

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

Μεταπτυχιακή Διπλωματική Εργασία

**«Αξιολόγηση της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την
κατάθλιψη με τη χρήση των κριτηρίων ελέγχου CLAIM»**

**«Assessing the reporting quality of deep learning studies on depression using the
CLAIM screening criteria»**

υπό

Κουτσόγιαννη Νικολάου, Ιατρού

Υπεβλήθη για την εκπλήρωση μέρους των

απαιτήσεων για την απόκτηση του

Διπλώματος Μεταπτυχιακών Σπουδών

«Φυσικές Αρχές Βιοϊατρικής Απεικόνισης και Ακτινοπροστασία»

Τριμελής Εξεταστική Επιτροπή

Καψαλάκη Ευτυχία, Καθηγήτρια ΠΘ

Βλυχού Μαριάννα, Καθηγήτρια ΠΘ

Τσούγκος Ιωάννης, Καθηγητής ΠΘ

Λάρισα, 2023

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ	3
ΠΕΡΙΛΗΨΗ.....	4
ABSTRACT.....	5
ΕΙΣΑΓΩΓΗ	6
ΣΚΟΠΟΣ	11
ΜΕΘΟΔΟΛΟΓΙΑ	12
Στρατηγική αναζήτησης.....	12
Κριτήρια καταλληλότητας ερευνητικών μελετών	12
Εξαγωγή δεδομένων	13
Ανάλυση δεδομένων	14
ΑΠΟΤΕΛΕΣΜΑΤΑ	15
Αποτελέσματα αναζήτησης.....	15
Αποδεκτές μελέτες.....	15
Αποτελέσματα ποιοτικής αξιολόγησης των ερευνητικών άρθρων	16
ΣΥΖΗΤΗΣΗ	25
ΠΕΡΙΟΡΙΣΜΟΙ	27
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	28
ΔΗΛΩΣΗ ΣΥΓΚΡΟΥΣΗΣ ΣΥΜΦΕΡΟΝΤΩΝ.....	28
ΒΙΒΛΙΟΓΡΑΦΙΑ	29
ΠΑΡΑΡΤΗΜΑ Α	38
Αποδεκτές ερευνητικές εργασίες (άρθρα) παρούσας μελέτης	38
ΠΑΡΑΡΤΗΜΑ Β	40
Checklist for Artificial Intelligence in Medical Imaging (CLAIM):.....	40

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της μεταπτυχιακής διπλωματικής μου εργασίας, θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν στην εκπόνησή της με οποιονδήποτε τρόπο.

Αρχικά, αισθάνομαι την ανάγκη να εκφράσω τις ειλικρινείς και θερμές ευχαριστίες μου στην επιβλέπουσα καθηγήτριά μου κα Ευτυχία Καψαλάκη, καθώς και στα υπόλοιπα μέλη της τριμελούς επιτροπής, την καθηγήτριά μου κα Μαριάννα Βλυχού και τον καθηγητή μου κο Ιωάννη Τσούγκο, για την επιστημονική τους καθοδήγηση και την πολύτιμη συμβολή τους στην ολοκλήρωση της παρούσας μελέτης.

Ακόμα, οφείλω ένα μεγάλο ευχαριστώ στην κα Δήμητρα Τσιβάκα για την άριστη συνεργασία, βοήθεια και εκπαίδευση σε όλη την περίοδο των σπουδών του παρόντος μεταπτυχιακού προγράμματος, καθώς και στην κα Κλαίρη Παπαδημητρίου.

Επίσης, θα ήθελα να ευχαριστήσω τον συντονιστή διευθυντή μου κο Τριαντάφυλλο Γερούκη και τη διευθύντριά μου κα Βασιλική Καραμπέκου.

Ιδιαίτερες ευχαριστίες θα ήθελα να εκφράσω στο προσωπικό της βιβλιοθήκης της Ιατρικής Λάρισας του Πανεπιστημίου Θεσσαλίας για την άμεση και αποτελεσματική συμβολή τους, για μία ακόμα φορά, σε όλη τη διάρκεια της ερευνητικής διαδικασίας της παρούσας εργασίας.

Τέλος, θέλω να ευχαριστήσω θερμά την αδερφή μου Μαρία και τους γονείς μου Αγγελική και Αθανάσιο για τη συνεχή στήριξή τους σε όλη τη διαδρομή των σπουδών μου μέχρι σήμερα.

ΠΕΡΙΛΗΨΗ

Εισαγωγή: Η μείζων καταθλιπτική διαταραχή αποτελεί μία από τις πιο συχνά εμφανιζόμενες ψυχικές νόσους σε παγκόσμια κλίμακα και εμφανίζεται σε άτομα διαφόρων ηλικιών, με σοβαρές κοινωνικές και οικονομικές επιπτώσεις. Το γεγονός αυτό, σε συνδυασμό με τη ραγδαία εξέλιξη της τεχνολογίας, κατευθύνουν την επιστημονική κοινότητα σε ερευνητικά πρωτόκολλα που σχετίζονται με τις διάφορες πτυχές της, ιδίως τη διάγνωση και τη θεραπεία της, με την βοήθεια τεχνικών τεχνητής νοημοσύνης και τα τελευταία χρόνια τεχνικών βαθιάς μάθησης.

Σκοπός μελέτης: Αξιολόγηση της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την κατάθλιψη, οι οποίες δημοσιεύθηκαν από το 2012 έως και το 2022, με τη χρήση των κριτηρίων ελέγχου CLAIM.

Μεθοδολογία: Πραγματοποιήθηκε αναζήτηση ερευνητικών μελετών στην ηλεκτρονική βάση δεδομένων MEDLINE (via PubMed), με λέξεις – κλειδιά «βαθιά μάθηση» και «μείζων καταθλιπτική διαταραχή» από 01/01/2012 έως και 31/12/2022. Τα άρθρα της παραπάνω αναζήτησης αξιολογήθηκαν ως προς την καταλληλότητά τους με βάση τον τίτλο, την περίληψη και το πλήρες κείμενό τους. Τα πλήρη κείμενα των αποδεκτών άρθρων αξιολογήθηκαν με τη χρήση των κριτηρίων ελέγχου CLAIM. Η ορθή αναφορά για κάθε ένα από τα κριτήρια αξιολογήθηκε και βαθμολογήθηκε, καθώς επίσης πραγματοποιήθηκε και στατιστική ανάλυση όλων των παραπάνω δεδομένων.

Αποτελέσματα: Η ποιότητα αναφοράς μελετών στα πρώτα άρθρα που δημοσιεύθηκαν, ελέγχεται σε χαμηλά ποσοστά, ενώ μετά τη δημοσίευση των κριτηρίων CLAIM, τα ποσοστά εμφανίζουν μικρού βαθμού αύξηση σε σταθερό ικανοποιητικό επίπεδο, χωρίς όμως στατιστικά σημαντική βελτίωση ($p\text{-value} > 0,05$). Η μέση τιμή ποιοτικής αναφοράς κριτηρίων όλων των άρθρων είναι 68,8%, ενώ χαμηλά ποσοστά ποιότητας αναφοράς εντοπίζονται στο τμήμα της μεθοδολογίας και των αποτελεσμάτων.

Συμπεράσματα: Παρουσιάζονται μεγάλες διακυμάνσεις στα ποσοστά της αξιολόγησης της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την κατάθλιψη με τη χρήση των κριτηρίων ελέγχου CLAIM και χρειάζονται περαιτέρω μελέτες για την επαλήθευση, την αξιολόγηση και την γενίκευση των αποτελεσμάτων.

Λέξεις κλειδιά: μείζων καταθλιπτική διαταραχή, κατάθλιψη, τεχνητή νοημοσύνη, βαθιά μάθηση, κριτήρια CLAIM

ABSTRACT

Introduction: Major depressive disorder (MDD) is a highly prevalent mental disorder worldwide and occurs to people of various ages, with enormous social and economic impact. This fact, combined with the rapid development of technology, are leading the scientific community on conducting research on major depressive disorder and all its aspects (e.g. diagnosis, therapy), with artificial intelligence and deep learning techniques.

Objective: Assessment of the reporting quality of studies on deep learning for depression published from 2012 to 2022 using the CLAIM checklist (Checklist for Artificial Intelligence in Medical Imaging).

Methods: A comprehensive electronic PubMed search was conducted (MEDLINE database) for articles referring to major depressive disorder and deep learning techniques, from January 01 2012 until December 31 2022. Search results were first screened for eligibility by title, then by abstract and finally by full text review. The full texts of the eligible for inclusion articles were assessed using the CLAIM checklist. The reporting of the checklist's items was evaluated and graded. Scores were calculated for each article and item (statistical analysis).

Results: The adequacy of reporting quality of studies on deep learning for depression, before the CLAIM checklist was low. After the publication of CLAIM checklist, the quality of studies on deep learning for MDD has shown a trend of increase (p-value >0,05). The overall compliance for the eligible articles was 68,8%. The compliance was particularly low for the methods and the results.

Conclusions: The reporting quality of studies on deep learning for depression ranges widely and further improvement is necessary to assess their validity.

Key words: major depressive disorder, depression, artificial intelligence, deep learning, CLAIM checklist

ΕΙΣΑΓΩΓΗ

Η μείζων καταθλιπτική διαταραχή αποτελεί μία από τις πιο συχνά εμφανιζόμενες ψυχικές νόσους παγκοσμίως και εμφανίζεται σε άτομα διαφόρων ηλικιών, με σοβαρές κοινωνικές και οικονομικές επιπτώσεις. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας ο παγκόσμιος επιπολασμός της νόσου ξεπερνά το 4%, με αποτέλεσμα περισσότερα από 320 εκατομμύρια άνθρωποι να πάσχουν από κατάθλιψη [1]. Εμφανίζεται περισσότερο σε γυναίκες από ότι σε άνδρες [2, 3], καθώς και σε άτομα τα οποία διανύουν τα εφηβικά τους έτη [4, 5]. Παρόλα αυτά, τα τελευταία χρόνια, παρατηρούμε, όλο και περισσότερο, βαθμιαία αύξηση των ποσοστών κατάθλιψης στους ηλικιωμένους, το οποίο αποδίδεται σε μεγάλο βαθμό, κυρίως σε κοινωνικούς παράγοντες [6].

Η καταθλιπτική νόσος, όπως έχει αποδειχθεί στο πέρασμα των χρόνων, διαθέτει γενετικό υπόβαθρο [7], το οποίο σε συνδυασμό με κοινωνικούς και περιβαλλοντικούς παράγοντες, μπορούν να οδηγήσουν σε ακόμα μεγαλύτερα ποσοστά εμφάνισης της νόσου. Στους κοινωνικούς παράγοντες περιλαμβάνονται η παραμέληση από τους γονείς, η πρόωμη απομάκρυνση-απώλεια ενός συγγενικού προσώπου, η μακρά περίοδος απόστασης από έναν γονέα κατά την παιδική ηλικία, η ανεργία και η έλλειψη διαπροσωπικών σχέσεων [8], ενώ κατά περιόδους μπορεί να προκύψουν συνθήκες εκτεταμένης απομόνωσης, όπως συνέβη και πρόσφατα με την παρουσία του ιού COVID-19 σε παγκόσμια κλίμακα [9].

Η κλινική εικόνα των πασχόντων περιλαμβάνει κυρίως συμπτώματα, τα οποία αφορούν στο συναίσθημα, όπως διάχυτη και επίμονη πτώση του συναισθήματος, μειωμένο ενδιαφέρον – ευχαρίστηση και κοινωνική απόσυρση. Επίσης, δύναται να περιλαμβάνει συμπτώματα που αφορούν σε βιολογικές επιπτώσεις, όπως είναι η σημαντική απώλεια σωματικού βάρους, διαταραχές ύπνου, ψυχοκινητική επιβράδυνση και ευερεθιστότητα, καθώς και συμπτώματα που σχετίζονται με γνωσιακές αλλαγές, όπως αισθήματα ενοχής, αναξιότητας, αυτοκατηγορίες, απελπισία και σκέψεις αυτοκτονίας. Πολύ συχνά διαλάθει της διάγνωσης, διότι μπορεί να εμφανιστεί με τη μορφή σωματικών συμπτωμάτων, όπως για παράδειγμα κεφαλαλγία, κόπωση, διαταραχή μνήμης και άτυπα θωρακικά άλγη, τα οποία μπορούν να αποπροσανατολίσουν τη διαγνωστική διαδικασία [10].

Η διάγνωση της κατάθλιψης στηρίζεται κυρίως στην κλινική εικόνα του εξεταζόμενου, με βάση τα κριτήρια DSM-5 (The Diagnostic and Statistical Manual of Mental Disorders 5th edition American Psychiatric Association), ενώ απαραίτητες είναι και οι εργαστηριακές και απεικονιστικές εξετάσεις (πχ. αιματολογικές εξετάσεις, μαγνητική τομογραφία εγκεφάλου), κυρίως για τον αποκλεισμό άλλων νοσημάτων. Το ατομικό και οικογενειακό αναμνηστικό προσφέρουν κάποιες επιπλέον πληροφορίες, οι οποίες συνήθως συνεισφέρουν σημαντικά στην τελική διάγνωση [11].

Η θεραπεία της καταθλιπτικής νόσου στηρίζεται κατά βάση στην ψυχοθεραπεία και την φαρμακοθεραπεία (πχ. SSRIs, SNRIs, τρικυκλικά αντικαταθλιπτικά, MAOIs) [12, 13]. Άλλες πιθανές συνδυαστικές θεραπείες αποτελούν η ηλεκτροσπασμοθεραπεία (ECT) και η Διακρανιακή Μαγνητική Διέγερση (TMS) [14, 15]. Επιπρόσθετα, οι ψυχοκοινωνικές παρεμβάσεις είναι σημαντικές για την ανακούφιση του ασθενούς σε περιπτώσεις προβληματικών διαπροσωπικών σχέσεων, οικονομικών προβλημάτων και δυσκολιών στέγασης. Τέλος, είναι μείζονος σημασία ο έλεγχος και η διαχείριση του υπαρκτού κινδύνου αυτοκτονικών τάσεων των ασθενών [16, 17].

Κάθε δεδομένο σχετικά με τη μείζονα καταθλιπτική διαταραχή, είτε αυτό αφορά για παράδειγμα στην αναζήτηση της γενετικής προδιάθεσης των ασθενών, είτε στη διάγνωση και τη θεραπεία τους, όπως έχει αποδειχθεί και με την ευρεία υποδιάγνωση της νόσου στο πέρασμα του χρόνου, χαρακτηρίζεται από συνθετότητα και πολυπλοκότητα [18]. Σε συνδυασμό με τη ραγδαία εξέλιξη της τεχνολογίας, τη συνεχή ανάπτυξη της υπολογιστικής ισχύος των υπολογιστών και την εμφάνιση της τεχνητής νοημοσύνης [19, 20], το επιστημονικό ενδιαφέρον στράφηκε ιδιαίτερα την τελευταία δεκαετία, στη διερεύνηση της μείζονας καταθλιπτικής νόσου με τεχνικές τεχνητής νοημοσύνης και τεχνικές μηχανικής και βαθιάς μάθησης [21, 22].

Η τεχνητή νοημοσύνη (artificial intelligence) αποτελεί έναν τομέα της επιστήμης των υπολογιστών που ασχολείται με τη σχεδίαση και τη δημιουργία ευφυών προγραμμάτων, τα οποία δύνανται να μιμηθούν τις ανθρώπινες γνωστικές ικανότητες [23,24], όπως την επίλυση προβλημάτων, μάθηση και εξαγωγή συμπερασμάτων, ενώ ταυτόχρονα να μπορούν να προσαρμόζονται στο περιβάλλον τους, με τρόπο παρόμοιο με αυτόν των ανθρώπων [25]. Τα πρώτα στοιχεία σχετικά με την ιδέα της τεχνητής νοημοσύνης απαντώνται στην ελληνική μυθολογία, όπως περιγράφονται στην Ιλιάδα (Όμηρος 9^{ος} – 8^{ος} αιώνας π.Χ.), με τον Ήφαιστο να εμφανίζεται υποβασταζόμενος από

δύο θεραπεινίδες - ανθρωπόμορφες μηχανές (ρομπότ), οι οποίες μοιάζουν σαν ζωντανές, καθώς διαθέτουν δύναμη, λογική και φωνή. Ένα άλλο ανθρωπόμορφο ρομπότ στην ελληνική μυθολογία είναι ο Τάλως, ένα κατασκεύασμα του Ήφαιστου, ο οποίος τον είχε δωρίσει στο βασιλιά Μίνωα για να φυλάει την Κρήτη [26]. Τα πρώτα ιστορικά στοιχεία σχετικά με την ιδέα της τεχνητής νοημοσύνης απαντώνται στον 4^ο αιώνα π.Χ., εποχή κατά την οποία έζησε ο Αριστοτέλης (384-322 π. Χ.) και πιο συγκεκριμένα στους συλλογισμούς του Αριστοτέλη (Αριστοτέλεια συλλογιστική) [27], ενώ ήδη από την αρχαιότητα , διατυπώνεται το φιλοσοφικό ερώτημα «τι είναι ευφυΐα και πως αυτή εκφράζεται» [28-34].

Η επίλυση προβλημάτων με ευφυείς προσεγγίσεις βασίζεται στην δυνατότητα των συστημάτων να μαθαίνουν, γνωστή ως μηχανική μάθηση (machine learning) [35]. Συνιστά ένα υποπεδίο της επιστήμης των υπολογιστών και αφορά στην κατασκευή αλγορίθμων, οι οποίοι με αυτόματο τρόπο (δηλαδή χωρίς προγραμματισμό) μαθαίνουν μέσω επαγωγικών μεθόδων (π.χ. δέντρα απόφασης), αναγνωρίζουν και εντοπίζουν διάφορα πρότυπα στα δεδομένα εισόδου (π.χ. νευρωνικά δίκτυα), τα οποία πρότυπα τελικά χρησιμοποιούν για να προβλέψουν μελλοντικά δεδομένα ή να εκτελέσουν άλλους τύπους αποφάσεων σε καθεστώς αβεβαιότητας (π.χ. γενετικοί αλγόριθμοι). Με άλλα λόγια είναι ένα σύστημα, το οποίο βελτιώνει την απόδοσή του κατά την εκτέλεση μιας συγκεκριμένης εργασίας μέσω εμπειρίας, χωρίς να υπάρχει ανάγκη να προγραμματιστεί εκ νέου [36].

Η μηχανική μάθηση μπορεί να διακριθεί σε τρεις κύριες κατηγορίες ανάλογα με τη φύση της εκπαίδευσης (ανάλογους τρόπους μάθησης με τους οποίους μαθαίνουν οι άνθρωποι) και το είδος της ανατροφοδότησης του συστήματος: επιτηρούμενη μάθηση (supervised learning), μη επιτηρούμενη μάθηση (unsupervised learning) και ενισχυτική μάθηση (reinforcement learning). Η δημιουργία και ο αναλυτικός σχεδιασμός απαιτητικών αλγορίθμων εμφανίζει πολλαπλές δυσκολίες για τους ερευνητές, με αποτέλεσμα να προκύπτει η ανάγκη δημιουργίας αλγορίθμων και συστημάτων, οι οποίοι να παρέχουν νέες λύσεις βασισμένες στην ικανότητα βελτίωσης μέσω της εμπειρίας είτε από τα ίδια – αρχικά δεδομένα είτε από νέα δεδομένα. Στο πεδίο της μηχανικής μάθησης, για κάθε πρόβλημα προς επίλυση, υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος, ο οποίος μπορεί να χρησιμοποιηθεί [37].

Η βαθιά μάθηση (deep learning) είναι μία τεχνική μηχανικής μάθησης, μέσω της οποίας οι αλγόριθμοι μπορούν να μάθουν και να αναγνωρίσουν τα διάφορα χαρακτηριστικά του συστήματος ενδιαφέροντος μέσω πολλαπλών επιπέδων μονάδων επεξεργασίας, χωρίς να απαιτείται ανθρώπινη παρέμβαση. Δηλαδή, η βαθιά μάθηση είναι εξ ολοκλήρου μια αυτόματη διαδικασία, μέσω της οποίας προκύπτουν συμπεράσματα, χωρίς να απαιτείται προεργασία από επιστήμονες ως προς τα διάφορα κριτήρια αξιολόγησης, με επακόλουθο τα αποτελέσματα να στηρίζονται σε περισσότερα χαρακτηριστικά, καθώς και να εντοπίζονται επιπλέον ευρήματα [38]. Η ιδέα αυτής της τεχνικής στηρίχτηκε στον τρόπο επεξεργασίας των οπτικών πληροφοριών από τον οφθαλμό (αμφιβληστροειδής χιτώνας) έως και τον εγκέφαλο (οπτικό νεύρο, ινιακός λοβός), όπου εντοπίζονται πολλαπλά επίπεδα επεξεργασίας [39]. Αντίστοιχα, οι αλγόριθμοι βαθιάς μάθησης προσπαθούν να εξάγουν συμπεράσματα, όπως ο ανθρώπινος εγκέφαλος μέσω πολύπλοκων διαδικασιών και είναι απαραίτητος μεγάλος όγκος δεδομένων, για αποτελέσματα μεγαλύτερης ακρίβειας και αξιοπιστίας [40].

Ο τομέας της βιοϊατρικής απεικόνισης διαθέτει σήμερα μεγάλο όγκο δεδομένων, με αποτέλεσμα να αποτελεί ιδανική επιλογή για τη χρήση εργαλείων βαθιάς μάθησης, με σκοπό την υποστήριξη των ιατρικών διαδικασιών (π.χ. πρόληψη, διάγνωση). Ένα τέτοιο παράδειγμα αποτελεί η χρήση τεχνικών βαθιάς μάθησης (πχ. νευρωνικά δίκτυα CNN) στην ανάλυση ιατρικών εικόνων [41].

Όπως και στην μηχανική μάθηση, η βαθιά μάθηση μπορεί να διακριθεί ανάλογα με τη φύση της εκπαίδευσης σε επιτηρούμενη μάθηση (supervised learning) [42], μη επιτηρούμενη μάθηση (unsupervised learning) και υβριδικού τύπου μάθηση (hybrid learning). Η αξιολόγηση των συστημάτων βαθιάς μάθησης στηρίζεται στην κατασκευή και την εκπαίδευση αρχιτεκτονικών βαθιάς μάθησης, οι οποίες είναι τεχνητά νευρωνικά δίκτυα πολλαπλών στρωμάτων και διακρίνονται ανάλογα με τα δεδομένα εισόδου και τους τελικούς ερευνητικούς στόχους σε: βαθιά νευρωνικά δίκτυα (Deep Neural Network - DNN), συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Network - CNN) και αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Network - RNN), ενώ τα τελευταία χρόνια παρουσιάζονται όλο και περισσότερες τεχνικές, όπως η χρήση αυτόματων κωδικοποιητών (Autoencoder) και μηχανών Boltzman (Boltzman Machine) [43, 44].

Η αποτύπωση του παραπάνω ενδιαφέροντος της επιστημονικής κοινότητας, για την χρήση τεχνικών τεχνητής νοημοσύνης στη μείζονα καταθλιπτική διαταραχή, εντοπίζεται ήδη σε ερευνητικές μελέτες, οι οποίες δημοσιεύονται στα τέλη του 20 αιώνα [45, 46]. Την τελευταία δεκαετία όμως, παρατηρείται έντονο ενδιαφέρον σχετικά με την βελτίωση της διάγνωσης των ατόμων που πάσχουν από κατάθλιψη, με αποτέλεσμα να πραγματοποιούνται όλο και περισσότερες έρευνες στον τομέα αυτό, βασισμένες στη χρήση διαφορετικών διαγνωστικών μεθόδων (π.χ. χρήση ηλεκτροεγκεφαλογραφήματος, ιατρικής απεικόνισης) [47,48].

Όπως και στο παρελθόν χρησιμοποιήθηκαν κριτήρια αξιολόγησης των διαφορετικών ερευνητικών μεθοδολογιών (π.χ. CONSORT) [49, 50], με την ίδια προσέγγιση πριν από λίγα χρόνια παρουσιάστηκε ένα πρότυπο ελέγχου και αξιολόγησης των μελετών τεχνικών τεχνητής νοημοσύνης και στην ιατρική απεικόνιση (Checklist for Artificial Intelligence in Medical Imaging -CLAIM) [51-61].

ΣΚΟΠΟΣ

Ο σκοπός της παρούσας μελέτης είναι η αξιολόγηση της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την κατάθλιψη, οι οποίες δημοσιεύθηκαν από το 2012 έως και το 2022, με τη χρήση των κριτηρίων ελέγχου CLAIM (Checklist for Artificial Intelligence in Medical Imaging) [62].

ΜΕΘΟΔΟΛΟΓΙΑ

Στρατηγική αναζήτησης

Η παρούσα μελέτη είναι βασισμένη σε βιβλιογραφική ανασκόπηση ερευνητικών εργασιών. Η αναζήτηση των ερευνητικών άρθρων πραγματοποιήθηκε στην ηλεκτρονική βάση δεδομένων MEDLINE (via PubMed), με κύριους όρους τις λέξεις – κλειδιά «deep learning and major depressive disorder». Χρησιμοποιήθηκαν επίσης, στην αναζήτηση και οι παρακάτω όροι / λέξεις - κλειδιά: «magnetic resonance imaging and depression and deep learning», «MRI and depression and deep learning», «imaging and depression and deep learning», «artificial intelligence and depression», «artificial intelligence and depression and imaging». Το αποδεκτό χρονικό διάστημα δημοσίευσης των ερευνητικών μελετών προσδιορίστηκε από 01/01/2012 έως και 31/12/2022, ενώ έγιναν αποδεκτές μελέτες, οι οποίες αφορούσαν μόνο σε ανθρώπους.

Κριτήρια καταλληλότητας ερευνητικών μελετών

Οι ερευνητικές εργασίες που έγιναν αποδεκτές για περαιτέρω ανάλυση και αξιολόγηση έπρεπε να πληρούν ταυτόχρονα όλα τα παρακάτω κριτήρια (inclusion criteria):

1. Θεματολογία σχετική μόνο με τη μείζονα καταθλιπτική διαταραχή (έγιναν αποδεκτές όλες οι εργασίες που αφορούσαν κάθε πτυχή της μείζονας καταθλιπτικής διαταραχής: γενετικό υπόβαθρο, επιδημιολογία, κλινική εικόνα, διάγνωση, θεραπεία, επιπλοκές, πρόληψη).
2. Θεματολογία σχετική με τεχνικές βαθιάς μάθησης (πιο συγκεκριμένα έγιναν αποδεκτές μελέτες που περιλάμβαναν τουλάχιστον μία κύρια τεχνική βαθιάς μάθησης).
3. Δημοσίευση εργασίας μόνο μεταξύ 01/01/2012 έως και 31/12/2022.
4. Εργασίες σχετικές μόνο με απεικόνιση σε ανθρώπους.
5. Εργασίες δημοσιευμένες μόνο στην αγγλική γλώσσα.

Επιπρόσθετα, δεν έγιναν αποδεκτές (exclusion criteria) συγκριτικές μελέτες, εργασίες παρατήρησης, ανασκοπήσεις, μετα-αναλύσεις, πειραματικές έρευνες σε ζώα και in vitro, καθώς και διπλότυπες εργασίες (duplication).

Τα αποτελέσματα της παραπάνω αναζήτησης αξιολογήθηκαν ως προς την καταλληλότητά τους αρχικά, με βάση τον τίτλο τους, στην συνέχεια με βάση την περίληψη και τέλος με βάση το πλήρες κείμενο των ερευνητικών εργασιών. Η διαλογή των άρθρων και η τελική επιλογή, με βάση τα παραπάνω κριτήρια καταλληλότητας, πραγματοποιήθηκαν από έναν ερευνητή (τον συγγραφέα της παρούσας μελέτης).

Εξαγωγή δεδομένων

Η παραπάνω αναζήτηση οδήγησε στην τελική επιλογή ερευνητικών εργασιών, οι οποίες αξιολογήθηκαν με τη χρήση των κριτηρίων ελέγχου CLAIM (Checklist for Artificial Intelligence in Medical Imaging).

Τα κριτήρια ελέγχου CLAIM (Checklist for Artificial Intelligence in Medical Imaging) υιοθετήθηκαν σχετικά πρόσφατα (25/03/2020) από την Ακτινολογική Εταιρεία της Βόρειας Αμερικής (RSNA). Ο έλεγχος των μελετών τεχνητής νοημοσύνης στην ιατρική απεικόνιση περιλαμβάνει συνολικά 42 κριτήρια και η ορθή χρήση τους στηρίχτηκε στις συνημμένες οδηγίες της μελέτης - αναφοράς τους [62].

Σαφείς και αναλυτικές πληροφορίες για κάθε κριτήριο στο αναμενόμενο και επιθυμητό τμήμα της εργασίας αντιστοιχεί σε θετικό εύρημα, με τιμή 1. Μη επαρκείς πληροφορίες για κάθε κριτήριο στο αναμενόμενο και επιθυμητό τμήμα της εργασίας, καθώς και επαρκείς πληροφορίες για κάθε κριτήριο σε μη αναμενόμενο και επιθυμητό τμήμα της εργασίας, αντιστοιχούν σε αρνητικό εύρημα, με τιμή 0. Εξαίρεση αποτελούν οι επαρκείς πληροφορίες στο τελικά κριτήρια (other information), οι οποίες υπολογίστηκαν με θετική τιμή 1, ακόμα και σε μη αναμενόμενο τμήμα του άρθρου.

Επίσης, τα κριτήρια 10, 27 και 31, τα οποία δεν είναι απαραίτητα να αναφέρονται – μελετώνται στις διάφορες εργασίες, στα πλαίσια διαφορετικής μεθοδολογίας κάθε εργασίας, υπολογίστηκαν στο τελικό άθροισμα των κριτηρίων στην στατιστική ανάλυση με τιμή 0, ακόμα και σε περίπτωση που οι εργασίες δεν περιλάμβαναν τέτοιου είδους πληροφορίες.

Ανάλυση δεδομένων

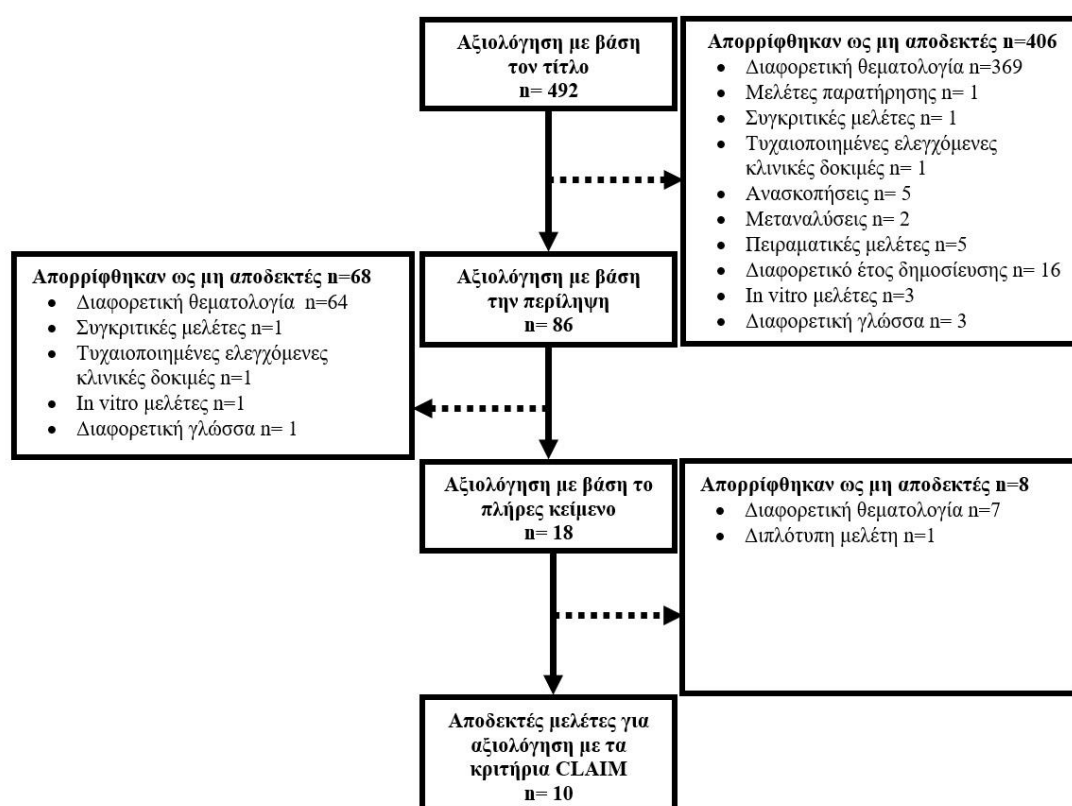
Στο σύνολο των τελικών αποδεκτών άρθρων διενεργήθηκε περιγραφική στατιστική ανάλυση, βασισμένη στα 42 κριτήρια αξιολόγησης CLAIM και ως προς το σύνολό τους, αλλά και ως προς τις διάφορες ξεχωριστές κατηγορίες – υποκατηγορίες (π.χ. ως προς τα συνολικά κριτήρια της μεθοδολογίας, των αποτελεσμάτων κ.α.), καθώς επίσης συναξιολογήθηκαν τα περιοδικά δημοσίευσης των άρθρων και το έτος δημοσίευσής τους.

Η ανάλυση των δεδομένων πραγματοποιήθηκε με τη χρήση του ηλεκτρονικού προγράμματος Microsoft Excel 365, καθώς και με τη βοήθεια του στατιστικού πακέτου - προγράμματος IBM SPSS v29.

ΑΠΟΤΕΛΕΣΜΑΤΑ

Αποτελέσματα αναζήτησης

Η αναζήτηση των ερευνητικών άρθρων που πραγματοποιήθηκε στην ηλεκτρονική βάση δεδομένων MEDLINE (via PubMed), απέδωσε συνολικά 492 πιθανές, τελικώς αποδεκτές, ερευνητικές μελέτες, οι οποίες στη συνέχεια αξιολογήθηκαν ως προς την καταλληλότητά τους με βάση τον τίτλο τους, με βάση την περίληψή τους και τέλος, με βάση το πλήρες κείμενό τους. Η στρατηγική αναζήτησης των ερευνητικών μελετών με βάση τα κριτήρια καταλληλότητας, καθώς και ο αριθμός των τελικώς αποδεκτών ερευνητικών εργασιών, συνοψίζονται στην Εικόνα 1.



Εικόνα 1: Διάγραμμα ροής για την εξαγωγή των αποδεκτών άρθρων με βάση τα κριτήρια καταλληλότητας.

Αποδεκτές μελέτες

Συνολικά, έγιναν αποδεκτές δέκα [63-72] ερευνητικές μελέτες (αναφορά των ερευνητικών μελετών ως αριθμημένα άρθρα στους πίνακες των αποτελεσμάτων, όπως αναλύεται στο Παράρτημα Α), οι οποίες στη συνέχεια αξιολογήθηκαν με τα κριτήρια CLAIM (Παράρτημα Β) (Πίνακας 1).

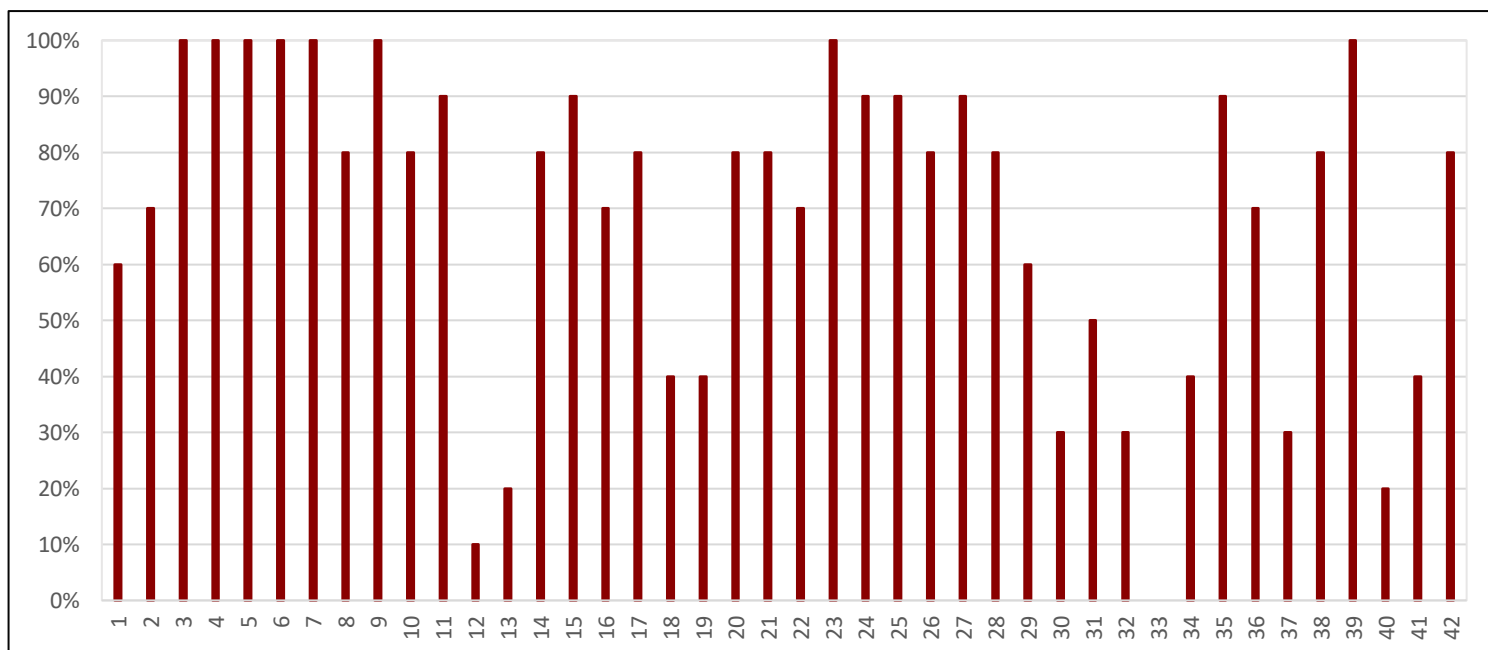
Αποτελέσματα ποιοτικής αξιολόγησης των ερευνητικών άρθρων

Κριτήρια CLAIM		Ερευνητικές Εργασίες										
		#	01	02	03	04	05	06	07	08	09	10
Τίτλος	1	6	0	1	1	1	1	0	1	0	0	1
Περίληψη	2	7	1	1	0	1	1	1	0	1	1	0
Εισαγωγή	3	10	1	1	1	1	1	1	1	1	1	1
	4	10	1	1	1	1	1	1	1	1	1	1
Μεθοδολογία												
Σχεδιασμός μελέτης	5	10	1	1	1	1	1	1	1	1	1	1
	6	10	1	1	1	1	1	1	1	1	1	1
Δεδομένα	7	10	1	1	1	1	1	1	1	1	1	1
	8	8	1	1	0	1	1	0	1	1	1	1
	9	10	1	1	1	1	1	1	1	1	1	1
	10	8	0	1	1	1	1	1	1	1	0	1
	11	9	1	1	1	1	1	0	1	1	1	1
	12	1	0	0	0	0	0	0	0	0	1	0
	13	2	0	0	0	1	0	0	0	1	0	0
Εμπειρικά θεμελιωμένα δεδομένα (ground truth)	14	8	0	1	0	1	1	1	1	1	1	1
	15	9	0	1	1	1	1	1	1	1	1	1
	16	7	1	1	0	0	1	0	1	1	1	1
	17	8	1	1	0	0	1	1	1	1	1	1
	18	4	0	0	0	0	1	1	1	0	1	0
Διαμερισμός δεδομένων	19	4	0	1	0	0	1	0	0	0	1	1
	20	8	0	1	1	1	1	1	1	1	0	1
	21	8	0	1	1	1	1	1	1	1	1	0
Μοντέλο Τεχνητής νοημοσύνης	22	7	0	1	0	1	1	1	1	1	1	0
	23	10	1	1	1	1	1	1	1	1	1	1
	24	9	1	1	1	1	1	1	1	1	0	1
Εκπαίδευση μοντέλου	25	9	1	1	1	1	1	1	1	1	0	1
	26	8	1	1	1	0	1	1	1	1	0	1
	27	9	1	1	1	1	1	1	1	1	0	1
Αξιολόγηση μοντέλου	28	8	1	1	1	0	1	1	0	1	1	1
	29	6	1	1	0	0	1	1	0	1	1	0
	30	3	1	0	0	0	0	1	1	0	0	0
	31	5	0	1	0	0	1	1	1	1	0	0
Αποτελέσματα	32	3	1	1	0	0	0	1	0	0	0	0
Δεδομένα	33	0	0	0	0	0	0	0	0	0	0	0
	34	4	0	1	0	0	1	0	0	0	1	1
Απόδοση μοντέλου	35	9	0	1	1	1	1	1	1	1	1	1
	36	7	0	1	0	0	1	1	1	1	1	1
	37	3	0	0	1	1	1	0	0	0	0	0
Συζήτηση	38	8	0	1	1	1	1	1	1	1	0	1
	39	10	1	1	1	1	1	1	1	1	1	1
Άλλες πληροφορίες	40	2	0	1	0	0	0	0	0	1	0	0
	41	4	0	0	0	0	1	0	0	1	1	1
	42	8	0	1	0	1	1	1	1	1	1	1
Σύνολο CLAIM	42		21	35	22	26	36	30	30	33	27	29

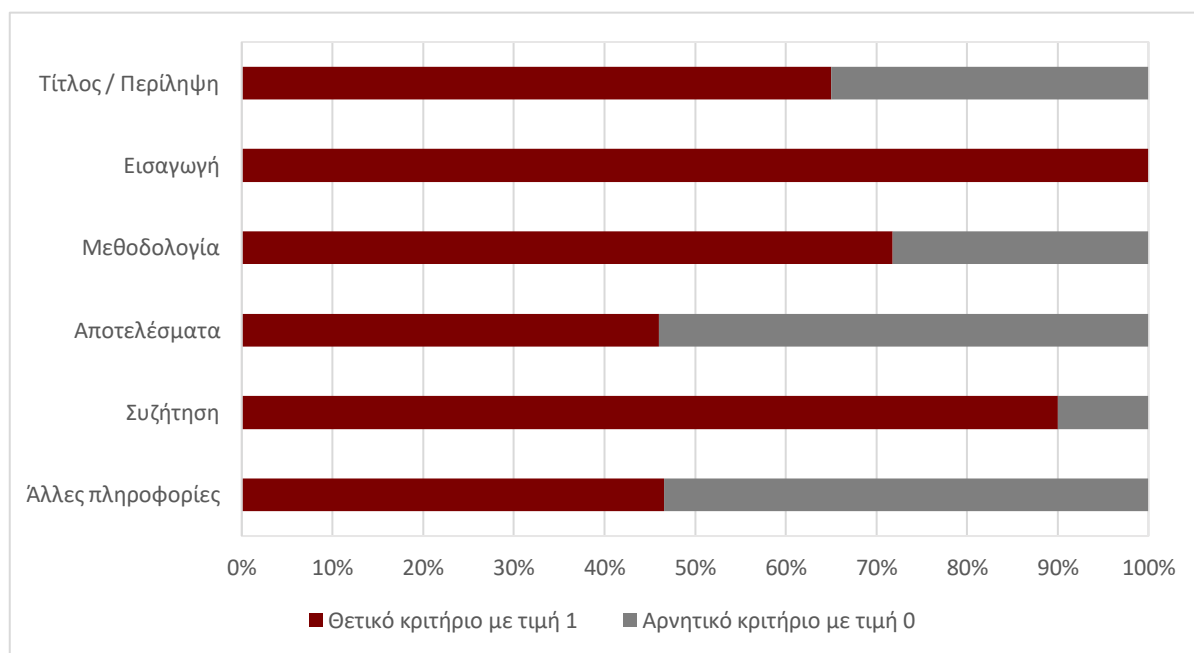
Πίνακας 1: Αποτελέσματα ποιοτικής αξιολόγησης των ερευνητικών εργασιών.

Κριτήρια CLAIM		Συνολικός αριθμός άρθρων	Αριθμός άρθρων με θετικό κριτήριο	Ποσοστιαία αναλογία (%)
Τίτλος και Περίληψη				
1	Τίτλος	10	6	60%
2	Περίληψη	10	7	70%
Εισαγωγή				
3	Εισαγωγή	10	10	100%
4		10	10	100%
Μεθοδολογία				
5	Σχεδιασμός Μελέτης	10	10	100%
6		10	10	100%
7	Δεδομένα	10	10	100%
8		10	8	80%
9		10	10	100%
10		10	8	80%
11		10	9	90%
12		10	1	10%
13		10	2	20%
14	Εμπειρικά θεμελιωμένα δεδομένα (ground truth)	10	8	80%
15		10	9	90%
16		10	7	70%
17		10	8	80%
18		10	4	40%
19	Διαμερισμός δεδομένων	10	4	40%
20		10	8	80%
21		10	8	80%
22	Μοντέλο Τεχνητής νοημοσύνης	10	7	70%
23		10	10	100%
24		10	9	90%
25	Εκπαίδευση μοντέλου	10	9	90%
26		10	8	80%
27		10	9	90%
28	Αξιολόγηση μοντέλου	10	8	80%
29		10	6	60%
30		10	3	30%
31		10	5	50%
32		10	3	30%
Αποτελέσματα				
33	Δεδομένα	10	0	0%
34		10	4	40%
35	Απόδοση μοντέλου	10	9	90%
36		10	7	70%
37		10	3	30%
Συζήτηση				
38	Συζήτηση	10	8	80%
39		10	10	100%
Άλλες πληροφορίες				
40	Άλλες πληροφορίες	10	2	20%
41		10	4	40%
42		10	8	80%

Πίνακας 2: Αποτελέσματα συνολικής ποιοτικής αξιολόγησης των ερευνητικών εργασιών.

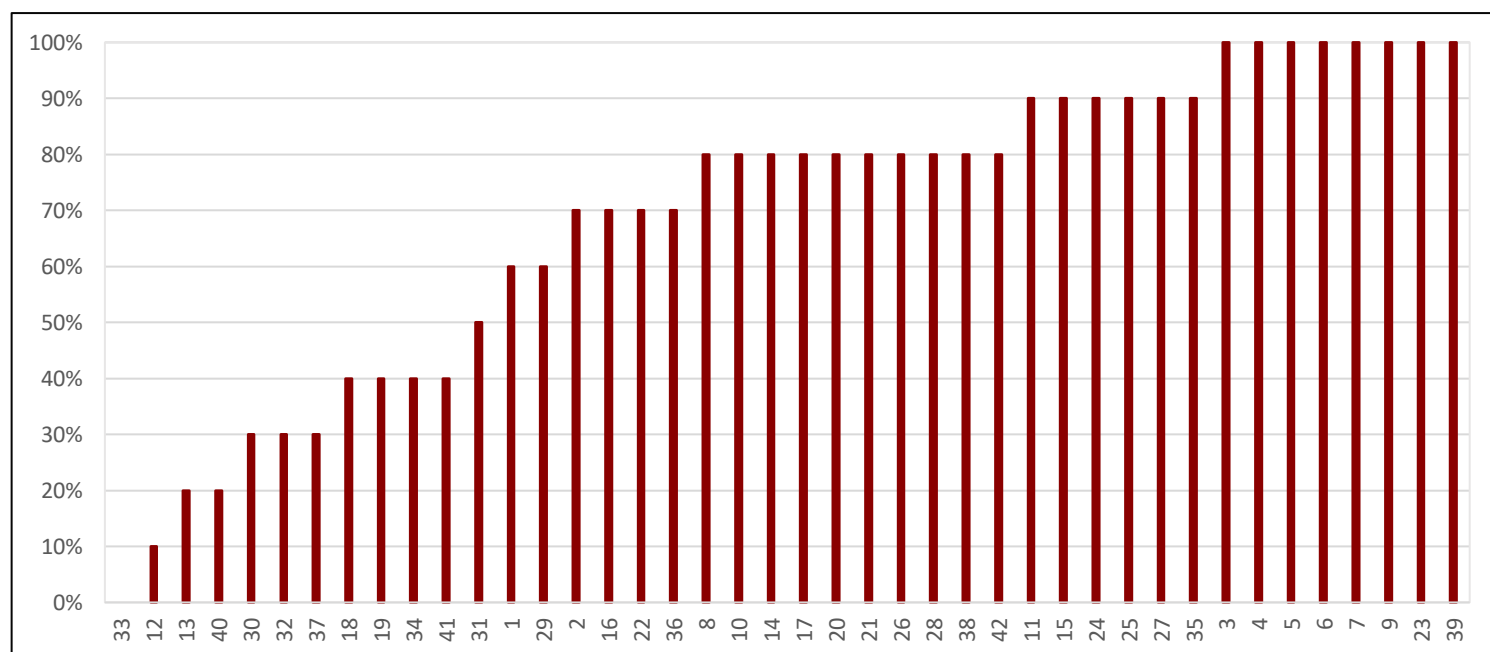


Εικόνα 2: Συνολική αξιολόγηση κάθε κριτηρίου (σε σειρά παρουσίασης στο CLAIM) στο σύνολο των άρθρων.



Εικόνα 3: Συνολική ποσοστιαία αναλογία κριτηρίων συνολικά στα άρθρα ανά τμήμα εργασίας.

Αναλύοντας τα αποτελέσματα του Πίνακα 2, της γραφικής παράστασης της Εικόνας 2 και της Εικόνας 3, παρατηρούμε τα χαμηλότερα ποσοστά και την μεγαλύτερη διακύμανση ποσοστών ανά τμήμα των άρθρων, στον τίτλο / περίληψη (κριτήρια 1-2), στην μεθοδολογία (κριτήρια 12-19 και 29-32), ιδίως στα αποτελέσματα (κριτήρια 33-37), καθώς και τις λοιπές πληροφορίες (40-42).



Εικόνα 4: Συνολική αξιολόγηση κάθε κριτηρίου (κατά αύξουσα ποσοστιαία αναλογία).

Αναλύοντας τα αποτελέσματα της συνολικής αξιολόγησης κάθε κριτηρίου σε αύξουσα ποσοστιαία αναλογία, όπως παρουσιάζονται στην γραφική παράσταση της Εικόνας 4, προκύπτουν οι παρακάτω παρατηρήσεις:

Το κριτήριο 33, το οποίο αφορά στα κριτήρια ένταξης και αποκλεισμού ασθενών, εξετάσεων ασθενών και οποιοδήποτε άλλο στοιχείο των συμμετεχόντων, καθώς και χαρακτηριστικών των βάσεων δεδομένων, δεν εντοπίστηκε σε καμία από τις εξεταζόμενες ερευνητικές εργασίες (0%). Στην πραγματικότητα αρκετές εργασίες περιλάμβαναν τις παραπάνω πληροφορίες, κυρίως με μορφή πινάκων, στο τμήμα της μεθοδολογίας και όχι των αποτελεσμάτων, όπως προτείνεται από το CLAIM checklist.

Πολύ μικρό ποσοστό παρατηρήθηκε επίσης στα κριτήρια 12 και 13, τα οποία αφορούν στην ανωνυμοποίηση και την διαχείριση των δεδομένων και τα οποία τελικά δεν αναλύθηκαν επαρκώς στις ερευνητικές εργασίες (10% και 20% αντίστοιχα).

Ιδιαίτερα χαμηλό ποσοστό εντοπίζεται και στο κριτήριο 40 (20%), το οποίο αποτελεί την ταυτότητα κάθε μελέτης και αντιστοιχεί στον αριθμό μητρώου – πρωτοκόλλου. Μερικές από τις εργασίες που εξετάστηκαν αποτελούν τμήμα ευρύτερων κλινικών ερευνών, στοιχείο το οποίο αυτόματα οδηγεί σε συγκεκριμένο αριθμό μητρώου – πρωτοκόλλου, χωρίς όμως να διευκρινίζεται σε κάθε εργασία ξεχωριστά.

Μικρό ποσοστό παρατηρείται στο κριτήριο 30 σχετικά με την ανθεκτικότητα (robustness) και την ευαισθησία (sensitivity) του κάθε συστήματος, σημαντικά χαρακτηριστικά κάθε μοντέλου (30%). Αντίστοιχο ποσοστό παρατηρείται και στο κριτήριο 32, καθώς δεν διαπιστώθηκαν αναλυτικές πληροφορίες σχετικά με την αξιολόγηση κάθε εξεταζόμενου μοντέλου βαθιάς μάθησης με νέα δεδομένα (external data) (30%). Τέλος, αντίστοιχο ποσοστό παρατηρείται και στο κριτήριο 37 (30%), καθώς στις περισσότερες εργασίες δεν αναλύονται οι πιθανές αιτίες λανθασμένης κατανομής των δεδομένων από τα μοντέλα βαθιάς μάθησης.

Ποσοστό λίγο κάτω από 50%, παρατηρείται στο κριτήριο 18 (40%), το οποίο αφορά στην διαβαθμολογική (interrater) και ενδοβαθμολογική (intrarater) αξιοπιστία (variability) και στην αξιολόγηση πιθανών αποκλίσεων στις εφαρμοζόμενες μεθοδολογίες. Αντίστοιχο ποσοστό εντοπίζεται στο κριτήριο 19, σχετικά με τον τρόπο υπολογισμού του απαιτούμενου μεγέθους δείγματος και στο κριτήριο 34, σχετικά με τα δημογραφικά και κλινικά χαρακτηριστικά των συμμετεχόντων στις έρευνες. Αναφορικά με το μέγεθος του δείγματος που χρησιμοποιήθηκε σε κάθε μελέτη, παρουσιάζεται η δυσκολία που υπάρχει στις περισσότερες εργασίες, δηλαδή να χρησιμοποιηθούν τα δεδομένα ασθενών είτε για λόγους προστασίας δεδομένων είτε για λόγους περιορισμένων πληροφοριών για το εν λόγω εξεταζόμενο δείγμα με μείζονα καταθλιπτική διαταραχή, ενώ ταυτόχρονα, όπως σχολιάστηκε και παραπάνω στο κριτήριο 33, τα δεδομένα αυτά εντοπίζονται συχνότερα στο τμήμα της μεθοδολογίας και όχι στο τμήμα των αποτελεσμάτων. Τέλος, αντίστοιχο χαμηλό ποσοστό ελέγχεται και για το κριτήριο 41 σχετικά με τις πηγές για την ανεύρεση του πλήρους πρωτοκόλλου της μελέτης, καθώς σε αρκετές εργασίες αναφέρεται ότι είναι διαθέσιμο μόνο ύστερα από συνεννόηση με τους αντίστοιχους συγγραφείς της.

Το μοναδικό κριτήριο με ποσοστό 50% αντιστοιχεί στο κριτήριο 31, το οποίο αφορά στην επεξήγηση και ερμηνεία των αποτελεσμάτων του εξεταζόμενου μοντέλου βαθιάς μάθησης στο τμήμα των αποτελεσμάτων. Με άλλα λόγια μόνο οι μισές υπό εξέταση εργασίες παρέχουν τις παραπάνω απαιτούμενες αναλυτικές πληροφορίες για τα διάφορα μοντέλα.

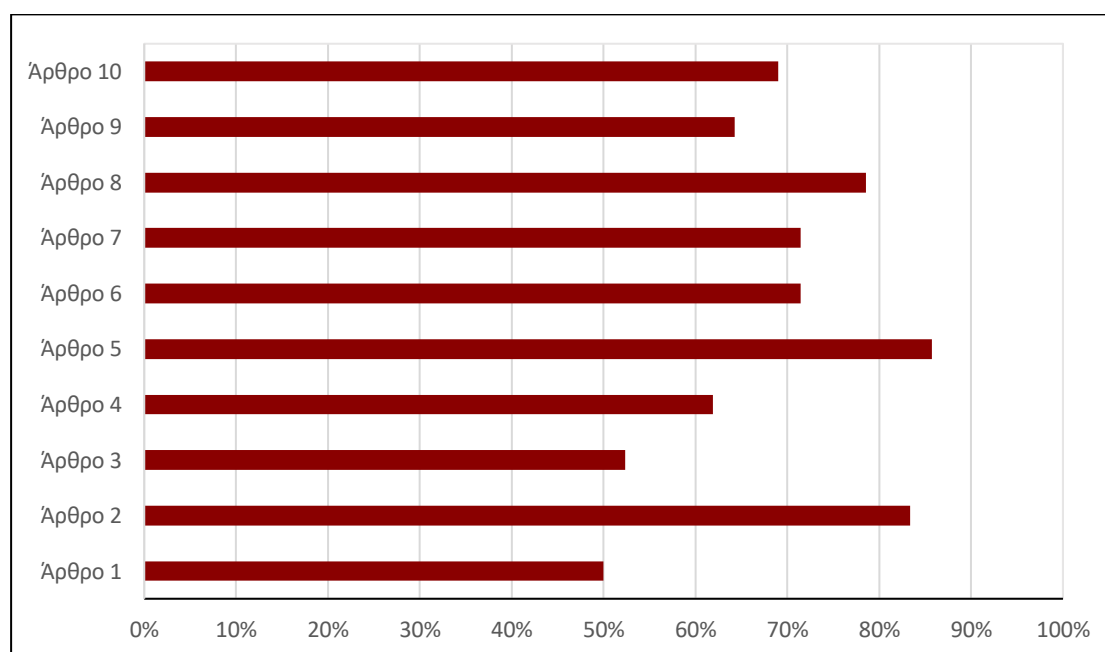
Το κριτήριο 1 που αναφέρεται στην υπόδειξη συγκεκριμένης μεθοδολογίας τεχνητής νοημοσύνης στον τίτλο, παρουσιάζει μέτριο ποσοστό (60%), για την αμεσότητα και την άμεση αποσαφήνιση που μπορεί να προσδώσει ο εκάστοτε τίτλος εργασίας.

Αντίστοιχα μέτριο ποσοστό παρατηρείται και στο κριτήριο 29 (60%), το οποίο σχετίζεται με την στατιστική αξιολόγηση του μοντέλου βαθιάς μάθησης.

Τα κριτήρια 2, 16, 22, 36, 8, 10, 14, 17, 20, 21, 26, 28, 38, 42, 11, 15, 24, 25, 27, 35 παρουσιάζουν ικανοποιητικά ποσοστά (70-90%) ποιοτικής αναφοράς σχετικά με την CLAIM checklist.

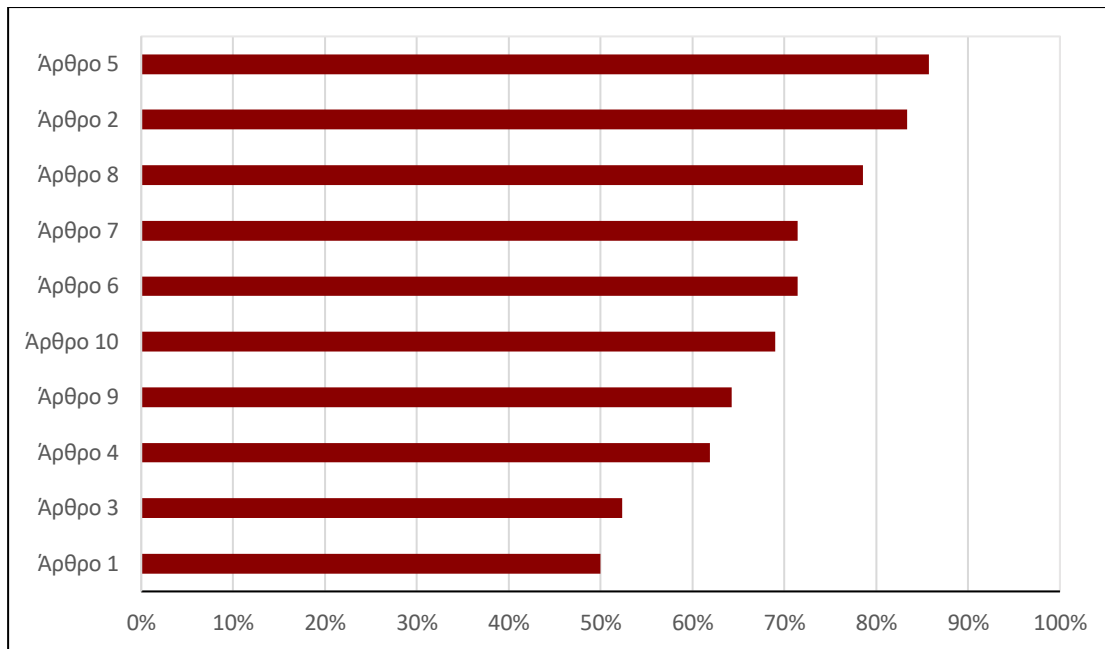
Τα κριτήρια 3, 4, 5, 6, 7, 9, 23, 39 εμφανίζουν το μέγιστο ποσοστό (100%) ποιοτικής αναφοράς σχετικά με την CLAIM checklist.

Συγκρίνοντας συνολικά τα κριτήρια αναφοράς, παρατηρείται ότι ~ 26% των κριτηρίων (11 από τα 42 συνολικά κριτήρια) εμφανίζουν ποσοστό ποιοτικής αναφοράς $\leq 50\%$, ~54% των κριτηρίων (23 από τα 42 συνολικά κριτήρια) εμφανίζουν ποσοστό ποιοτικής αναφοράς $>50\%$ και $<100\%$, ενώ ~ 19% των κριτηρίων παρουσιάζουν ποσοστό ποιοτικής αναφοράς 100% (8 από τα 42 συνολικά κριτήρια).



Εικόνα 5: Συνολική αξιολόγηση κάθε άρθρου με βάση τα κριτήρια CLAIM (σε σειρά).

Με βάση την αξιολόγηση αναφοράς των συνολικών κριτηρίων κάθε άρθρου, όπως παρουσιάζεται στην γραφική παράσταση της Εικόνας 5, είναι εμφανές το γεγονός ότι όλα τα άρθρα εμφανίζουν ποσοστό ποιοτικής αναφοράς $\geq 50\%$.



Εικόνα 6: Συνολική αξιολόγηση κάθε άρθρου με βάση τα κριτήρια CLAIM (σε αύξουσα ποσοστιαία αναλογία).

Πιο συγκεκριμένα, όπως αναδεικνύεται στην γραφική παράσταση της Εικόνας 6, το Άρθρο 1 παρουσιάζει το μικρότερο ποσοστό ποιοτικής αναφοράς κριτηρίων (50%), ενώ το άρθρο 5 παρουσιάζει το μεγαλύτερο ποσοστό ποιοτικής αναφοράς κριτηρίων (~85%).

Τίτλος Άρθρου	Περιοδικό δημοσίευσης	Έτος δημοσίευσης
Άρθρο 1	Springer	2019
Άρθρο 2	Journal of clinical medicine	2020
Άρθρο 3	SPIE proceedings – journal (international society for optics and photonics)	2020
Άρθρο 4	IEEE - Open Journal of Engineering in Medicine and Biology	2020
Άρθρο 5	Nature	2021
Άρθρο 6	IEEE - Open Journal of Engineering in Medicine and Biology	2021
Άρθρο 7	Journal of integrative neuroscience	2021
Άρθρο 8	Biological psychiatry	2022
Άρθρο 9	Contrast media & molecular imaging	2022
Άρθρο 10	Journal of Affective Disorders Reports	2022

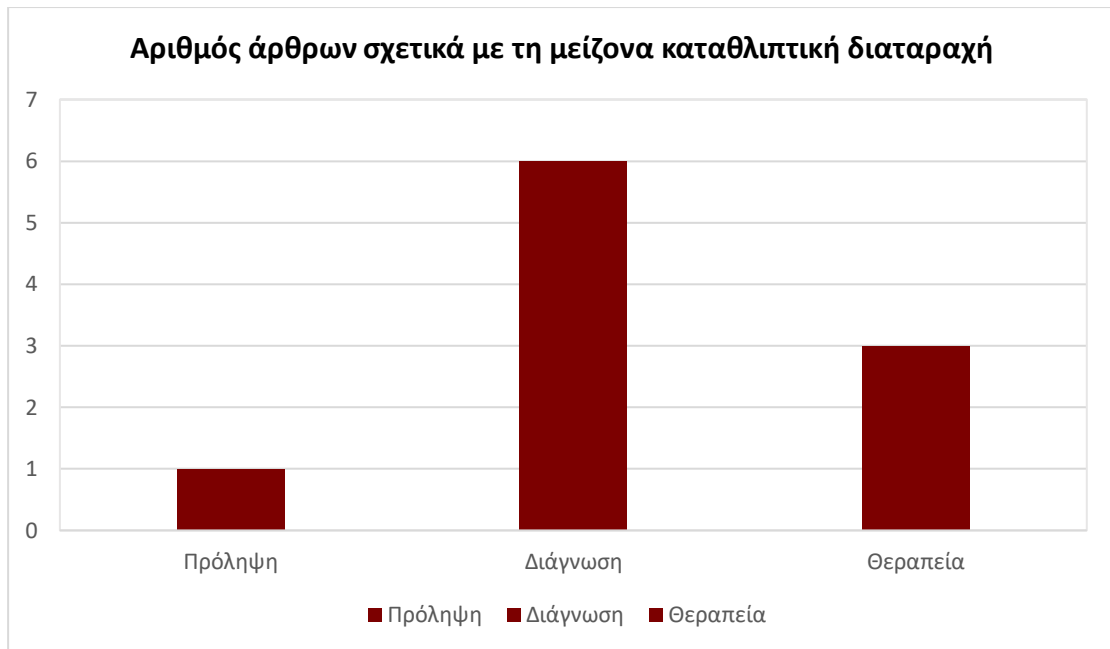
Πίνακας 3: Δεδομένα άρθρων (περιοδικό δημοσίευσης και έτος δημοσίευσης).

Συνδυάζοντας τα αποτελέσματα της γραφικής παράστασης της Εικόνας 5, Εικόνας 6 και τα δεδομένα του Πίνακα 3, παρατηρείται ότι τα άρθρα, τα οποία δημοσιεύθηκαν μετά την εισαγωγή των κριτηρίων CLAIM από το RSNA [62], εμφανίζουν υψηλότερα ποσοστά ποιοτικής αναφοράς κριτηρίων (>60%) (Άρθρο 5, 6, 7, 8, 9, 10), από τα αντίστοιχα πριν τη δημοσίευση των κριτηρίων (Άρθρο 1, 2, 3, 4), χωρίς όμως στατιστικά σημαντική βελτίωση ($p\text{-value} > 0,05$), ενώ η μέση τιμή ποιοτικής αναφοράς κριτηρίων όλων των υπό μελέτη άρθρων είναι 68,8%.

Τα δεδομένα από τον Πίνακα 3 οδηγούν στο συμπέρασμα, ότι τα περιοδικά στα οποία δημοσιεύθηκαν οι εργασίες σχετικά με τις τεχνικές βαθιάς μάθησης στη μείζονα καταθλιπτική διαταραχή είναι ποικίλου επιστημονικού περιεχομένου, δηλαδή περιοδικά κλινικής ιατρικής, περιοδικά μηχανικών εφαρμογών στην ιατρική και τη βιολογία και περιοδικά φυσικών επιστημών.

Τίτλος Άρθρου	Τύπος δικτύου βαθιάς μάθησης	Μείζων καταθλιπτική διαταραχή
Άρθρο 1	Deep Neural Network	Θεραπεία
Άρθρο 2	Convolutional Neural Network – based Autoencoder	Πρόληψη επιπλοκών
Άρθρο 3	Connected Neural Network	Θεραπεία
Άρθρο 4	Convolutional Neural Network	Διάγνωση
Άρθρο 5	Convolutional Neural Network	Διάγνωση
Άρθρο 6	Convolutional Neural Network (BNCPL)	Διάγνωση
Άρθρο 7	Convolutional Neural Network (3D-DenseNet)	Διάγνωση
Άρθρο 8	Connected Neural Network	Θεραπεία
Άρθρο 9	Deep Neural Network	Διάγνωση
Άρθρο 10	Deep Neural Network	Διάγνωση

Πίνακας 4: Δεδομένα άρθρων (τύπος δικτύου βαθιάς μάθησης και πτυχή - στόχος της μείζονας καταθλιπτικής διαταραχής).



Εικόνα 7: Αριθμός άρθρων σχετικά με τη μείζονα καταθλιπτική διαταραχή.

Σύμφωνα με τα δεδομένα του Πίνακα 4 και τη γραφική παράσταση της Εικόνας 7, στις ερευνητικές εργασίες που μελετήθηκαν, χρησιμοποιούνται διάφορες τεχνικές βαθιάς μάθησης για όλες τις πτυχές της μείζονος καταθλιπτικής διαταραχής, με ιδιαίτερη έμφαση σε μελέτες με στόχο την άμεση και έγκαιρη διάγνωσή της.

ΣΥΖΗΤΗΣΗ

Τα κριτήρια ελέγχου CLAIM (Checklist for Artificial Intelligence in Medical Imaging) δημοσιεύθηκαν από την Ακτινολογική Εταιρεία της Βόρειας Αμερικής (RSNA) τον Μάρτιο του 2020, για την αξιολόγηση της ποιότητας αναφοράς μελετών τεχνητής νοημοσύνης στην ιατρική απεικόνιση [62]. Σκοπός της ακτινολογικής εταιρείας, όπως αναφέρεται και στο άρθρο της δημοσίευσης, είναι η παροχή προτεινόμενων οδηγιών, σχετικών με τον τρόπο οργάνωσης και παρουσίασης ερευνητικών εργασιών τεχνητής νοημοσύνης στην ιατρική απεικόνιση, ώστε να συντάσσονται δομημένα όλα τα απαραίτητα στοιχεία κάθε έρευνας, ενώ ταυτόχρονα να παρέχεται η δυνατότητα σε κάθε ερευνητή να εξερευνήσει σε σειρά τα στοιχεία της μελέτης, να τα αξιολογήσει, καθώς και να επαληθεύσει τα αποτελέσματά της. Ιδίως η σαφής καταγραφή των χαρακτηριστικών των μοντέλων τεχνητής νοημοσύνης (μηχανικής και βαθιάς μάθησης), δύναται να συμβάλλει στο μέγιστο βαθμό στην αναπαραγωγιμότητα (reproducibility) και την επαναληψιμότητα (repeatability) των ερευνητικών εργασιών [73]. Στη βιβλιογραφία υπάρχουν αντίστοιχες εργασίες με αξιολόγηση ερευνών με τα κριτήρια CLAIM, οι οποίες όμως δεν αφορούν σε ψυχιατρικά νοσήματα [74]. Όπως προέκυψε και από τα αποτελέσματα της παρούσας μελέτης, η ολοκληρωμένη και ακριβής αναφορά των ερευνητικών δεδομένων ποικίλει από άρθρο σε άρθρο.

Αναλύοντας τα κριτήρια στις αποδεκτές έρευνες της μελέτης [63-72], προκύπτει ότι η μέση τιμή των ποσοστών της ποιοτικής αναφοράς των κριτηρίων όλων των υπό μελέτη άρθρων είναι 68,8%, δηλαδή ένα ποσοστό παρόμοιο με αντίστοιχες πρόσφατες έρευνες CLAIM κριτηρίων [75]. Επίσης, είναι εμφανές ότι τα χαμηλότερα ποσοστά και η μεγαλύτερη διακύμανση ποσοστών ανά τμήμα των άρθρων, εντοπίζονται στο τμήμα της μεθοδολογίας (κριτήρια 12-19 και 29-32) και ιδίως στο τμήμα των αποτελεσμάτων (κριτήρια 33-37). Το εύρημα αυτό αναγνωρίζεται και σε άλλες αντίστοιχες εργασίες με αξιολόγησή τους με την CLAIM checklist, με αποτέλεσμα να προκύπτει η ανάγκη για βελτίωση της καταγραφής των χαρακτηριστικών - δεδομένων του μοντέλου βαθιάς μάθησης (πιο συγκεκριμένα για τα δεδομένα ground truth και τα στοιχεία της αξιολόγησης του μοντέλου - σημαντικά δεδομένα για την αναπαραγωγιμότητα της εργασίας), καθώς και των αποτελεσμάτων (παρουσίαση και ανάλυση των αρχικών δεδομένων των μοντέλων) [75-78].

Αντίστοιχα χαμηλά ποσοστά ποιότητας αναφοράς κριτηρίων εντοπίζονται στον τίτλο/περίληψη (κριτήρια 1-2) και τις λοιπές πληροφορίες (40-42) παρόμοιων εργασιών[75]. Ο τίτλος αποτελεί το πρώτο σημείο ανάγνωσης μιας εργασίας και είναι απαραίτητο να παρέχει τις κατάλληλες και επαρκείς πληροφορίες στον αναγνώστη, όπως αντίστοιχα και η περίληψη. Επιπλέον, οι πληροφορίες σχετικά με τις πηγές για την ανεύρεση του πλήρους πρωτοκόλλου της μελέτης παρουσιάζονται με χαμηλά ποσοστά αναφοράς, καθώς σε αρκετές εργασίες αναφέρεται ότι είναι διαθέσιμες μόνο ύστερα από συνεννόηση με τους αντίστοιχους συγγραφείς της. Σταδιακά, όπως παρατηρήθηκε στις πιο πρόσφατες αποδεκτές μελέτες, όλο και περισσότεροι συγγραφείς προσθέτουν παραρτήματα στα άρθρα τους για πλήρη πρόσβαση στα πρωτόκολλα των ερευνών τους (Πίνακας 1) [70-72].

Σχετικά με τα κριτήρια ένταξης και αποκλεισμού ασθενών, εξετάσεων ασθενών και οποιοδήποτε άλλο στοιχείο των συμμετεχόντων, καθώς και χαρακτηριστικών των βάσεων δεδομένων (κριτήριο 33), δεν εντοπίστηκε σε καμία από τις εξεταζόμενες ερευνητικές εργασίες (0%). Στην πραγματικότητα αρκετές εργασίες περιλαμβάνουν τις παραπάνω πληροφορίες, κυρίως με μορφή πινάκων, στο τμήμα όμως της μεθοδολογίας και όχι των αποτελεσμάτων, όπως προτείνεται από το CLAIM checklist. Πολύ μικρό ποσοστό παρατηρήθηκε επίσης στα, τα οποία αφορούν στην ανωνυμοποίηση και την διαχείριση των δεδομένων (κριτήριο 12 και 13) τα οποία δεν χρησιμοποιήθηκαν τελικά στις εργασίες αντίστοιχα (10% και 20% αντίστοιχα). Τα παραπάνω ευρήματα είναι υπαρκτά και σε άλλα αντίστοιχα άρθρα [79].

Επιπρόσθετα, τα κριτήρια σχετικά με τις υποομάδες δεδομένων (κριτήριο 10) και τις πρόσθετες τεχνικές τεχνητής νοημοσύνης (κριτήριο 27 και 31), δεν είναι απαραίτητο να αναφέρονται – μελετώνται στις διάφορες εργασίες, (στα πλαίσια διαφορετικής μεθοδολογίας κάθε εργασίας). Παρόλα αυτά υπολογίστηκαν στο τελικό άθροισμα των κριτηρίων στην στατιστική ανάλυση με τιμή 0, ακόμα και σε περίπτωση που οι εργασίες δεν περιλάμβαναν τέτοιου είδους πληροφορίες.

Συνδυάζοντας τα αποτελέσματα της γραφικής παράστασης της Εικόνας 5, της Εικόνας 6 και τα δεδομένα του Πίνακα 3, παρατηρείται ότι τα άρθρα, τα οποία δημοσιεύθηκαν μετά την εισαγωγή των κριτηρίων CLAIM από το RSNA [57], εμφανίζουν υψηλότερα ποσοστά ποιοτικής αναφοράς κριτηρίων >60% (Άρθρο 5, 6, 7, 8, 9, 10), από τα

αντίστοιχα πριν τη δημοσίευση των κριτηρίων (Άρθρο 1, 2 ,3 ,4), ένα εύρημα για το οποίο απαιτούνται περισσότερες μελέτες για τη στατιστική σημαντικότητά του.

Τα δεδομένα από τον Πίνακα 3 οδηγούν στο συμπέρασμα ότι τα περιοδικά στα οποία δημοσιεύθηκαν οι εργασίες σχετικά με τις τεχνικές βαθιάς μάθησης στη μείζονα καταθλιπτική διαταραχή είναι ποικίλου επιστημονικού περιεχομένου, δηλαδή περιοδικά κλινικής ιατρικής, περιοδικά μηχανικών εφαρμογών στην ιατρική και τη βιολογία και περιοδικά φυσικών επιστημών, εύρημα το οποίο εντοπίζεται και σε άλλες αντίστοιχες μελέτες [75].

Συνολικά, εξετάζοντας τις συνημμένες οδηγίες της μελέτης - αναφοράς των CLAIM κριτηρίων, είναι ιδιαίτερα εμφανής η πληθώρα στοιχείων που παρουσιάζονται σε κάθε κριτήριο [62]. Πολλά κριτήρια αξιολογούνται με αρνητικό κριτήριο και τιμή 0, ακόμα και αν περιγράφουν μεγάλο τμήμα των προτεινόμενων στοιχείων από την Ακτινολογική Εταιρεία της Βόρειας Αμερικής (RSNA), με αποτέλεσμα να δίνεται η εντύπωση ότι δεν αναφέρονται αρκετές πληροφορίες σχετικά με το κάθε τμήμα του άρθρου.

Η ανεπαρκής καταγραφή πληροφοριών στο σύνολό της, δυσχεραίνει την επαναληψιμότητα και την αναπαραγωγιμότητα ενός μοντέλου τεχνητής νοημοσύνης από έναν ερευνητή, καθώς και την αναζήτηση των άρθρων (π.χ. ανεπαρκείς πληροφορίες στον τίτλο και την περίληψη) σε μελέτες συστηματικής ανασκόπησης και μεταanalύσεις.

Η παρούσα μελέτη αποτελεί την πρώτη καταγεγραμμένη (βιβλιογραφικά) ερευνητική εργασία αξιολόγησης της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την κατάθλιψη με τη χρήση των κριτηρίων ελέγχου CLAIM και χρειάζονται περαιτέρω μελέτες για την επαλήθευση και αξιολόγηση των αποτελεσμάτων της.

ΠΕΡΙΟΡΙΣΜΟΙ

Η παρούσα μελέτη διενεργήθηκε μόνο από έναν ερευνητή (τον συγγραφέα της μελέτης). Η αναζήτηση των ερευνητικών άρθρων πραγματοποιήθηκε μόνο στην ηλεκτρονική βάση δεδομένων MEDLINE (via PubMed) και χρησιμοποιήθηκαν άρθρα, τα οποία είναι γραμμένα μόνο στην αγγλική γλώσσα. Ακόμα, ο συνολικός αριθμός των

αποδεκτών άρθρων που αναλύθηκε είναι περιορισμένος, με επακόλουθο τα αποτελέσματα που προκύπτουν να παρουσιάζουν χαμηλή στατιστική σημαντικότητα.

Η παρούσα εργασία αφορά σε συγκεκριμένες τεχνικές τεχνητής νοημοσύνης, δηλαδή μόνο σε τεχνικές βαθιάς μάθησης και μόνο σε εργασίες σχετικές με την μείζονα καταθλιπτική διαταραχή, με αποτέλεσμα να είναι ολιγάριθμα τα τελικά υπό εξέταση άρθρα και να απαιτούνται περισσότερα δεδομένα για την γενίκευση των αποτελεσμάτων.

Επίσης, η αξιολόγηση κάθε κριτηρίου αποτελεί μία σύνθετη διαδικασία, καθώς οι συνημμένες οδηγίες της μελέτης αναφοράς των CLAIM κριτηρίων που παρουσιάζονται σε κάθε κριτήριο, παρουσιάζουν πληθώρα στοιχείων, με αποτέλεσμα πολλά κριτήρια να αξιολογούνται με αρνητικό κριτήριο και τιμή 0, ακόμα και αν περιγράφουν μεγάλο τμήμα των προτεινόμενων στοιχείων. Τα παραπάνω δεδομένα υπολογίστηκαν και αναλύθηκαν εκτενώς, ώστε να περιοριστούν στο μέγιστο δυνατό πιθανά συστηματικά σφάλματα (selection and measurement bias).

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ύπαρξη κριτηρίων ελέγχου CLAIM (Checklist for Artificial Intelligence in Medical Imaging) είναι ουσιώδης και απαραίτητη για την ορθή παρουσίαση μίας επιστημονικής εργασίας που βασίζεται σε τεχνικές βαθιάς μάθησης, παρέχοντας ολοκληρωμένη και ακριβή αναφορά ερευνητικών στοιχείων. Με άλλα λόγια, η ύπαρξη τέτοιων οδηγιών προσφέρει μεθοδικότητα όχι μόνο στη συγγραφή των εργασιών (διαφάνεια μελέτης), αλλά και στην ανάγνωση και αξιολόγηση αυτών των τεχνικών μελετών, με απώτερο σκοπό τη βελτίωση και ταυτόχρονα την κριτική αξιολόγηση των άρθρων. Με την παρούσα μελέτη παρουσιάζονται μεγάλων διακυμάνσεων ποσοστά στην αξιολόγηση της ποιότητας αναφοράς μελετών βαθιάς μάθησης για την κατάθλιψη με τη χρήση των κριτηρίων ελέγχου CLAIM και χρειάζονται περαιτέρω μελέτες για την επαλήθευση, την αξιολόγηση και την γενίκευση των αποτελεσμάτων της.

ΔΗΛΩΣΗ ΣΥΓΚΡΟΥΣΗΣ ΣΥΜΦΕΡΟΝΤΩΝ

Δεν υπάρχει σύγκρουση συμφερόντων σε σχέση με τη παρούσα μελέτη.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. World Health Organization. (2017). Depression and other common mental disorders: global health estimates. World Health Organization. WHO/MSD/MER/2017.2
2. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022 Feb;9(2):137-150. doi: 10.1016/S2215-0366(21)00395-3. Epub 2022 Jan 10. PMID: 35026139; PMCID: PMC8776563.
3. Kendler KS, Gardner CO. Sex differences in the pathways to major depression: a study of opposite-sex twin pairs. *Am J Psychiatry*. 2014 Apr;171(4):426-35. doi: 10.1176/appi.ajp.2013.13101375. PMID: 24525762; PMCID: PMC3972260.
4. Shorey S, Ng ED, Wong CHJ. Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *Br J Clin Psychol*. 2022 Jun;61(2):287-305. doi: 10.1111/bjc.12333. Epub 2021 Sep 26. PMID: 34569066.
5. Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith DJ, Thapar AK, Thapar A. Adolescent and adult differences in major depression symptom profiles. *J Affect Disord*. 2019 Jan 15;243:175-181. doi: 10.1016/j.jad.2018.09.015. Epub 2018 Sep 11. PMID: 30243197.
6. Sousa S, Paúl C, Teixeira L. Predictors of Major Depressive Disorder in Older People. *Int J Environ Res Public Health*. 2021 Nov 12;18(22):11894. doi: 10.3390/ijerph182211894. PMID: 34831649; PMCID: PMC8624368.
7. Shadrina M, Bondarenko EA, Slominsky PA. Genetics Factors in Major Depression Disease. *Front Psychiatry*. 2018 Jul 23;9:334. doi: 10.3389/fpsy.2018.00334. PMID: 30083112; PMCID: PMC6065213.
8. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, Grant BF. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry*. 2018 Apr 1;75(4):336-346. doi: 10.1001/jamapsychiatry.2017.4602. PMID: 29450462; PMCID: PMC5875313.
9. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to

- the COVID-19 pandemic. *Lancet*. 2021 Nov 6;398(10312):1700-1712. doi: 10.1016/S0140-6736(21)02143-7. Epub 2021 Oct 8. PMID: 34634250; PMCID: PMC8500697.
10. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. *World Psychiatry*. 2013 Jun;12(2):92-8. doi: 10.1002/wps.20050. PMID: 23737408; PMCID: PMC3683251.
 11. Uher R, Payne JL, Pavlova B, Perlis RH. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. *Depress Anxiety*. 2014 Jun;31(6):459-71. doi: 10.1002/da.22217. Epub 2013 Nov 22. PMID: 24272961.
 12. McCarron RM, Shapiro B, Rawles J, Luo J. Depression. *Ann Intern Med*. 2021 May;174(5):ITC65-ITC80. doi: 10.7326/AITC202105180. Epub 2021 May 11. PMID: 33971098.
 13. Paris J. The mistreatment of major depressive disorder. *Can J Psychiatry*. 2014 Mar;59(3):148-51. doi: 10.1177/070674371405900306. PMID: 24881163; PMCID: PMC4079242.
 14. Health Quality Ontario. Repetitive Transcranial Magnetic Stimulation for Treatment-Resistant Depression: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Ont Health Technol Assess Ser*. 2016 Mar 1;16(5):1-66. PMID: 27099642; PMCID: PMC4808719.
 15. Perera T, George MS, Grammer G, Janicak PG, Pascual-Leone A, Wirecki TS. The Clinical TMS Society Consensus Review and Treatment Recommendations for TMS Therapy for Major Depressive Disorder. *Brain Stimul*. 2016 May-Jun;9(3):336-346. doi: 10.1016/j.brs.2016.03.010. Epub 2016 Mar 16. PMID: 27090022; PMCID: PMC5612370.
 16. Trivedi MH. Major Depressive Disorder in Primary Care: Strategies for Identification. *J Clin Psychiatry*. 2020 Mar 17;81(2):UT17042BR1C. doi: 10.4088/JCP.UT17042BR1C. PMID: 32220155.
 17. Rafeyan R, Papakostas GI, Jackson WC, Trivedi MH. Inadequate Response to Treatment in Major Depressive Disorder: Augmentation and Adjunctive Strategies. *J Clin Psychiatry*. 2020 May 12;81(3):OT19037BR3. doi: 10.4088/JCP.OT19037BR3. PMID: 32412697.
 18. Squires M, Tao X, Elangovan S, Gururajan R, Zhou X, Acharya UR, Li Y. Deep learning and machine learning in psychiatry: a survey of current progress in

- depression detection, diagnosis and treatment. *Brain Inform.* 2023 Apr 24;10(1):10. doi: 10.1186/s40708-023-00188-6. PMID: 37093301; PMCID: PMC10123592.
19. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism.* 2017 Apr;69S:S36-S40. doi: 10.1016/j.metabol.2017.01.011. Epub 2017 Jan 11. PMID: 28126242.
 20. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial Intelligence in Medicine: Where Are We Now? *Acad Radiol.* 2020 Jan;27(1):62-70. doi: 10.1016/j.acra.2019.10.001. Epub 2019 Oct 19. PMID: 31636002.
 21. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep.* 2019 Nov 7;21(11):116. doi: 10.1007/s11920-019-1094-0. PMID: 31701320; PMCID: PMC7274446.
 22. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Healthcare (Basel).* 2023 Jan 17;11(3):285. doi: 10.3390/healthcare11030285. PMID: 36766860; PMCID: PMC9914523.
 23. Turing MA. Computing machinery and intelligence. *Mind.* 1950. 59 (236):433.
 24. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955 (2006). *AI Magazine*, 27(4), 12.
 25. Haenlein, M., & Kaplan, A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review.* 2019. 61(4), 5–14.
 26. Adrienne Mayor - *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology.* Princeton University Press. 2018.
 27. Coeckelbergh M. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Sci Eng Ethics.* 2020 Aug;26(4):2051-2068. doi: 10.1007/s11948-019-00146-8. Epub 2019 Oct 24. PMID: 31650511; PMCID: PMC7417397.
 28. Palanca-Castan N, Sánchez Tajadura B, Cofré R. Towards an interdisciplinary framework about intelligence. *Heliyon.* 2021 Feb 17;7(2):e06268. doi: 10.1016/j.heliyon.2021.e06268. PMID: 33665435; PMCID: PMC7902546.

29. Kovacs Kristof, Conway Andrew R.A. Process overlap theory: a unified account of the general factor of intelligence. *Psychol. Inq.* 2016
30. Knight Rex, Piaget Jean, Piercy M., Berlyne D.E. The psychology of intelligence. *Philos. Q.* 1951
31. Flynn James R. Cambridge University Press; 2007. What Is Intelligence?: Beyond the Flynn Effect.
32. Heuer Richards J. Center for the Study of Intelligence; 1999. Psychology of Intelligence Analysis.
33. Hampshire Adam, Highfield Roger R., Parkin Beth L., Owen Adrian M. Fractionating human intelligence. *Neuron.* 2012
34. Haier Richard J. Cambridge University Press; 2016. The Neuroscience of Intelligence.
35. Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. Machine Learning: A Historical and Methodological Analysis. *AI Magazine.* 1983. 4(3), 69.
36. Fradkov LA. Early History of Machine Learning. *IFAC- PapersOnLine.* 2020. 53(2) 1385-1390
37. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci.* 2021;2(3):160. doi: 10.1007/s42979-021-00592-x. Epub 2021 Mar 22. PMID: 33778771; PMCID: PMC7983091.
38. Wang, H., & Raj, B. On the Origin of Deep Learning On the Origin of Deep Learning. 2017. arXiv:1702.07800v4.
39. Lukanov H, König P, Pipa G. Biologically Inspired Deep Learning Model for Efficient Foveal-Peripheral Vision. *Front Comput Neurosci.* 2021 Nov 22;15:746204. doi: 10.3389/fncom.2021.746204. PMID: 34880741; PMCID: PMC8645638.
40. Najafabadi, MM. Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2015. 2: 1-21.
41. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol.* 2020;1213:3-21. doi: 10.1007/978-3-030-33128-3_1. PMID: 32030660; PMCID: PMC7442218.
42. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015).
43. Sarker IH. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Comput Sci.*

- 2021;2(5):377. doi: 10.1007/s42979-021-00765-8. Epub 2021 Jul 12. PMID: 34278328; PMCID: PMC8274472.
44. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput Sci.* 2021;2(6):420. doi: 10.1007/s42979-021-00815-1. Epub 2021 Aug 18. PMID: 34426802; PMCID: PMC8372231.
 45. Servan-Schreiber D. Artificial intelligence and psychiatry. *J Nerv Ment Dis.* 1986 Apr;174(4):191-202. doi: 10.1097/00005053-198604000-00001. PMID: 3514790.
 46. Haslam N, Beck AT. Categorization of major depression in an outpatient sample. *J Nerv Ment Dis.* 1993 Dec;181(12):725-31. doi: 10.1097/00005053-199312000-00003. PMID: 8254323.
 47. Thoduparambil PP, Dominic A, Varghese SM. EEG-based deep learning model for the automatic detection of clinical depression. *Phys Eng Sci Med.* 2020 Dec;43(4):1349-1360. doi: 10.1007/s13246-020-00938-4. Epub 2020 Oct 22. PMID: 33090373.
 48. Gao S, Calhoun VD, Sui J. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci Ther.* 2018 Nov;24(11):1037-1052. doi: 10.1111/cns.13048. Epub 2018 Aug 23. PMID: 30136381; PMCID: PMC6324186.
 49. Arundel C, James S, Northgraves M, Booth A. Study reporting guidelines: How valid are they? *Contemp Clin Trials Commun.* 2019 Mar 11;14:100343. doi: 10.1016/j.conctc.2019.100343. PMID: 30923775; PMCID: PMC6421355.
 50. Catalá-López F, Alonso-Arroyo A, Page MJ, Hutton B, Ridao M, Tabarés-Seisdedos R, Aleixandre-Benavent R, Moher D. Reporting guidelines for health research: protocol for a cross-sectional analysis of the EQUATOR Network Library. *BMJ Open.* 2019 Mar 4;9(3):e022769. doi: 10.1136/bmjopen-2018-022769. PMID: 30837245; PMCID: PMC6429992.
 51. Kouli O, Hassane A, Badran D, Kouli T, Hossain-Ibrahim K, Steele JD. Automated brain tumor identification using magnetic resonance imaging: A systematic review and meta-analysis. *Neurooncol Adv.* 2022 May 27;4(1):vdac081. doi: 10.1093/noajnl/vdac081. PMID: 35769411; PMCID: PMC9234754.

52. Karabacak M, Ozkara BB, Mordag S, Bisdas S. Deep learning for prediction of isocitrate dehydrogenase mutation in gliomas: a critical approach, systematic review and meta- analysis of the diagnostic test performance using a Bayesian approach. *Quant Imaging Med Surg.* 2022;12(8):4033-4046.
53. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain metastasis detection using machine learning: a systematic review and meta-analysis. *Neuro Oncol.* 2021;23(2):214-225.
54. Farook TH, Jamayet NB, Abdullah JY, Alam MK. Machine Learning and Intelligent Diagnostics in Dental and Orofacial Pain Management: A Systematic Review. *Pain Res Manag.* 2021 Apr 24;2021:6659133. doi: 10.1155/2021/6659133. PMID: 33986900; PMCID: PMC8093041.
55. Alabed S, Maiter A, Salehi M, Mahmood A, Daniel S, Jenkins S, Goodlad M, Sharkey M, Mamalakis M, Rakocevic V, Dwivedi K, Assadi H, Wild JM, Lu H, O'Regan DP, van der Geest RJ, Garg P, Swift AJ. Quality of reporting in AI cardiac MRI segmentation studies - A systematic review and recommendations for future studies. *Front Cardiovasc Med.* 2022 Jul 15;9:956811. doi: 10.3389/fcvm.2022.956811. PMID: 35911553; PMCID: PMC9334661.
56. Belue MJ, Turkbey B. Tasks for artificial intelligence in prostate MRI. *Eur Radiol Exp.* 2022 Jul 31;6(1):33. doi: 10.1186/s41747-022-00287-9. PMID: 35908102; PMCID: PMC9339059.
57. Lans A, Pierik RJB, Bales JR, Fourman MS, Shin D, Kanbier LN, Rifkin J, DiGiovanni WH, Chopra RR, Moeinzad R, Verlaan JJ, Schwab JH. Quality assessment of machine learning models for diagnostic imaging in orthopaedics: A systematic review. *Artif Intell Med.* 2022 Oct;132:102396. doi: 10.1016/j.artmed.2022.102396. Epub 2022 Sep 6. PMID: 36207080.
58. Bleker J, Kwee TC, Yakar D. Quality of Multicenter Studies Using MRI Radiomics for Diagnosing Clinically Significant Prostate Cancer: A Systematic Review. *Life (Basel).* 2022 Jun 23;12(7):946. doi: 10.3390/life12070946. PMID: 35888036; PMCID: PMC9324573.
59. Sushentsev N, Moreira Da Silva N, Yeung M, Barrett T, Sala E, Roberts M, Rundo L. Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. *Insights Imaging.* 2022 Mar 28;13(1):59. doi: 10.1186/s13244-022-01199-3. PMID: 35347462; PMCID: PMC8960511.

60. Zhong J, Hu Y, Zhang G, Xing Y, Ding D, Ge X, Pan Z, Yang Q, Yin Q, Zhang H, Zhang H, Yao W. An updated systematic review of radiomics in osteosarcoma: utilizing CLAIM to adapt the increasing trend of deep learning application in radiomics. *Insights Imaging*. 2022 Aug 20;13(1):138. doi: 10.1186/s13244-022-01277-6. PMID: 35986808; PMCID: PMC9392674.
61. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, Arnaout R, Kohane IS, Saria S, Topol E, Obermeyer Z, Yu B, Butte AJ. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med*. 2020 Sep;26(9):1320-1324. doi: 10.1038/s41591-020-1041-y. PMID: 32908275; PMCID: PMC7538196.
62. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell*. 2020 Mar 25;2(2):e200029. doi: 10.1148/ryai.2020200029. PMID: 33937821; PMCID: PMC8017414.
63. Nguyen KP, Fatt CC, Treacher A, Mellema C, Trivedi MH, Montillo A. Predicting Response to the Antidepressant Bupropion using Pretreatment fMRI. *Predict Intell Med*. 2019 Oct;11843:53-62. doi: 10.1007/978-3-030-32281-6_6. Epub 2019 Oct 10. PMID: 31709423; PMCID: PMC6839715.
64. Weng JC, Lin TY, Tsai YH, Cheok MT, Chang YE, Chen VC. An Autoencoder and Machine Learning Model to Predict Suicidal Ideation with Brain Structural Imaging. *J Clin Med*. 2020 Feb 29;9(3):658. doi: 10.3390/jcm9030658. PMID: 32121362; PMCID: PMC7141277.
65. Nguyen KP, Fatt CC, Treacher A, Mellema C, Trivedi MH, Montillo A. Anatomically-Informed Data Augmentation for Functional MRI with Applications to Deep Learning. *Proc SPIE Int Soc Opt Eng*. 2020 Feb;11313:113130T. doi: 10.1117/12.2548630. Epub 2020 Mar 10. PMID: 33767520; PMCID: PMC7990266.
66. Chen GY, Huang CM, Liu HL, Lee SH, Lee TM, Lin C, Wu SC. Depression Scale Prediction with Cross-Sample Entropy and Deep Learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020 Jul;2020:120-123. doi: 10.1109/EMBC44109.2020.9175816. PMID: 33017945.
67. Lim SH, Yoon J, Kim YJ, Kang CK, Cho SE, Kim KG, Kang SG. Reproducibility of automated habenula segmentation via deep learning in major depressive disorder and normal controls with 7 Tesla MRI. *Sci Rep*. 2021 Jun

- 29;11(1):13445. doi: 10.1038/s41598-021-92952-z. PMID: 34188141; PMCID: PMC8241874.
68. Zhi D, Calhoun VD, Wang C, Li X, Ma X, Lv L, Yan W, Yao D, Qi S, Jiang R, Zhao J, Yang X, Lin Z, Zhang Y, Chung YC, Zhuo C, Sui J. BNCPL: Brain-Network-based Convolutional Prototype Learning for Discriminating Depressive Disorders. *Annu Int Conf IEEE Eng Med Biol Soc.* 2021 Nov;2021:1622-1626. doi: 10.1109/EMBC46164.2021.9630010. PMID: 34891596; PMCID: PMC9021005.
 69. Wang Y, Gong N, Fu C. Major depression disorder diagnosis and analysis based on structural magnetic resonance imaging and deep learning. *J Integr Neurosci.* 2021 Dec 30;20(4):977-984. doi: 10.31083/j.jin2004098. PMID: 34997720.
 70. Nguyen KP, Chin Fatt C, Treacher A, Mellema C, Cooper C, Jha MK, Kurian B, Fava M, McGrath PJ, Weissman M, Phillips ML, Trivedi MH, Montillo AA. Patterns of Pretreatment Reward Task Brain Activation Predict Individual Antidepressant Response: Key Results From the EMBARC Randomized Clinical Trial. *Biol Psychiatry.* 2022 Mar 15;91(6):550-560. doi: 10.1016/j.biopsych.2021.09.011. Epub 2021 Sep 22. PMID: 34916068; PMCID: PMC8857018.
 71. Cheng Y, Liao T, Jia N. Classification Algorithms for Brain Magnetic Resonance Imaging Images of Patients with End-Stage Renal Disease and Depression. *Contrast Media Mol Imaging.* 2022 Jul 6;2022:4795307. doi: 10.1155/2022/4795307. PMID: 35854766; PMCID: PMC9279039.
 72. Liu Z, Nichol M.L. Wong, Robin Shao, Shwu-Hua Lee, Chih-Mao Huang, Ho-Ling Liu, Chemin Lin, Tatia M.C. Lee. Classification of Major Depressive Disorder using Machine Learning on brain structure and functional connectivity. *J Affect Disord.* 2022 Dec (10):100428;ISSN 2666-9153.
 73. Plesser HE. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Front Neuroinform.* 2018 Jan 18;11:76. doi: 10.3389/fninf.2017.00076. PMID: 29403370; PMCID: PMC5778115.
 74. Sivanesan U, Wu K, McInnes MDF, Dhindsa K, Salehi F, van der Pol CB. Checklist for Artificial Intelligence in Medical Imaging Reporting Adherence in Peer-Reviewed and Preprint Manuscripts With the Highest Altmetric Attention Scores: A Meta-Research Study. *Can Assoc Radiol J.* 2023 May;74(2):334-342. doi: 10.1177/08465371221134056. Epub 2022 Oct 27. PMID: 36301600.

75. Si L, Zhong J, Huo J, Xuan K, Zhuang Z, Hu Y, Wang Q, Zhang H, Yao W. Deep learning in knee imaging: a systematic review utilizing a Checklist for Artificial Intelligence in Medical Imaging (CLAIM). *Eur Radiol*. 2022 Feb;32(2):1353-1361. doi: 10.1007/s00330-021-08190-4. Epub 2021 Aug 4. PMID: 34347157.
76. Hickman SE, Woitek R, Le EPV, Im YR, Mouritsen Luxhøj C, Aviles-Rivero AI, Baxter GC, MacKay JW, Gilbert FJ. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology*. 2022 Jan;302(1):88-104. doi: 10.1148/radiol.2021210391. Epub 2021 Oct 19. PMID: 34665034; PMCID: PMC8717814.
77. Jia LL, Zhao JX, Pan NN, Shi LY, Zhao LP, Tian JH, Huang G. Artificial intelligence model on chest imaging to diagnose COVID-19 and other pneumonias: A systematic review and meta-analysis. *Eur J Radiol Open*. 2022;9:100438. doi: 10.1016/j.ejro.2022.100438. Epub 2022 Aug 18. PMID: 35996746; PMCID: PMC9385733.
78. Bhandari A, Scott L, Weilbach M, Marwah R, Lasocki A. Assessment of artificial intelligence (AI) reporting methodology in glioma MRI studies using the Checklist for AI in Medical Imaging (CLAIM). *Neuroradiology*. 2023 May;65(5):907-913. doi: 10.1007/s00234-023-03126-9. Epub 2023 Feb 7. PMID: 36746792; PMCID: PMC10105653.
79. Belue MJ, Harmon SA, Lay NS, Daryanani A, Phelps TE, Choyke PL, Turkbey B. The Low Rate of Adherence to Checklist for Artificial Intelligence in Medical Imaging Criteria Among Published Prostate MRI Artificial Intelligence Algorithms. *J Am Coll Radiol*. 2023 Feb;20(2):134-145. doi: 10.1016/j.jacr.2022.05.022. Epub 2022 Jul 31. PMID: 35922018; PMCID: PMC9887098.

ΠΑΡΑΡΤΗΜΑ Α

Αποδεκτές ερευνητικές εργασίες (άρθρα) παρούσας μελέτης.

Άρθρο 1: Nguyen KP, Fatt CC, Treacher A, Mellema C, Trivedi MH, Montillo A. Predicting Response to the Antidepressant Bupropion using Pretreatment fMRI. Predict Intell Med. 2019 Oct;11843:53-62. doi: 10.1007/978-3-030-32281-6_6. Epub 2019 Oct 10. PMID: 31709423; PMCID: PMC6839715.

Άρθρο 2: Weng JC, Lin TY, Tsai YH, Cheok MT, Chang YE, Chen VC. An Autoencoder and Machine Learning Model to Predict Suicidal Ideation with Brain Structural Imaging. J Clin Med. 2020 Feb 29;9(3):658. doi: 10.3390/jcm9030658. PMID: 32121362; PMCID: PMC7141277.

Άρθρο 3: Nguyen KP, Fatt CC, Treacher A, Mellema C, Trivedi MH, Montillo A. Anatomically-Informed Data Augmentation for Functional MRI with Applications to Deep Learning. Proc SPIE Int Soc Opt Eng. 2020 Feb;11313:113130T. doi: 10.1117/12.2548630. Epub 2020 Mar 10. PMID: 33767520; PMCID: PMC7990266.

Άρθρο 4: Chen GY, Huang CM, Liu HL, Lee SH, Lee TM, Lin C, Wu SC. Depression Scale Prediction with Cross-Sample Entropy and Deep Learning. Annu Int Conf IEEE Eng Med Biol Soc. 2020 Jul;2020:120-123. doi: 10.1109/EMBC44109.2020.9175816. PMID: 33017945.

Άρθρο 5: Lim SH, Yoon J, Kim YJ, Kang CK, Cho SE, Kim KG, Kang SG. Reproducibility of automated habenula segmentation via deep learning in major depressive disorder and normal controls with 7 Tesla MRI. Sci Rep. 2021 Jun 29;11(1):13445. doi: 10.1038/s41598-021-92952-z. PMID: 34188141; PMCID: PMC8241874.

Άρθρο 6: Zhi D, Calhoun VD, Wang C, Li X, Ma X, Lv L, Yan W, Yao D, Qi S, Jiang R, Zhao J, Yang X, Lin Z, Zhang Y, Chung YC, Zhuo C, Sui J. BNCPL: Brain-Network-based Convolutional Prototype Learning for Discriminating Depressive Disorders. Annu Int Conf IEEE Eng Med Biol Soc. 2021 Nov;2021:1622-1626. doi: 10.1109/EMBC46164.2021.9630010. PMID: 34891596; PMCID: PMC9021005.

Αρθρο 7: Wang Y, Gong N, Fu C. Major depression disorder diagnosis and analysis based on structural magnetic resonance imaging and deep learning. J Integr Neurosci. 2021 Dec 30;20(4):977-984. doi: 10.31083/j.jin2004098. PMID: 34997720.

Αρθρο 8: Nguyen KP, Chin Fatt C, Treacher A, Mellema C, Cooper C, Jha MK, Kurian B, Fava M, McGrath PJ, Weissman M, Phillips ML, Trivedi MH, Montillo AA. Patterns of Pretreatment Reward Task Brain Activation Predict Individual Antidepressant Response: Key Results From the EMBARC Randomized Clinical Trial. Biol Psychiatry. 2022 Mar 15;91(6):550-560. doi: 10.1016/j.biopsych.2021.09.011. Epub 2021 Sep 22. PMID: 34916068; PMCID: PMC8857018.

Αρθρο 9: Cheng Y, Liao T, Jia N. Classification Algorithms for Brain Magnetic Resonance Imaging Images of Patients with End-Stage Renal Disease and Depression. Contrast Media Mol Imaging. 2022 Jul 6;2022:4795307. doi: 10.1155/2022/4795307. PMID: 35854766; PMCID: PMC9279039.

Αρθρο 10: Liu Z, Nichol M.L. Wong, Robin Shao, Shwu-Hua Lee, Chih-Mao Huang, Ho-Ling Liu, Chemin Lin, Tatia M.C. Lee. Classification of Major Depressive Disorder using Machine Learning on brain structure and functional connectivity. J Affect Disord. 2022 Dec (10):100428;ISSN 2666-9153.

ΠΑΡΑΡΤΗΜΑ Β

Checklist for Artificial Intelligence in Medical Imaging (CLAIM):

Section / Topic	No.	Item	
TITLE / ABSTRACT			
	1	Identification as a study of AI methodology, specifying the category of technology used (e.g., deep learning)	
	2	Structured summary of study design, methods, results, and conclusions	
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the AI approach	
	4	Study objectives and hypotheses	
METHODS			
<i>Study Design</i>	5	Prospective or retrospective study	
	6	Study goal, such as model creation, exploratory study, feasibility study, non-inferiority trial	
<i>Data</i>	7	Data sources	
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (e.g., symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)	
	9	Data pre-processing steps	
	10	Selection of data subsets, if applicable	
	11	Definitions of data elements, with references to Common Data Elements	
	12	De-identification methods	
	13	How missing data were handled	
<i>Ground Truth</i>	14	Definition of ground truth reference standard, in sufficient detail to allow replication	
	15	Rationale for choosing the reference standard (if alternatives exist)	

	16	Source of ground-truth annotations; qualifications and preparation of annotators	
	17	Annotation tools	
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies	
Data Partitions	19	Intended sample size and how it was determined	
	20	How data were assigned to partitions; specify proportions	
	21	Level at which partitions are disjoint (e.g., image, study, patient, institution)	
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections	
	23	Software libraries, frameworks, and packages	
	24	Initialization of model parameters (e.g., randomization, transfer learning)	
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained	
	26	Method of selecting the final model	
	27	Ensembling techniques, if applicable	
Evaluation	28	Metrics of model performance	
	29	Statistical measures of significance and uncertainty (e.g., confidence intervals)	
	30	Robustness or sensitivity analysis	
	31	Methods for explainability or interpretability (e.g., saliency maps), and how they were validated	
	32	Validation or testing on external data	
RESULTS			
Data	33	Flow of participants or cases, using a diagram to indicate inclusion and exclusion	
	34	Demographic and clinical characteristics of cases in each partition	
Model performance	35	Performance metrics for optimal model(s) on all data partitions	

	36	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	
	37	Failure analysis of incorrectly classified cases	
DISCUSSION			
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability	
	39	Implications for practice, including the intended use and/or clinical role	
OTHER INFORMATION			
	40	Registration number and name of registry	
	41	Where the full study protocol can be accessed	
	42	Sources of funding and other support; role of funders	