του εγκεφαλικού φλοιού που βρίσκεται πάνω από αυτές. Το Xception, εστιάζει περισσότερο στο κάτω και αριστερό μέρος του φλοιού και όχι τόσο στις κοιλίες, γεγονός που γίνεται εύκολα αντιληπτό και από το saliency map. Στο saliency map του Xception, παρατηρείται έντονη ενεργοποίηση στην περιοχή του αριστερού κροταφικού λοβού η οποία παρουσιάζει συρρίκνωση προς τα έσω. Τέλος στο EfficientNet, παρατηρείται ενεργοποίηση στην ευρύτερη περιοχή των κοιλιών. Συνεπώς, συμπεραίνουμε ότι τα μοντέλα εστιάζουν την προσοχή τους στις περιοχές που εμφανίζουν αλλοιώσεις, ενισχύοντας έτσι την αξιοπιστία τους.

5.2.2 Λανθασμένες Ταξινομήσεις



Σε αυτή την περίπτωση δεν έγινε επιλογή της ίδιας εικόνας για όλα τα μοντέλα. Οι περισσότερες από τις λανθασμένα ταξινομημένες εικόνες μοιάζουν με την τελευταία εικόνα, δηλαδή απεικονίζουν διαφορετική αξονική τομή του εγκεφάλου και έχουν διαφορετικό χρωματισμό. Κατά την δημιουργία του ADNI έχουν χρησιμοποιηθεί διαφορετικοί μαγνητικοί τομογράφοι και διαφορετικές τεχνικές επεξεργασίας των εικόνων. Αυτό έχει ως αποτέλεσμα να υπάρχουν μέσα στο σύνολο δεδομένων κάποιες εικόνες οι οποίες χρειάζονται ειδική μεταχείριση και προεπεξεργασία ανάλογα με την προέλευση τους. Κάτι τέτοιο δεν ήταν εφικτό να γίνει στα πλαίσια της παρούσας εργασίας και δεδομένου ότι αυτές οι περιπτώσεις δεν ήταν τόσο συχνές, επιλέχθηκε μια ενιαία προσέγγιση στο κομμάτι της προεπεξεργασίας των δεδομένων.

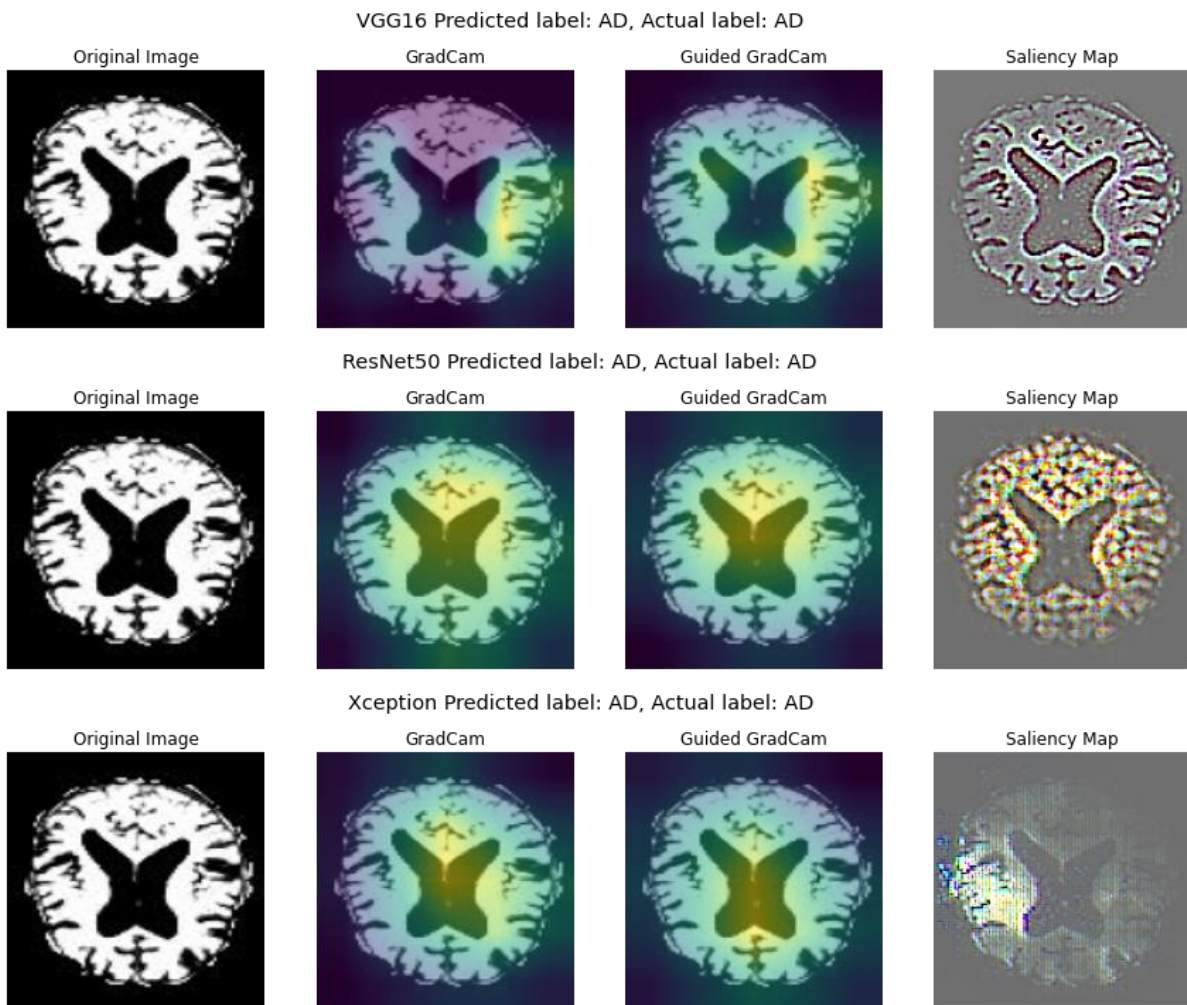
Ειδικότερα, αξίζει να αναφερθεί ότι για το μοντέλο EfficientNetV2M η εικόνα που παρουσιάζεται είναι η μοναδική που δεν ταξινομήθηκε σωστά στην κλάση MCI. Η εικόνα αυτή δεν κατάφερε να

ταξινομηθεί σωστά και από τα υπόλοιπα μοντέλα, παρόλα αυτά κρίθηκε σκόπιμο να δούμε παραδείγματα τα οποία δεν είχαν αυτή την ιδιαιτερότητα και δεν ταξινομήθηκαν σωστά.

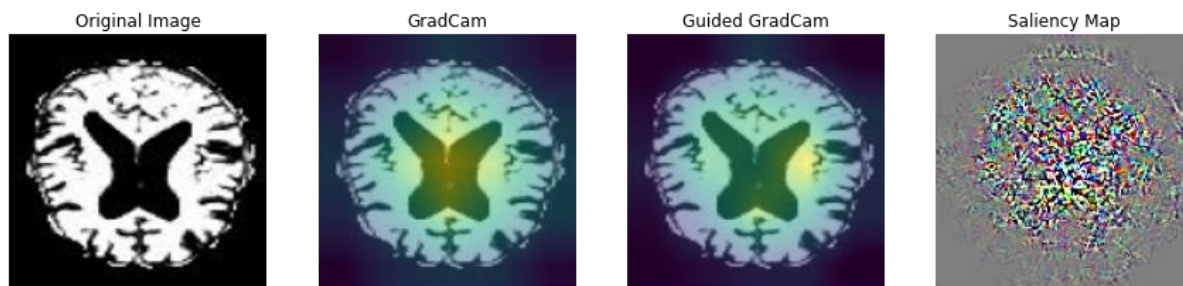
Το μοντέλο VGG16 φαίνεται να ενεργοποιείται κυρίως στην περιοχή των κοιλιών του εγκεφάλου. Επίσης βλέπουμε ότι οι κοιλίες στο συγκεκριμένο παράδειγμα φαίνονται να έχουν φυσιολογικές διαστάσεις. Έτσι μπορούμε να υποθέσουμε ότι η εστίαση της προσοχής του μοντέλου στην περιοχή των κοιλιών ήταν ο λόγος που οδήγησε στην λήψη της απόφασης να ταξινομηθεί ως CN αντί για MCI. Από την άλλη, το παράδειγμα το οποίο ταξινομήθηκε λανθασμένα από το ResNet, φαίνεται να μην έχει προεπεξεργαστεί βέλτιστα. Αυτό μπορεί να γίνει αντιληπτό, από την πληροφορία του κρανίου που έχει παραμείνει στα δεξιά και αριστερά της εικόνας. Επιπλέον, βλέπουμε ότι το ύψος της αξονικής τομής δεν μοιάζει να είναι στη μέση (συγκρίνοντας το με τα υπόλοιπα παραδείγματα). Τα στοιχεία αυτά πιθανώς δυσχεραίνουν την διαδικασία λήψης αποφάσεων για το μοντέλο. Το Xception δείχνει να εστιάζει στο αριστερό ημισφαίριο αλλά και στις κοιλίες. Παρόλα αυτά δεν καταφέρνει να ταξινομήσει σωστά το συγκεκριμένο παράδειγμα. Τέλος, το EfficientNetV2M δεν κατάφερε να ταξινομήσει σωστά το τελευταίο παράδειγμα, λόγω της ιδιαιτερότητας του, παρόλα αυτά κατάφερε και ταξινόμησε σωστά όλα τα υπόλοιπα, γεγονός που είναι εντυπωσιακό.

5.3. Κλάση AD

5.3.1. Σωστές Ταξινομήσεις



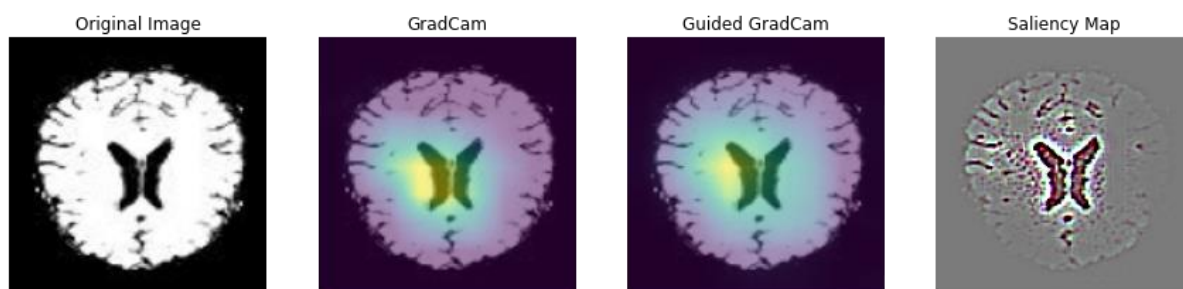
EfficientNetV2M Predicted label: AD, Actual label: AD



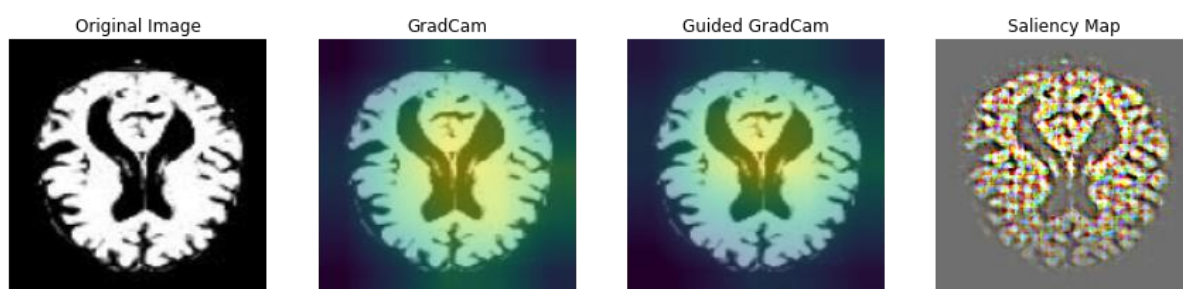
Παρατηρούμε ότι το VGG16 φαίνεται να εστιάζει εντονότερα αυτή τη φορά σε σημεία ατροφίας του εγκεφαλικού φλοιού. Πιο συγκεκριμένα παρατηρούμε έντονη ενεργοποίηση στην περιοχές του δεξιού ημισφαιρίου που παρουσιάζουν στοιχεία διάβρωσης. Τα υπόλοιπα μοντέλα φαίνεται να εστιάζουν στην ευρύτερη περιοχή των κοιλιών του εγκεφάλου, με επιμέρους ενεργοποιήσεις σε περιοχές που παρουσιάζουν στοιχεία διάβρωσης. Στο saliency map του Xception διακρίνουμε ενεργοποίηση στον αριστερό κροταφικό λοβό, αλλά και σε άλλες περιοχές που εμφανίζουν ατροφία π.χ. κάτω δεξιά της εικόνας.

5.3.2. Λανθασμένες Ταξινομήσεις

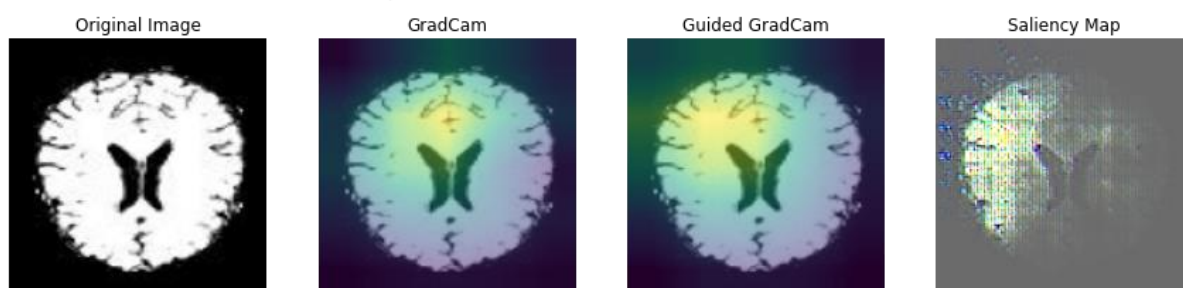
VGG16 Predicted label: CN, Actual label: AD

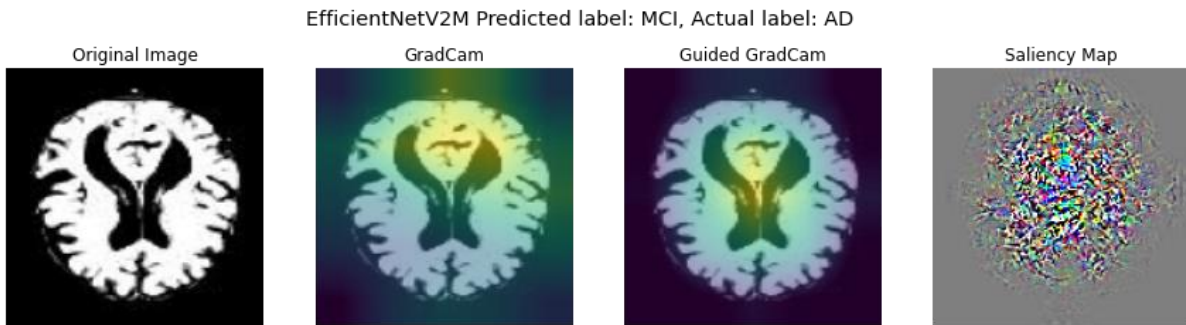


ResNet50 Predicted label: CN, Actual label: AD



Xception Predicted label: CN, Actual label: AD





Αρχικά, έγινε επιλογή παραδειγμάτων τα οποία ταξινομήθηκαν λανθασμένα ως CN και ανήκουν στην κατηγορία AD. Αυτό αποτελεί ένα σημαντικό πρόβλημα για ένα σύστημα αυτόματης διάγνωσης και μπορεί να αποβεί μοιραίο. Το μόνο μοντέλο που δεν έκανε misclassify ως CN κάποια εικόνα με ετικέτα AD είναι το EfficientNet, γεγονός που ενισχύει την αξιοπιστία του μοντέλου και το καθιστά καλύτερο για το πρόβλημα μας.

Βλέπουμε ότι το VGG16 και το Xception αντιμετώπισαν δυσκολία στην ταξινόμηση του ίδιου παραδείγματος. Το συγκεκριμένο παράδειγμα μοιάζει για outlier καθώς δεν φαίνεται να παρουσιάζει στοιχεία διάβρωσης, ούτε διεύρυνσης των κοιλιών. Η ενεργοποίηση του VGG16 ήταν έντονη στο σημείο των κοιλιών, όπως φαίνεται και στο saliency map. Από την άλλη το Xception εστίασε περισσότερο στο τμήμα του φλοιού πάνω και αριστερά από τις κοιλίες. Το ResNet50 και το EfficientNet συνάντησαν και αυτά δυσκολία στην κατηγοριοποίηση του παραδείγματος που φαίνεται στις αντίστοιχες απεικονίσεις. Η περιοχή των κοιλιών του συγκεκριμένου παραδείγματος, φαίνεται να παρουσιάζει αλλοιώσεις οι οποίες πιθανώς να προήλθαν από την προεπεξεργασία των δεδομένων και κατά συνέπεια να χρειαζόταν ειδική μεταχείριση. Η ιδιαιτερότητα του παραδείγματος σε συνδυασμό με την εστίαση των μοντέλων στην περιοχή των κοιλιών είχε ως αποτέλεσμα την λανθασμένη ταξινόμηση του.

6. Επίλογος

Σε αυτή την ενότητα θα γίνει μια σύνοψη της δουλειάς αυτής της διπλωματικής εργασίας σε συνδυασμό με τα τελικά συμπεράσματα που προέκυψαν από αυτή την μελέτη. Τέλος θα αναφερθούν μερικές προτάσεις για μελλοντικές προεκτάσεις της παρούσας εργασίας.

6.1 Σύνοψη

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε την αναγνώριση της νόσου του Alzheimer με χρήση μεθόδων βαθιάς μάθησης και ερμηνεύσιμης τεχνητής νοημοσύνης σε εικόνες μαγνητικής τομογραφίας εγκεφάλου. Πρόκειται για ένα δύσκολο και απαιτητικό πρόβλημα, καθώς αυτή η διαδικασία πραγματοποιείται από ειδικευμένο κλινικό προσωπικό, όπως γιατρούς ή ακτινολόγους.

Όπως σε κάθε πρόβλημα που γίνεται εφαρμογή μεθόδων βαθιάς μάθησης, η ποιότητα των αποτελεσμάτων εξαρτάται από την ποιότητα και την επάρκεια των δεδομένων εκπαίδευσης. Το σύνολο δεδομένων που χρησιμοποιήθηκε ήταν το ADNI, και περιείχε 2180 εικόνες μαγνητικής τομογραφίας σε T1-weighted μορφή χωρισμένες σε 3 κατηγορίες (AD, MCI, CN).

Οι ερευνητές του ADNI είχαν φροντίσει, ώστε κάθε εικόνα μαγνητικής τομογραφίας να είναι υψηλής ποιότητας και να έχει την σωστή ετικέτα ανάλογα με την κατηγορία στην οποία ανήκει. Παρόλα αυτά, υπήρχε ποικιλία ως προς τις διαστάσεις των εικόνων και των μεθόδων επεξεργασίας που είχαν προηγηθεί, λόγω των διαφορετικών μαγνητικών τομογράφων που χρησιμοποιήθηκαν για την παραγωγή των εικόνων. Συνεπώς η προεπεξεργασία των εικόνων κρίθηκε απαραίτητη, ούτως ώστε να μετασχηματιστούν σε μορφή κατάλληλη για είσοδο στα μοντέλα βαθιάς μάθησης που χρησιμοποιήθηκαν.

Το πρώτο βήμα της προεπεξεργασίας των δεδομένων ήταν ο μετασχηματισμός τους σε ένα κοινό σύστημα συντεταγμένων, εφαρμόζοντας αρχικά resampling και στη συνέχεια registration των εικόνων σε ένα κοινό atlas. Η ιστροπική ανάλυση που επιλέχθηκε για το resampling ήταν 2mm^3 , γεγονός που μείωσε τις διαστάσεις των εικόνων στο μισό και κατά συνέπεια έκανε το πρόβλημα περισσότερο υπολογιστικά διαχειρίσιμο. Για το registration χρησιμοποιήθηκε ο MNI 305 atlas ο οποίος έχει προκύψει από τον μέσο όρο 305 T1-weighted εικόνων μαγνητικής τομογραφίας μετασχηματισμένων γραμμικά στον χώρο Talairach. Έπειτα υλοποιήθηκε μια διαδικασία για την εξαγωγή της πληροφορίας του κρανίου από τις εικόνες, χρησιμοποιώντας το FSL BET. Τέλος έγινε επιλογή διαστάσεων slices από τις εικόνες, ούτως ώστε να μπορέσουν να αποτελέσουν είσοδο στα προ-εκπαιδευμένα συνελκτικά νευρωνικά δίκτυα που χρησιμοποιήθηκαν.

Σημαντικό κομμάτι ενός μοντέλου βαθιάς μάθησης αποτελεί η αμερόληπτη αξιολόγηση του με σκοπό να αξιολογηθεί η πιθανή κλινική του αξία. Για τον σκοπό αυτό, έγινε διαχωρισμός στα δεδομένα τόσο σε επίπεδο υποκειμένου, όσο και σε επίπεδο εγκάρσιων τομών. Έτσι έγινε μελέτη της επίδρασης του διαχωρισμού των δεδομένων στην απόδοση των μοντέλων.

Για την ταξινόμηση των εικόνων επιλέχθηκαν 4 μοντέλα από διαφορετικές οικογένειες δικτύων, πιο συγκεκριμένα τα: VGG16, ResNet50V2, Xception και EfficientNetV2-M. Η διαδικασία εκπαίδευσης των μοντέλων ήταν η ίδια για όλα τα πειράματα ανεξάρτητα από τον διαχωρισμό των δεδομένων ούτως ώστε η αναπαραγωγή των πειραμάτων και η σύγκριση των αποτελεσμάτων να γίνεται ευκολότερα.

Τέλος έγινε εφαρμογή μεθόδων επεξηγήσιμης τεχνητής νοημοσύνης πιο συγκεκριμένα του Grad-CAM και του Guided Grad-CAM. Έτσι παράγονται αναπαραστάσεις με την μορφή χαρτών θερμότητας οι οποίοι αναπαριστούν τις μέγιστες ενεργοποιήσεις των νευρωνικών δικτύων κατά την λήψη μιας

απόφασης. Αυτό προσδίδει μια επεξήγηση στην διαδικασία με την οποία λαμβάνουν αποφάσεις τα μοντέλα, ενισχύοντας έτσι την αξιοπιστία τους.

6.2 Τελικά Συμπεράσματα

Ο βασικός σκοπός αυτής της διπλωματικής εργασίας ήταν η εξοικείωση με τις μεθόδους βαθιάς μάθησης και επεξηγήσιμης τεχνητής νοημοσύνης, καθώς και η μελέτη των χαρακτηριστικών που οδηγούν σε ένα καλό μοντέλο. Θεωρώ ότι όλα τα παραπάνω καλύφθηκαν κατά την εκπόνηση της παρούσας εργασίας κι επιπλέον τα αποτελέσματα τα οποία επιτεύχθηκαν, κρίνονται ικανοποιητικά σε σύγκριση με άλλες εργασίες. Τα συμπεράσματα στα οποία καταλήξαμε προκύπτουν τόσο από τα πειράματα που εκτελέστηκαν όσο και από τη σύγκριση με συγγενείς εργασίες στον ίδιο τομέα εφαρμογής και παρουσιάζονται επιγραμματικά παρακάτω.

Αρχικά παρατηρήθηκε ότι η προεπεξεργασία των εικόνων παίζει καίριο και καθοριστικό ρόλο στην ποιότητα των αποτελεσμάτων, καθώς οι εικόνες που δεν είχαν προεπεξεργαστεί κατάλληλα συχνά ταξινομούσαν λανθασμένα.

Επιπλέον η εφαρμογή της μεταφοράς μάθησης έδωσε την δυνατότητα στα μοντέλα να επιτύχουν πολύ υψηλές επιδόσεις, παρά τον περιορισμένο αριθμό εκπαιδευτικών παραδειγμάτων, με ελάχιστη ρύθμιση στις παραμέτρους και χωρίς να απαιτείται μεγάλος χρόνος εκπαίδευσης.

Επίσης η εφαρμογή τυχαίου data augmentation στις εικόνες εκπαίδευσης, βοήθησε στην αποφυγή του overfitting, βελτιώνοντας σημαντικά την ικανότητα των μοντέλων να αποδίδουν εξίσου καλά σε παραδείγματα που δεν είχαν δει προηγουμένως.

Εκτός αυτού, παρατηρήθηκε ότι ο διαχωρισμός των δεδομένων είναι εξαιρετικά σημαντικός, όπως είχε γίνει ήδη σαφές από την βιβλιογραφία. Πιο συγκεκριμένα παρατηρήθηκε μεγάλη απόκλιση στις τιμές των επιδόσεων μεταξύ διαχωρισμού σε επίπεδο υποκειμένου και επίπεδο εγκάρσιας τομής όταν πραγματοποιήθηκε ταξινόμηση σε 3 κατηγορίες. Οι τιμές των αποτελεσμάτων έγιναν συγκρίσιμες μόνο όταν εφαρμόσαμε δυαδική ταξινόμηση των δεδομένων που είχαν χωριστεί σε επίπεδο υποκειμένου, με όλους τους δυνατούς συνδυασμούς. Επιπλέον, η επιλογή των slices ήταν εξίσου σημαντική καθώς η επίδοση στον subject-level διαχωρισμό δεν ήταν ικανοποιητική μέχρις ότου να βρεθεί ο σωστός συνδυασμός από slices.

Τέλος, όσον αφορά τα αποτελέσματα που προέκυψαν από τις μεθόδους επεξηγήσιμης τεχνητής νοημοσύνης, παρατηρήθηκε ότι οι περιοχές εστίασης των μοντέλων ταυτίζονται με τις περιοχές του εγκεφάλου που επηρεάζονται περισσότερο από την νόσο του Alzheimer, ενισχύοντας περαιτέρω την αξιοπιστία και την εμπιστοσύνη των μοντέλων που υλοποιήθηκαν. Από την άλλη οι περισσότερες από τις εικόνες που ταξινομήθηκαν λανθασμένα, συνήθως δεν είχαν προεπεξεργαστεί σωστά. παρουσιάζοντας αλλοιώσεις ή είχαν σημαντικές οπτικές ομοιότητες με τις εικόνες της λανθασμένης κατηγορίας.

6.3 Μελλοντικές Προεκτάσεις

Η δουλειά της παρούσας διπλωματικής, θα μπορούσε να αποτελέσει την βάση για περαιτέρω έρευνα στην αναγνώριση της νόσου του Alzheimer. Στην συνέχεια θα αναφερθούν επιγραμματικά μερικές προτάσεις για μελλοντικές προεκτάσεις.

Αρχικά, θα είχε ενδιαφέρον η χρήση διαφορετικών επιπέδων από τις μαγνητικές τομογραφίες εγκεφάλου. Στην παρούσα εργασία έγινε χρήση μόνο του εγκάρσιου επιπέδου, αλλά θα μπορούσαμε να δημιουργήσουμε διαφορετικά μοντέλα και για τα υπόλοιπα επίπεδα (στεφανιαίο και οβελιαίο). Εκτός αυτού θα μπορούσε να υλοποιηθεί και ένα γενικό μοντέλο το οποίο θα χρησιμοποιούσε την πληροφορία και από τα τρία επίπεδα. Έτσι θα μπορούσε να γίνει μια συγκριτική αξιολόγηση των αποτελεσμάτων που θα προέκυπταν από κάθε μοντέλο, καταλήγοντας στο καλύτερο επίπεδο ή συνδυασμό επιπέδων για το παρόν πρόβλημα.

Επιπλέον στην παρούσα εργασία δεν πραγματοποιήθηκε hyperparameter tuning λόγω περιορισμένων υπολογιστικών πόρων. Είναι σημαντικό να υλοποιηθεί ένας συστηματικός τρόπος βελτιστοποίησης των παραμέτρων (π.χ. Grid Search, Random Search), ούτως ώστε να καταλήξουμε σε ακόμα καλύτερα αποτελέσματα. Επίσης θα μπορούσαν να πραγματοποιηθούν παραπάνω πειράματα με σκοπό την ανίχνευση των βέλτιστων slices για το παρόν πρόβλημα αλλά και της βέλτιστης ισοτροπικής ανάλυσης των εικόνων.

Όλα τα μοντέλα που χρησιμοποιήθηκαν σε αυτή την διπλωματική ήταν προ-εκπαιδευμένα στο σύνολο δεδομένων ImageNet. Θα μπορούσε να γίνει εφαρμογή του Big Transfer (BiT)⁹², μιας state-of-the-art εκδοχής της μετάδοσης γνώσης για την ταξινόμηση εικόνας. Το BiT επανεξετάζει την διαδικασία της προεκπαίδευσης δικτύων σε μεγάλα επιβλεπόμενα σύνολα δεδομένων τα οποία στη συνέχεια γίνονται fine-tuned στο πρόβλημα που μας ενδιαφέρει. Όσο αυξάνεται ο όγκος των δεδομένων προεκπαίδευσης, τόσο αυξάνεται και η σημασία της κατάλληλης επιλογής επιπέδων κανονικοποίησης και της κλιμάκωσης της αρχιτεκτονικής του δικτύου. Υπάρχουν BiT μοντέλα προεκπαιδευμένα στο ImageNet και το ImageNet-21k⁹³ διαθέσιμα στο TFHub, με μεγάλη ποικιλία μεγεθών από ένα τυπικό ResNet50 μέχρι ένα ResNet152x4 (με 152 επίπεδα βάθος και 4 φορές πιο πλατύ από ένα τυπικό ResNet50) για χρήστες με υψηλότερες απαιτήσεις ακρίβειας που διαθέτουν μεγάλη υπολογιστική ισχύ και μνήμη.

Τέλος, όπως αναφέρθηκε και προηγουμένως, ορισμένες εικόνες δεν είχαν προεπεξεργαστεί με κατάλληλο τρόπο με αποτέλεσμα συχνά να παρουσιάζουν αλλοιώσεις, και κατά συνέπεια να μην ταξινομούνται σωστά,. Μια μελλοντική προέκταση της παρούσας εργασίας θα μπορούσε να μεριμνήσει για αυτές τις περιπτώσεις, εφαρμόζοντας εξατομικευμένη προεπεξεργασία, ανάλογα με συγκεκριμένα χαρακτηριστικά των εικόνων (π.χ. τον μαγνητικό τομογράφο προέλευσης).

Βιβλιογραφία

1. Scheltens, P. *et al.* Alzheimer's disease. *The Lancet* vol. 388 505–517 Preprint at [https://doi.org/10.1016/S0140-6736\(15\)01124-1](https://doi.org/10.1016/S0140-6736(15)01124-1) (2016).
2. Prince, M. *et al.* The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's and Dementia* vol. 9 63-75.e2 Preprint at <https://doi.org/10.1016/j.jalz.2012.11.007> (2013).
3. Τσολάκη, Μ. *Alzheimer: Νέες διαστάσεις στην προσέγγιση της νόσου.* (2007).
4. Goedert, M. & Spillantini, M. G. A century of Alzheimer's disease. *Science* vol. 314 777–781 Preprint at <https://doi.org/10.1126/science.1132814> (2006).
5. Joe, E. & Ringman, J. M. Cognitive symptoms of Alzheimer's disease: Clinical management and prevention. *The BMJ* vol. 367 Preprint at <https://doi.org/10.1136/bmj.l6217> (2019).
6. Jill Jin. Alzheimer Disease. *JAMA* **313**, (2015).
7. Thies, W. & Bleiler, L. Alzheimer's disease facts and figures. (2013) doi:10.1016/j.jalz.2013.02.003.
8. Cummings, J. L. & Cole, G. *Alzheimer Disease*. <http://jama.jamanetwork.com/>.
9. Medical Association, A. Screening for Cognitive Impairment in Older Adults. (2020) doi:10.1001/jama.2020.0435.
10. Nasreddine, Z. S. *et al.* *The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment.* www.mocatest.com.
11. Langa, K. M. & Levine, D. A. The diagnosis and management of mild cognitive impairment: A clinical review. *JAMA - Journal of the American Medical Association* vol. 312 2551–2561 Preprint at <https://doi.org/10.1001/jama.2014.13806> (2014).
12. Correia, S. & Behavior, H. Alzheimer disease: Time to improve its diagnosis and treatment. *CLEVELAND CLINIC JOURNAL OF MEDICINE* **76**, 49 (2009).
13. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* vol. 6 67–77 Preprint at <https://doi.org/10.1038/nrneurol.2009.215> (2010).
14. Vemuri, P. & Jack, C. R. Role of structural MRI in Alzheimer's disease. *Alzheimer's Research and Therapy* **2**, 1–10 (2010).
15. Matthews, B. R. *ALZHEIMER DISEASE UPDATE.* *Continuum Lifelong Learning Neurol* vol. 16 <http://journals.lww.com/continuum> (2010).
16. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* vol. 521 436–444 Preprint at <https://doi.org/10.1038/nature14539> (2015).
17. Shen, D., Wu, G. & Suk, H.-I. Deep Learning in Medical Image Analysis. (2017) doi:10.1146/annurev-bioeng-071516.
18. Nair, V. & Hinton, G. E. *Rectified Linear Units Improve Restricted Boltzmann Machines.*
19. Srivastava, N., Hinton, G., Krizhevsky, A. & Salakhutdinov, R. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* *Journal of Machine Learning Research* vol. 15 (2014).

20. Ioffe, S. & Szegedy, C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*.
21. Everingham, M. *et al.* The PASCAL Visual Object Classes (VOC) Challenge. *Int J Comput Vis* **88**, 303–338 (2010).
22. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* **115**, 211–252 (2015).
23. de Brébisson, A. & Montana, G. *Deep Neural Networks for Anatomical Brain Segmentation*.
24. Wu, G., Kim, M., Wang, Q., Munsell, B. C. & Shen, D. Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning; Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning. *IEEE Transactions on Biomedical Engineering* **63**, (2016).
25. Suk, H. il, Lee, S. W. & Shen, D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage* **101**, 569–582 (2014).
26. Shin, H.-C. *et al.* *Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation*.
27. Jo, T., Nho, K. & Saykin, A. J. Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience* **11**, 220 (2019).
28. Liu, S. *et al.* Early diagnosis of Alzheimer’s disease with deep learning. *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014* 1015–1018 (2014) doi:10.1109/ISBI.2014.6868045.
29. Yan, K., Wang, X., Lu, L. & Summers, R. M. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. <https://doi.org/10.1117/1.JMI.5.3.036501> **5**, 036501 (2018).
30. Cireşan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **8150 LNCS**, 411–418 (2013).
31. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* **115**, 211–252 (2015).
32. Decuyper, M., Stockhoff, M., Vandenberghe, S., -, al & Ying, X. An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series* **1168**, 022022 (2019).
33. Post-NIPS*95 Workshop on Transfer in Inductive Systems. http://socrates.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/transfer.workshop.1995.html.
34. Zhuang, F. *et al.* A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE* vol. 109 43–76 Preprint at <https://doi.org/10.1109/JPROC.2020.3004555> (2021).
35. Ribani, R. & Marengoni, M. A Survey of Transfer Learning for Convolutional Neural Networks; A Survey of Transfer Learning for Convolutional Neural Networks. (2019) doi:10.1109/SIBGRABI-T.2019.00010.
36. Transfer learning & fine-tuning. https://keras.io/guides/transfer_learning/.

37. Fine-tuning with Keras and Deep Learning - PyImageSearch.
<https://pyimagesearch.com/2019/06/03/fine-tuning-with-keras-and-deep-learning/>.
38. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2014).
39. Szegedy, C. *et al.* *Going Deeper with Convolutions*. (2015).
40. Szegedy, C., Vanhoucke, V., Ioffe, S. & Shlens, J. *Rethinking the Inception Architecture for Computer Vision*. (2016).
41. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition*.
<http://image-net.org/challenges/LSVRC/2015/> (2016).
42. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. (2016).
43. Tan, M. & Le, Q. v. *EfficientNetV2: Smaller Models and Faster Training*.
<https://github.com/google/> (2021).
44. Tan, M. & Le, Q. v. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*.
45. Xu, F. *et al.* Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11839 LNAI**, 563–574 (2019).
46. Barredo Arrieta, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **58**, 82–115 (2020).
47. Samek, W., Wiegand, T. & Müller, K.-R. *EXPLAINABLE ARTIFICIAL INTELLIGENCE: UNDERSTANDING, VISUALIZING AND INTERPRETING DEEP LEARNING MODELS*.
48. Tjoa, E. & Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS* **32**, 4793 (2021).
49. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. *Learning Deep Features for Discriminative Localization*. <http://cnnlocalization.csail.mit.edu>.
50. Selvaraju, R. R. *et al.* *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. <http://gradcam.cloudev.org> (2017).
51. Petersen, R. C. *et al.* Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Neurology* **74**, 201–209 (2010).
52. Jahn Andy. Andy’s Brain Book 1.0 documentation.
<https://andysbrainbook.readthedocs.io/en/latest/FrequentlyAskedQuestions/FrequentlyAskedQuestions.html>.
53. Plasencia, O. D. Alzheimer Diagnosis with Deep Learning: Data Preprocessing. *Towards Data Science* <https://towardsdatascience.com/alzheimer-diagnosis-with-deep-learning-data-preprocessing-4521d6e6eb> (2021).
54. Evans, A. C. *et al.* Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage* **1**, 43–53 (1992).

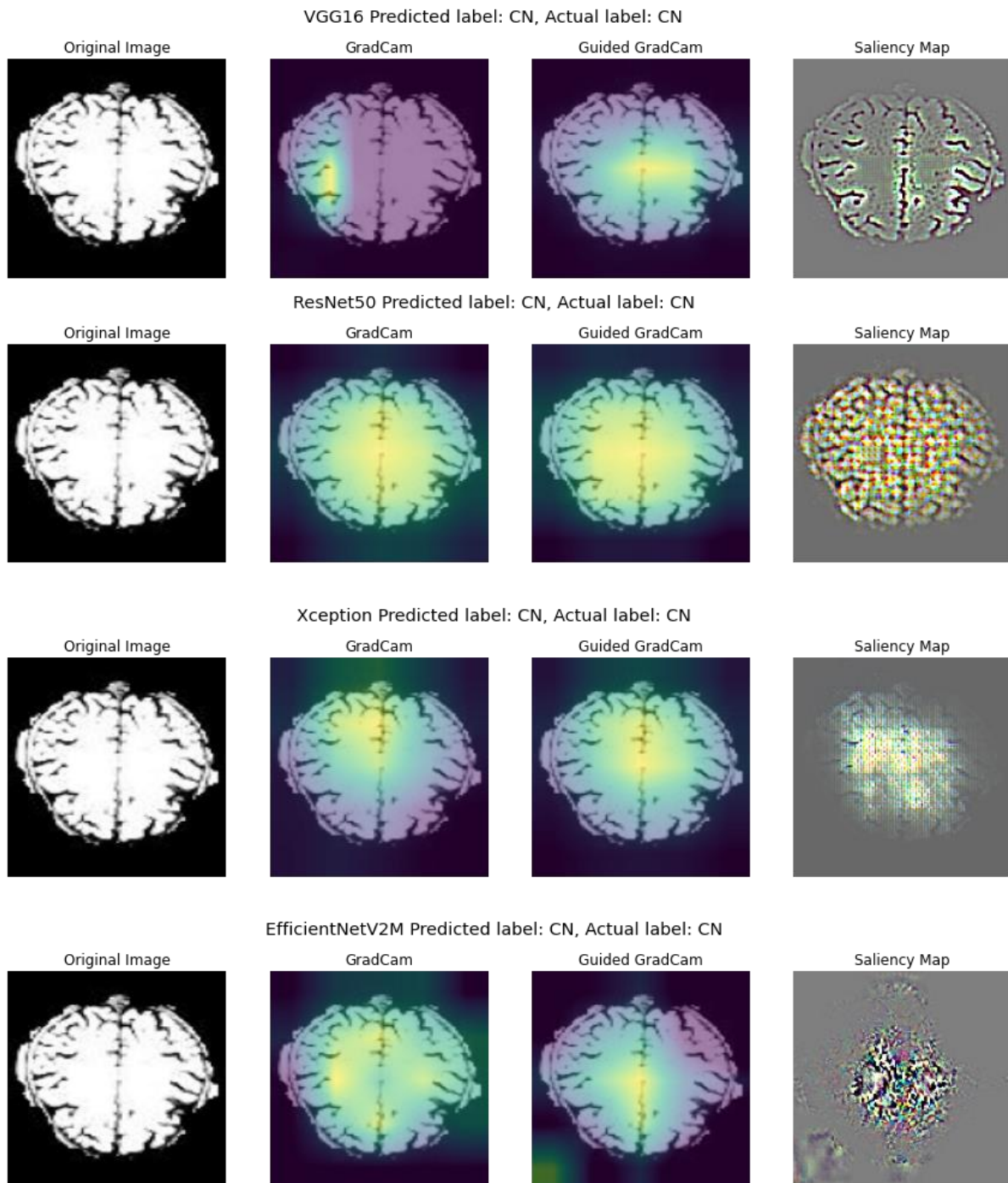
55. Collins, D. L., Neelin, P., Peters, T. M. & Evans, A. C. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography* **18**, 192–205 (1994).
56. Evans, A. C. *et al.* 3D statistical neuroanatomical models from 305 MRI volumes. *IEEE Nuclear Science Symposium & Medical Imaging Conference* 1813–1817 (1994) doi:10.1109/NSSMIC.1993.373602.
57. Pawlowski, N. *et al.* DLTk: State of the Art Reference Implementations for Deep Learning on Medical Images. (2017) doi:10.48550/arxiv.1711.06853.
58. Marstal, K., Berendsen, F., Staring, M. & Klein, S. SimpleElastix: A user-friendly, multi-lingual library for medical image registration.
59. Yaniv, Z., Lowekamp, B. C., Johnson, H. J. & Beare, R. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *Journal of Digital Imaging* **31**, 290–303 (2018).
60. Lowekamp, B. C., Chen, D. T., Ibáñez, L. & Blezek, D. The design of simpleITK. *Frontiers in Neuroinformatics* **7**, 45 (2013).
61. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
62. Gorgolewski, K. *et al.* Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics* **5**, 13 (2011).
63. Ding, Y. *et al.* A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain. *Radiology* **290**, 456–464 (2019).
64. Wen, J. *et al.* Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation. *Medical Image Analysis* **63**, (2020).
65. Bäckström, K., Mahmood Nazari, Irene Yu-Hua Gu & Asgeir Store Jakola. *An efficient 3D deep convolutional network for Alzheimer’s disease diagnosis using MR images.* (IEEE, 2018).
66. Yagis, E. *et al.* Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Scientific Reports /* **11**, 22544 (123AD).
67. Valliani, A. & Soni, A. Deep residual nets for improved Alzheimer’s diagnosis. in *ACM-BCB 2017 - Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 615 (Association for Computing Machinery, Inc, 2017). doi:10.1145/3107411.3108224.
68. Aderghal, K., Boissenin, M., Benois-Pineau, J., Catheline, G. & Afdel, K. Classification of sMRI for AD diagnosis with convolutional neuronal networks: A pilot 2-D+ ϵ study on ADNI. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **10132 LNCS**, 690–701 (2017).
69. Aderghal, K. *et al.* Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning. in *Proceedings - IEEE Symposium on Computer-Based Medical Systems* vols. 2018-June 345–350 (Institute of Electrical and Electronics Engineers Inc., 2018).
70. Korolev, S., Safiullin, A., Belyaev, M. & Dodonova, Y. Residual and plain convolutional neural networks for 3D brain MRI classification. *Proceedings - International Symposium on Biomedical Imaging* 835–838 (2017) doi:10.1109/ISBI.2017.7950647.

71. Islam, J. & Zhang, Y. Deep convolutional neural networks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using 3D brain MRI. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11309 LNAI**, 359–369 (2018).
72. Qiu, S. *et al.* Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **10**, 737–749 (2018).
73. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *COMMUNICATIONS OF THE ACM* **60**, (2017).
74. VGG16 architecture. <https://iq.opengenus.org/vgg16/>.
75. Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification) | by Sik-Ho Tsang | Towards Data Science. <https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568>.
76. tf.keras.layers.Rescaling | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Rescaling.
77. Transfer learning & fine-tuning. https://keras.io/guides/transfer_learning/.
78. tf.keras.layers.Dropout | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dropout.
79. Keras Applications. <https://keras.io/api/applications/>.
80. tf.keras.layers.GlobalAveragePooling2D | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/layers/GlobalAveragePooling2D.
81. tf.keras.optimizers.Adam | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam.
82. tf.keras.losses.BinaryCrossentropy | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/losses/BinaryCrossentropy.
83. tf.keras.losses.SparseCategoricalCrossentropy | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy.
84. tf.keras.callbacks.ModelCheckpoint | TensorFlow Core v2.9.1. https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ModelCheckpoint.
85. Measuring Performance: The Confusion Matrix | by Rachel Draelos | Towards Data Science. <https://towardsdatascience.com/measuring-performance-the-confusion-matrix-25c17b78e516>.
86. Hoo, Z. H., Candlish, J. & Teare, D. What is an ROC curve? *Emergency Medicine Journal* **34**, 357–359 (2017).
87. Receiver Operating Characteristic (ROC) — scikit-learn 1.0.2 documentation. https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html.
88. Precision-Recall — scikit-learn 1.1.0 documentation. https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html.
89. Nakano, R. [reiinakano/scikit-plot: v0.2.1](https://doi.org/10.5281/ZENODO.293191). (2017) doi:10.5281/ZENODO.293191.

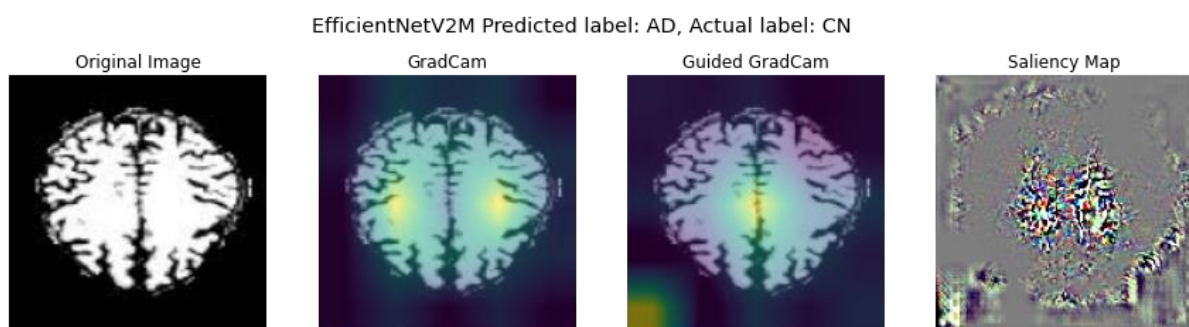
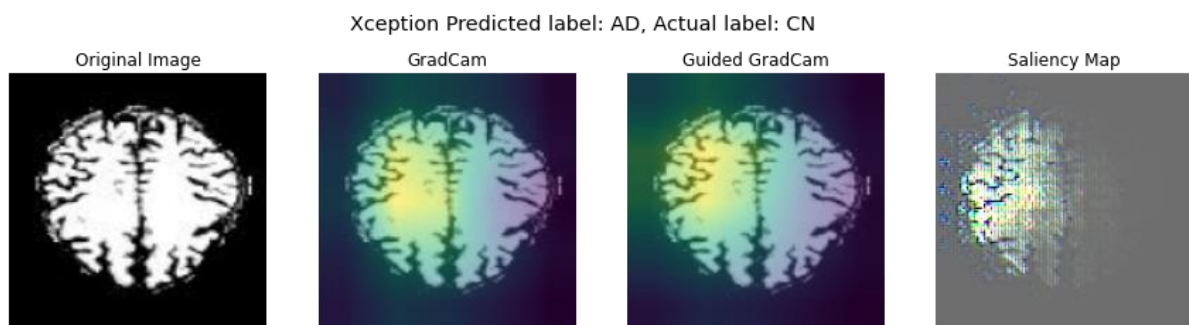
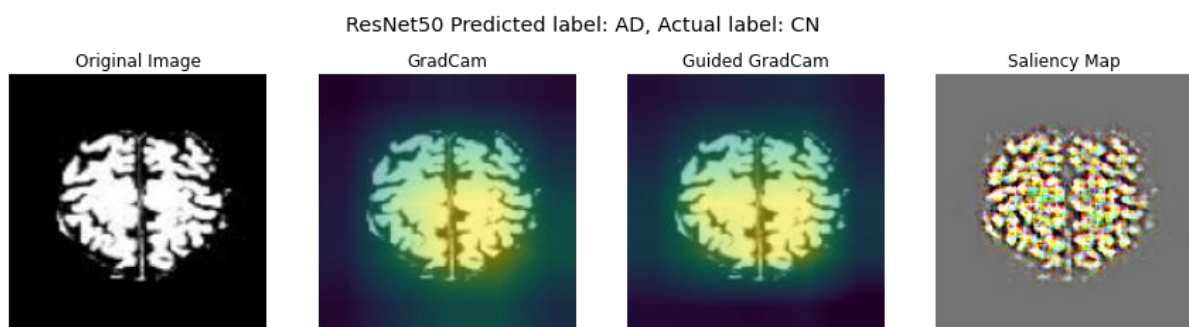
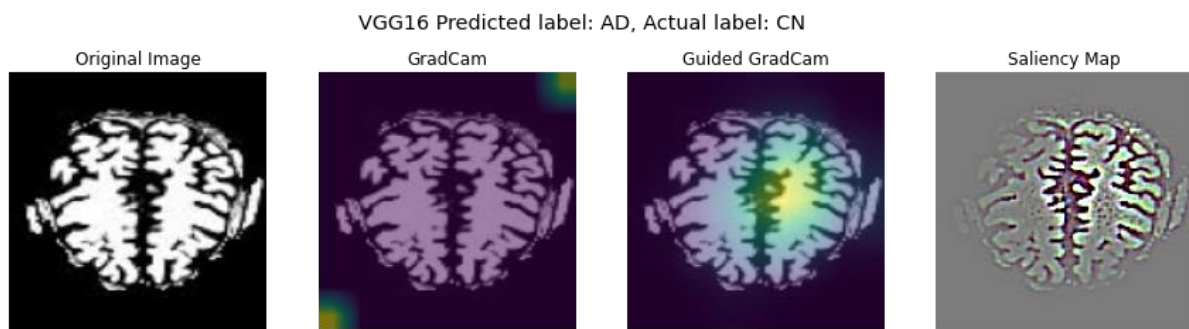
90. Tour of Evaluation Metrics for Imbalanced Classification. <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>.
91. ismailuddin/gradcam-tensorflow-2: Grad-CAM implementation using TensorFlow 2.X code. Including guided Grad-CAM and counterfactuals. <https://github.com/ismailuddin/gradcam-tensorflow-2>.
92. Kolesnikov, A. *et al.* *Big Transfer (BiT): General Visual Representation Learning*.
93. Ridnik, T., Ben-Baruch, E., Noy, A. & Zelnik-Manor, L. *ImageNet-21K Pretraining for the Masses*. <https://github.com/Alibaba-MIIL/ImageNet21K> (2021).

Παράρτημα Α: Αποτελέσματα Επεξηγησιμότητας (Subject Level Split)

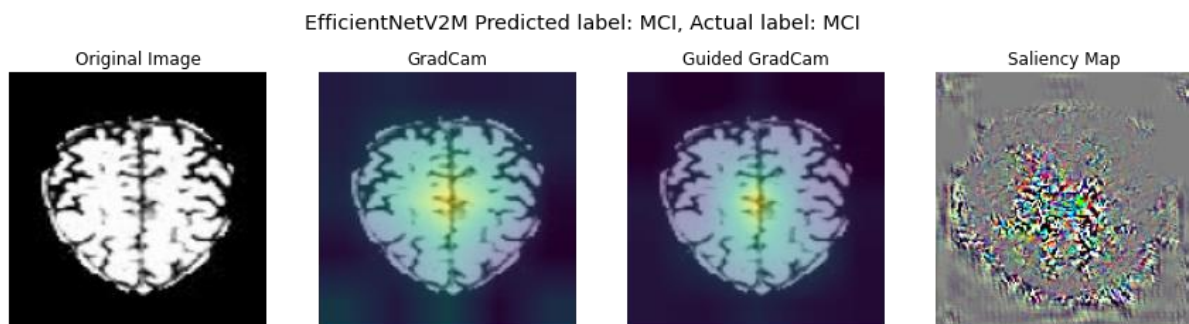
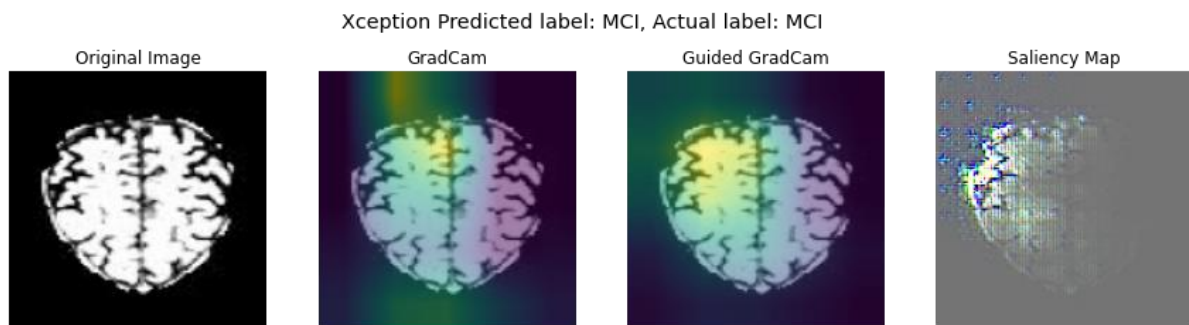
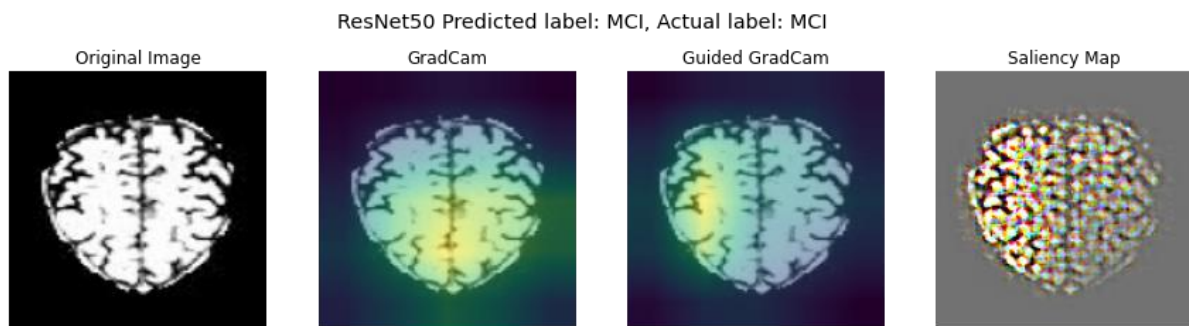
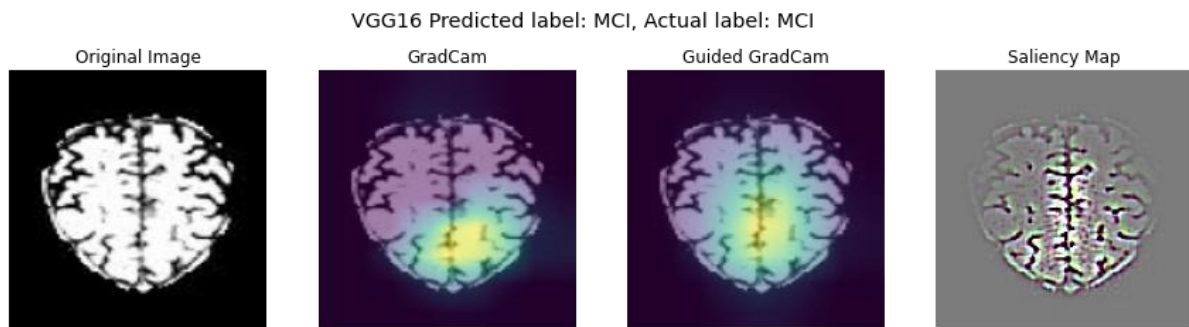
A.1 Σωστές ταξινομήσεις CN



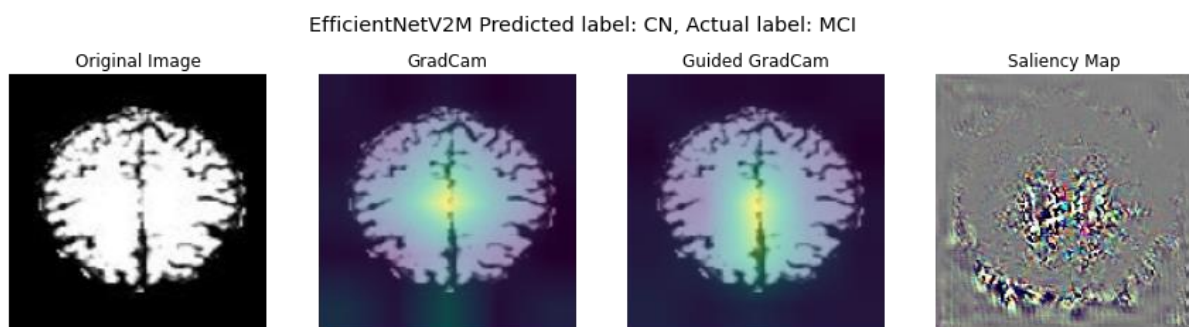
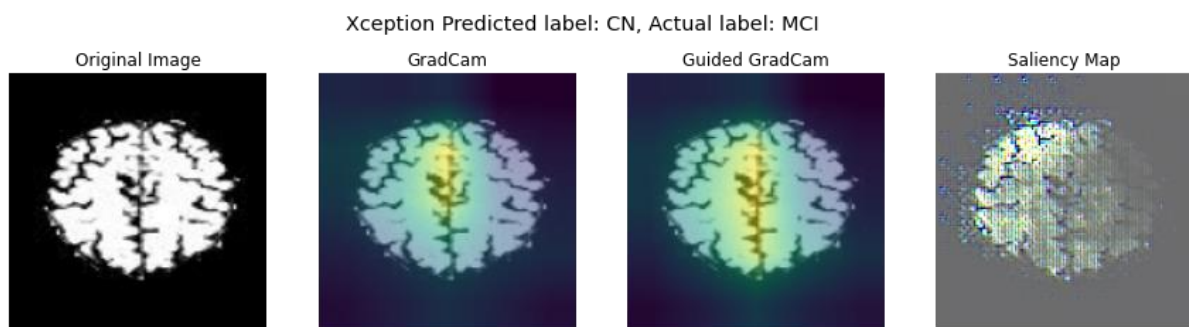
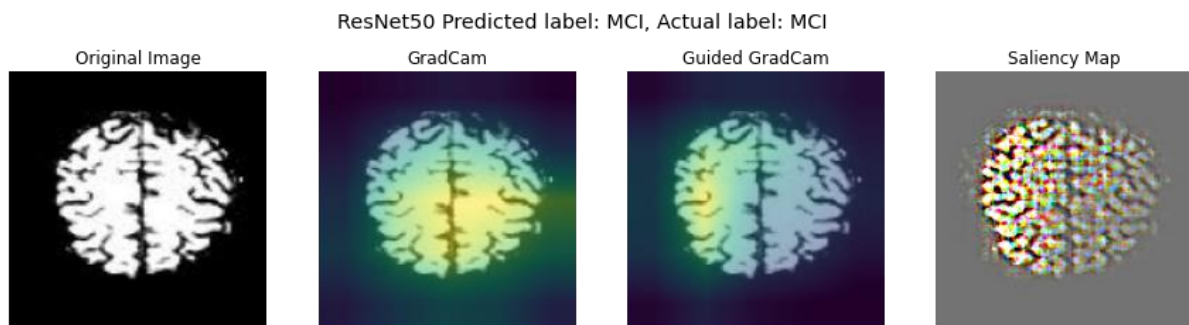
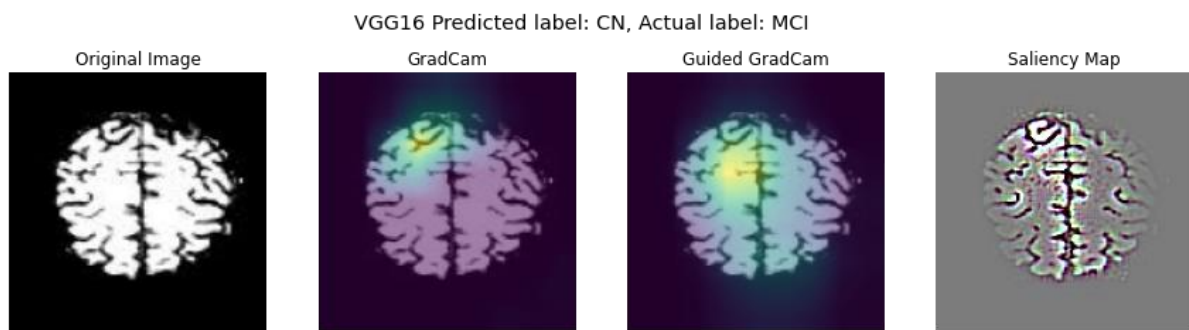
A.2 Λανθασμένες ταξινομήσεις CN



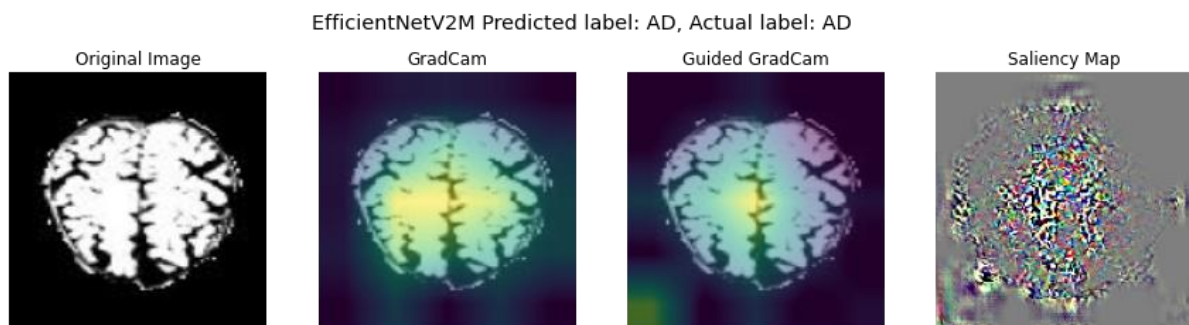
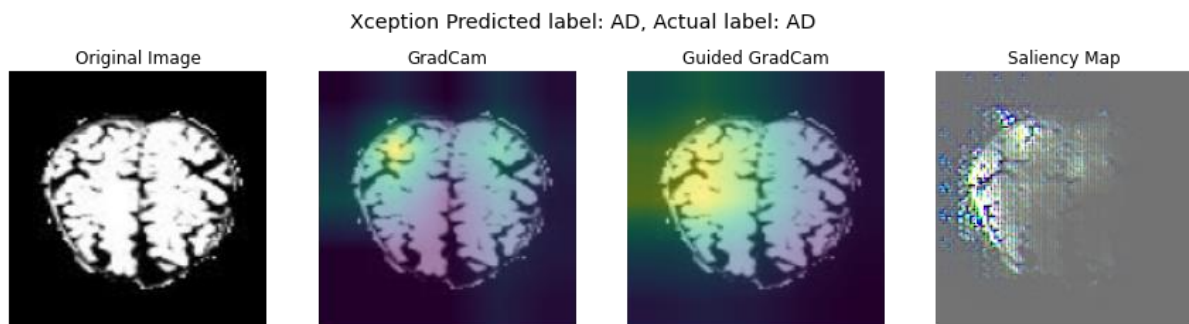
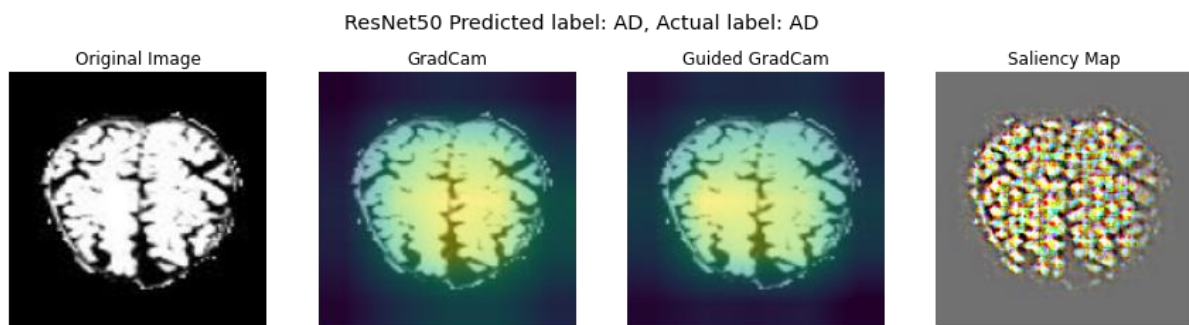
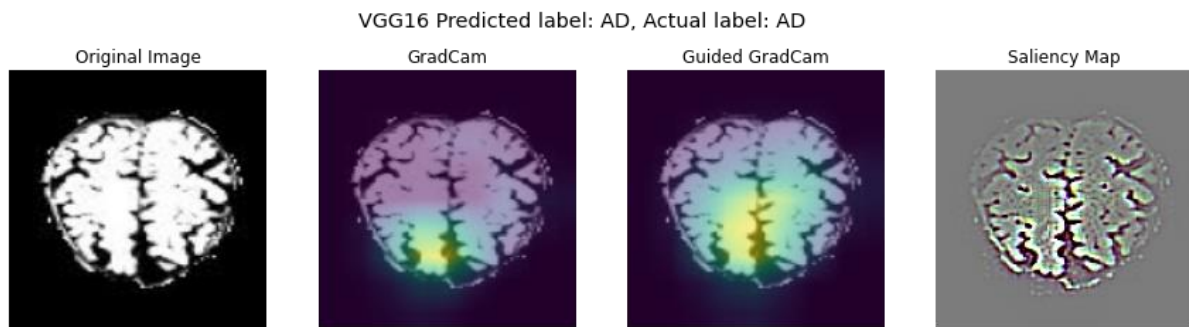
A.3 Σωστές Ταξινομήσεις MCI



A.4 Λανθασμένες Ταξινομήσεις MCI



A.5 Σωστές ταξινομήσεις AD



A.6 Λανθασμένες Ταξινομήσεις AD

