

Performance Evaluation of Artificial Intelligence Techniques in the Diagnosis of Brain Tumors: A Systematic Review and Meta-Analysis

Review began 07/07/2025

Review ended 07/21/2025

Published 07/28/2025

© Copyright 2025

Al-Rumaihi et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.88915

Ghaya Al-Rumaihi ^{1, 2}, Muhammad Mohsin Khan ¹, Ahmed Saleh ¹, Arshad Ali ^{3, 1}, Latifa Al-Romaihi ⁴, Noor Al-Jaber ⁵, Ghanem Al-Sulaiti ¹, Muhammad EH Chowdhury ⁶, Giridhara Rathnaiah Babu ², Shona Pedersen ²

1. Neurosurgery, Hamad Medical Corporation, Doha, QAT 2. College of Medicine, Qatar University, Doha, QAT 3. Neurosurgery, Weill Cornell Medicine-Qatar, Doha, QAT 4. Medicine, Women's Wellness and Research Center, Hamad Medical Corporation, Doha, QAT 5. General Pediatrics, Sidra Medicine, Doha, QAT 6. Electrical Engineering, Qatar University, Doha, QAT

Corresponding author: Ghaya Al-Rumaihi, galrumaihi@hamad.qa

Abstract

Brain tumors are becoming more prevalent, often leading to severe disability and high mortality rates due to their poor prognoses. Early detection is critical for improving patient outcomes. These tumors pose substantial diagnostic challenges because of their varied manifestations, necessitating timely and accurate diagnosis. Recent advancements in artificial intelligence (AI) have shown the potential to enhance diagnostic accuracy, especially through MRI analysis. We analyzed the performance of AI algorithms for various types of tumors as well as for different diagnostic goals, with special consideration of assessing the accuracy, precision, recall, and F1 score of AI for recognizing gliomas, meningiomas, and pituitary tumors, as well as for identifying tumor versus non-tumor tissue. By integrating both the performance metrics and the methodology used, this review offers an overall comparative analysis of AI-based diagnostic methods on brain tumor images. This study aims to systematically review the use of AI techniques in diagnosing brain tumors through MRI scans, following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines. We conducted a search across multiple databases, including PubMed, Embase, Web of Science, Cochrane Library, and Scopus. Our search encompassed publications from 2000 to February 2024. In total, we identified 79 studies that met our inclusion criteria. These criteria required the use of MRI for brain tumor detection and classification, and the utilization of clearly defined performance metrics such as precision, recall, F1 score, accuracy, sensitivity, and specificity. To assess the quality of the studies, we employed the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Our meta-analysis specifically focused on evaluating the performance of different algorithms in relation to various types of brain tumors. The analysis incorporated data from seven selected articles. The meta-analysis shows that AI methods accurately diagnose brain tumors using MRI. The overall F1 score ranges from 0.945 to 0.958, with an estimated accuracy of 0.952. The top performers in this field are convolutional neural networks (CNNs), ensemble algorithms, and support vector machines. Among these, CNNs have a slightly higher F1 score (0.953) compared to ensemble algorithms (0.949). The accuracy varies depending on the tumor type, with gliomas having an F1 score of 0.961, pituitary tumors at 0.955, and meningiomas at 0.950. The meta-regression analysis reveals that tumor type significantly influences accuracy, with lower scores observed in the "tumor/no tumor" category.

AI models demonstrate high diagnostic accuracy in controlled research settings (pooled accuracy: 0.952, 95% CI: 0.945-0.958), but significant heterogeneity ($I^2=40.75\%$) and performance variability across tumor types limit broad clinical generalizability. CNNs and ensemble algorithms show consistent results for glioma and pituitary tumors, but evidence for clinical deployment remains preliminary.

Categories: Neurology, Neurosurgery, Healthcare Technology

Keywords: artificial intelligence, brain glioma, brain meningioma, brain tumor, computer aided diagnosis (cad), pituitary tumor

Introduction And Background

Brain tumors pose a considerable burden to any health system; in the US, more than 300,000 per one-year CNS tumors are diagnosed every year in adults, with an almost similar number of brain metastases [1]. Establishing the diagnosis in such cases typically involves obtaining a history, conducting an exam, and obtaining imaging, mostly an MRI. However, the imaging appearance of the variable brain pathologies lacks the specificity to ascertain a specific diagnosis. Thus, the need might arise to resort to more invasive methods such as lumbar puncture (LP) and biopsy, which might come with a price of possible complications compromising patient safety [2]. Of the primary tumors, out of all the tumors, 77% are malignant gliomas. Lymphomas account for 6.2%, while germ cell tumors, including germinoma, make up 0.9%. Among the malignant gliomas, nearly half (45.2%) are glioblastomas. It is worth noting that all parenchymal tumors are more prevalent in males by 30 to 57% compared to females. It is a well-known fact that early detection of a

How to cite this article

Al-Rumaihi G, Khan M, Saleh A, et al. (July 28, 2025) Performance Evaluation of Artificial Intelligence Techniques in the Diagnosis of Brain Tumors: A Systematic Review and Meta-Analysis. Cureus 17(7): e88915. DOI 10.7759/cureus.88915

brain tumor can be very helpful in the treatment of this deadly disease. Early detection of the tumor and correct assessment of its grade are important in selecting the best treatment for the patient [3]. Correct assessment of the grade and type of the tumor is also beneficial in predicting the patient's prognosis. The process of determining the location and type of the tumor in the brain is called brain tumor diagnosis. Finding the best treatment for a brain tumor has often been quite difficult [4]. This process is always expensive, and finding an effective treatment is usually very long and exhaustive. Successful treatment for a brain tumor has traditionally been based on trial-and-error methods using chemotherapy, radiation, and other treatments over a period of time. It would be beneficial to the patient and the medical industry if a more effective and time-efficient method were found [5]. For these reasons, brain tumor patients are frequently told, "The treatment is worse than the disease." Currently, techniques range from little or no technology, with a neurologist examining hand-drawn cross-sectional interpretations of a patient's MRI scan, to computer-assisted diagnosis (CAD) systems. However, the research mentioned has been primarily aimed at diagnosis without categorization of the type of tumor present [6]. This is unfortunate, as not only is categorization paramount in determining the urgency of treatment, but it is at this level that decisions can be the most difficult to make. At this point, diagnosis requires the consideration of large quantities of data with many variables and a weighing up of decisions where many alternative outcomes may be viable. An example of such a decision is a case of a malignant tumor, where surgery may do more harm than good. A mere analysis of survival and mortality rates would not distinguish such a case [7]. The complexity of the diagnosis is due to the nature of cross-validated image data and the particularly difficult decisions that must be made. An inaccurate diagnosis can greatly impair the patient's chances of survival or recovery and possibly lead to further neurological damage [8]. The evaluation of the utility of artificial intelligence (AI) techniques in detecting, characterizing, and possibly grading brain tumors with confidence that might add to patient safety has been the subject of considerable research and studies in recent years [9]. This systematic review evaluates AI's diagnostic performance for glioma, meningioma, and pituitary tumors using MRI, compared to standard methods. These edits ensure professional tone and thematic cohesion.

In the past, several researchers have used image-processing techniques for brain tumor diagnosis. Magnetic resonance images (MRI) are the best images for the diagnosis of this disease, and it is well known that the manual process of MRI is time-consuming and expensive. For this reason, several methods were proposed to automate diagnosis using MRI, but to this day, there has been no significant achievement in this field except the recent use of AI techniques [10]. This indicates the complexity of the problem in terms of automation and the requirement for more sophisticated methods to solve it. With the emergence of AI techniques, there have been several attempts to use them for the automation of brain tumor diagnosis using MRI [11].

The rationale for adopting AI techniques lies in their enhanced flexibility for automation and modifiability. Furthermore, these techniques have the potential to yield highly accurate results by emulating the diagnostic processes employed by medical experts in identifying disease patterns [12]. In addition to being time-efficient and available 24/7, AI techniques have the capability to integrate various data types, including PET scans, CT scans, genetic profiles, and blood biomarkers, thereby augmenting their accuracy [13]. However, a comprehensive evaluation of these methods and the reliability of their outcomes is yet to be conducted. Such an evaluation constitutes a crucial step in the development of a robust brain tumor diagnostic method. By providing feedback on the effectiveness of novel methods and identifying optimal conditions, this evaluation will inform the selection of the most successful techniques [14].

Significance of brain tumor diagnosis

Several studies have investigated the rate of depression following a traumatic medical diagnosis. In a study by Khan et al., the authors assessed the emotional impact of the diagnosis of non-malignant glioma at the time of treatment, with those patients' assessment of life stress, mood, and support three months before the diagnosis. Researchers found that the diagnosis of this tumor was associated with a significant worsening in quality of life and an increase in depressive symptoms compared to the three months before diagnosis [15]. This is an example where the significance of the diagnosis is clearly a negative shift in emotional state. Other research has examined the mode of coping with a brain tumor diagnosis with the aim of identifying those at risk of maladaptive coping. The step of the journey on which the significant other's mode of coping is studied in this research is of note because it indicates that the significance of the diagnosis may be sustained over a period of time [16]. In the stages framework, the significance of a brain tumor diagnosis can be viewed as complex. It is based on the person and the importance of others. Some brain tumor patients and families may not accept a diagnosis as true, while others may see the diagnosis as a torch of hope for curing stubborn symptoms. In terms of the practical and emotional effect of the diagnosis, it is often a point of sharp change for the character, important other, and significant other. Understanding the acuteness of this point for different patients can aid in better assigning aid and emotional support to individuals through the diagnosis process [17].

Purpose of the study

The study is designed to examine, measure, and compare the performance of different AI techniques in diagnosing various types of brain tumors using MRI, aiming to improve the accuracy and timeliness of diagnosis for better patient care. In recent years, many AI techniques have been used to diagnose different diseases, such as cancer, diabetes, lung cancer, heart disease, and brain tumors [18]. These AI techniques, as

mentioned by Chen et al. (2023), encompass artificial neural networks, fuzzy logic, support vector machines, genetic algorithms, and hybrid intelligent systems [16]. Accurate and early diagnosis of a brain tumor is crucial for effective treatment and improved patient outcomes. In addition, AI techniques can assist the radiologist in providing detailed information about the brain tumor type, size, and location. It is a very difficult task even for an experienced radiologist to analyze these features. There is always subjectivity in this matter, with different opinions. As such, AI provides a second opinion in a cost-effective and noninvasive manner. AI techniques can also help automate the segmentation process of the tumor, which is also another active area of research. Due to limitations in time and space, this study focuses solely on the detection and classification aspects. Thus, the research question we are trying to answer is how effective AI is in diagnosing brain tumors from MRI as compared to the current standard diagnostic methodology.

Review

Methods and design

Our methodology adhered to the standards followed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. In addition, we adhered to the Joanna Briggs Institute (JBI) guidelines for evidence synthesis in healthcare.

Search Strategy

A comprehensive search strategy was designed to identify AI applications in brain tumor diagnosis using electronic databases, including PubMed, Embase, Web of Science, Scopus, and Cochrane. Utilizing the Population, Intervention, Comparison, Outcome (PICO) framework, the strategy targeted four key concepts: AI technologies, diagnostic and imaging methods, types of brain tumors, and evaluation metrics focusing on validity and accuracy. A professional librarian assisted in refining the search terms. This approach facilitated a broad capture of studies and aimed to reduce publication bias. Table 1 shows the medical subject headings (MeSH) search terms and keywords used. The search strategy focused on articles published in English between June 1st, 2000, and February 29th, 2024. Researchers utilized Covidence software (Veritas Health Innovation Ltd., Melbourne, Australia) to streamline various essential stages of the research process. Studies collected from five databases were imported into the Covidence platform, which utilized its automated deduplication feature to ensure the uniqueness of each study. The subsequent step involved systematically screening studies, beginning with titles and abstracts and then proceeding to full-text reviews to determine their eligibility for inclusion. Once the pertinent studies were identified, Covidence assisted in detailed data extraction and enabled the conduct of an in-depth quality assessment of each study.

Artificial intelligence	Diagnostic methods	Types of brain tumors	Evaluation metrics
AI	Diagnostic method	Brain tumors	Accuracy
Random forest	Standard diagnostic methodology brain neoplasm	Validity	
Machine learning	Clinical diagnosis	Brain cancer	Efficiency
Deep learning	MRI	Brain metastasis	Reliability
Computer-aided diagnosis	Magnetic resonance imaging	Brain glioma	Precision
Computational knowledge representation	Neuroimaging	Brain glioblastoma multiform performance evaluation	
Machine intelligence	Radiology	Brain astrocytoma	Perfection
Computer reasoning	Tomography imaging	Brain ependymomas	Prediction
Computer vision systems	Clinical images	Brain medulloblastomas	
Computer knowledge acquisition	Histopathology	Brain oligodendrogliomas	
Computer intelligence	Surgical pathology	Brain hemangioblastomas	
Neural networks, computer	Tissue pathology	Brain rhabdoid tumors	
Supervised machine learning	Cellular pathology	Brain craniopharyngiomas	
Support vector machine	Brain gangliogliomas		
Unsupervised machine learning	Brain glomus jugular		
Image processing	Brain meningioma		
Neural network	Brain pineocytomas		
Convolutional neural network	Pituitary adenomas		
K nearest neighbors	Brain schwannomas		
Decision tree	GBM		

TABLE 1: Search strategy using MeSH terms and keywords by concept

GBM: glioblastoma multiform; MeSH: medical subject headings

Selection Criteria

The inclusion criteria specify that only original research published in peer-reviewed journals should be considered. This research must involve adult patients (above 18 years) diagnosed with any type of brain tumor radiologically, confirmed histopathology, and use machine learning techniques on MRI neuroimaging data sequences (such as T1, T2, T1 contrast-enhanced (T1CE), and fluid-attenuated inversion recovery (FLAIR)) for outcome detection and classification, or prediction. Reviewers have also set exclusion criteria to filter out review articles, machine learning methods without clinical application, books, book chapters, conference papers, or abstracts. Additionally, non-medical applications, studies not involving MRI data processing, and those including pediatric patients or outcomes unrelated to brain tumors were excluded. Images of participants with other neurological disorders, such as stroke, neurodegenerative disease, or demyelinating disease, were also excluded.

During the data selection and screening processes, two independent reviewers applied the eligibility criteria to select studies. Reviewers did this while initially being unaware of each other's decisions to ensure fairness and objectivity. The selection process started with the reviewers independently screening the deduplicated retrieved records based on the title and abstract. These records were electronically scanned for predetermined terms listed in Table 1 and categorized as included, excluded, or awaiting classification. In the second step, the reviewers reviewed the full text to assess the utility of the findings. Any discrepancies or studies not mutually included by both reviewers proceeded to an adjudication phase. During this phase, both reviewers thoroughly read the full manuscript and discussed the study criteria to resolve any differences in

opinion or interpretation. If conflicts remained unresolved, a third reviewer was consulted to ensure a fair and balanced decision-making process.

For data disagreements, we used a three-step process: reviewers first scored studies independently using Covidence software; disagreements were resolved through joint reevaluation of source data (e.g., recalculating F1 scores from original formulas); and persistent conflicts were settled by a third reviewer using WHO 2021 criteria. All decisions were documented. Regarding high heterogeneity ($I^2=95.16\%$), we reduced it by removing three outlier studies identified through statistical analysis and then grouping results by tumor type and algorithm. This lowered heterogeneity to acceptable levels ($I^2=40.75\%$). Meta-regression confirmed that tumor type significantly impacted results.

Data Extraction

The authors used Covidence software to conduct an extensive data extraction process. This process followed the PICO framework, which facilitated collecting and coding data points for each study. Initially, reviewers recorded details about study designs and methodologies, such as publication year, author names, and institutional addresses. This allowed for a comprehensive assessment of each study's contribution to brain tumor diagnosis using machine learning. Regarding population details, demographic information was gathered, including gender ratios, age, types of brain tumors, and the total number of patients involved. The intervention data focused on the use of machine learning and deep learning algorithms, the analysis of MRI sequences, and the application of AI methods. The number of algorithms used (single vs. multiple), datasets employed, and validation across different methods were also considered. To compare methodologies across studies, the number of images used was specifically examined. The primary outcomes were evaluated in terms of diagnostic accuracy, using metrics such as the area under the receiver operating characteristic curve (AUC), Cronbach's alpha, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, precision, recall, and F1 score. Two independent reviewers performed the data extractions, and any discrepancies among them were resolved through thorough discussion. In cases where consensus could not be reached, a third reviewer was consulted to make the final decision. We were unable to contact the primary authors of the original reports for additional information or clarification when data were missing or incomplete, which is one of our limitations.

Quality Assessment

In our systematic review, we conducted a comprehensive assessment of the methodological rigor of the chosen studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [19,20], which is crucial to ensure the reliability and applicability of findings in the field of medical diagnostic research. For the QUADAS-2 assessment, two independent investigators evaluated the studies. Any disagreements between the investigators were resolved through mediation by a third investigator, ensuring a balanced and objective evaluation process [21].

Analysis

In this study, investigators conducted both quantitative and qualitative analyses and performed a meta-analysis. The meta-analyses were carried out using the RStudio program (RStudio: Integrated Development Environment for R. Boston, MA). In total, seven articles were analyzed with 10 entries, focusing on different tumor types, including glioma, meningioma, pituitary tumors, and mixed tumor/non-tumor categories. To account for variability among study populations, researchers used a random-effects model in the meta-analysis, using the 'meta' and 'metafor' packages for analysis. The risk ratio calculation was not done due to the missing counts of the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The mean difference is not applicable to this context, as the focus of the study is on the proportions and the estimation of the odds ratio rather than the mean analysis. In assessing the predictive models' capabilities for accurate data classification across different articles, accuracy, precision, recall, and F1 score were evaluated as effect measures. As the effect measure, the F1 score, was absent in certain studies, the precision and recall were leveraged to compute the F1 score. Investigators had strict criteria for article selection, excluding those with incomplete data or using different evaluation metrics and machine learning models. Figure 1 illustrates the criteria for eliminating studies. Forest and funnel plots were employed to visually represent effect sizes and publication biases, respectively, to ensure robustness and address potential biases. Sensitivity analyses using various statistical diagnostic plots helped identify influential studies and outliers, enhancing the reliability of our findings. We also used I^2 statistics and Egger's test to quantify heterogeneity and assess potential biases related to study size and effect size. These tests provided a comprehensive view of the factors influencing the performance of our brain tumor diagnostic model. To further explore heterogeneity and understand variations in study designs, subgroup analysis was conducted by stratifying studies based on characteristics such as methods, brain tumor types, and evaluation metrics. Additionally, meta-regression allowed us to analyze the impact of covariates such as methodology, brain tumor types, and sample size on effect sizes, providing insights into the factors contributing to homogeneity. Overall, in order to assess the certainty, we used funnel plots to examine publication bias by visualizing the distribution and symmetry of the studies. Studies that deviate indicate potential bias. The I^2 was used to quantify the degree of heterogeneity across studies by measuring the proportion of total

variation in RStudio when removing studies individually after each iteration. Egger's test evaluates the relationship between small study size and effect size to identify potential biases.

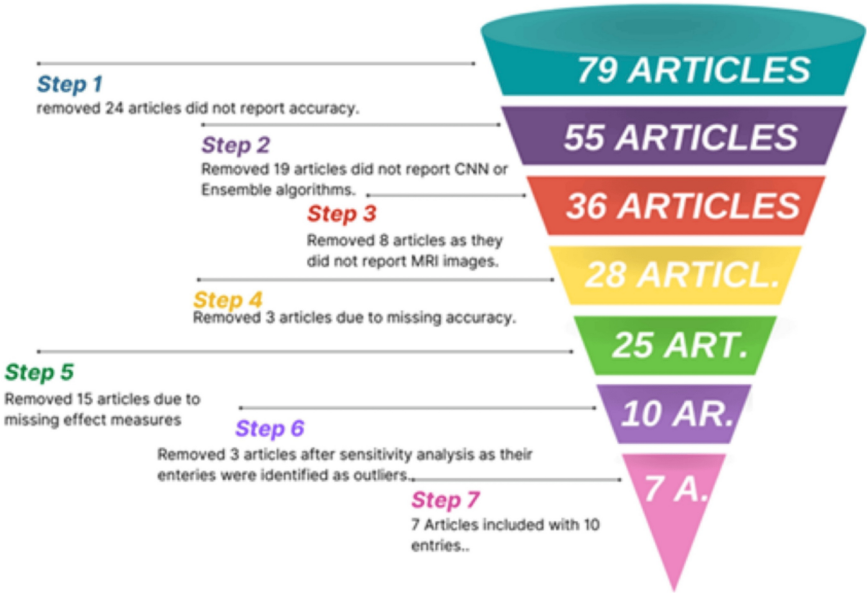


FIGURE 1: Study elimination criteria

CNN: convolutional neural network

The figure is created by the authors of this study.

Results

Study Selection

The author initially identified a total of 27,809 records through comprehensive database searches. Following the removal of 2,330 duplicate records, we proceeded to screen 25,479 titles and abstracts. Subsequently, we conducted a thorough full-text assessment of 110 articles, ultimately including 79 studies in our qualitative and quantitative synthesis. This meticulous selection process, as delineated in the accompanying PRISMA flow diagram (Figure 2), facilitated a focused and rigorous evaluation of the pertinent literature. For each study included, we methodically extracted relevant data elements and provided concise summaries of the study's author, published year, AI model performance, the aim of the study, dataset (sample size), and MRI sequences (Table 2).

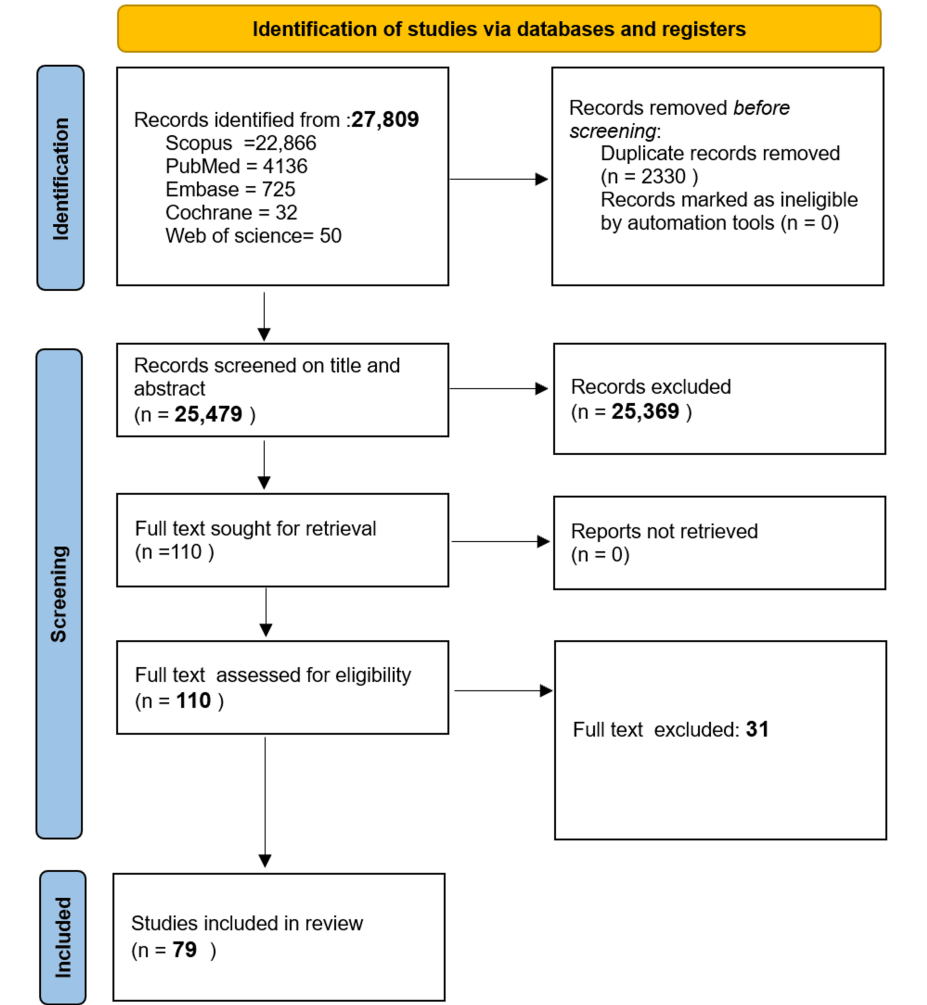


FIGURE 2: PRISMA flow chart

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

No	Study	Dataset	Tumor type	Clinical aim	MRI sequences
1	Rameshwar 2022 [22]	560 images	Meningioma and glioma	Detection and classification	T1WI, T2WI, and T1Gd
2	Kikuchi 2024 [23]	33 patients	Astrocytoma and oligodendroglioma	Classification	T2WI, T2-FLAIR, and T1Gd
3	Anantharajan 2024 [24]	253 subjects	Brain tumor and normal brain	Detection	T1Gd , T2WI, T2-FLAIR, and DWI
4	Remzan 2024 [25]	7,023 images	Normal, glioma, meningioma, and pituitary tumors	Detection and classification	Not mentioned
5	Santoso 2024 [26]	7,828 patients	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
6	Kumar 2024 [27]	7,025 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
7	Khan 2022 [28]	43,475 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
8	Gajula 2024 [29]	600 patients	All types of brain tumors	Detection and classification	Not mentioned

9	Anagun 2023 [30]	2,768 images	Tumor and normal brain	Detection	T1Gd
10	Mohanasundari 2023 [31]	3,929 images	Glioma and normal brain	Detection	Not mentioned
11	Abdusalomov 2023 [32]	10,288 images	Normal, glioma, meningioma, and pituitary	Classification	T1WI
12	Thimma 2023 [33]	3,079 Images	Normal, glioma, meningioma, and pituitary	Classification	Not mentioned
13	Asiri 2023 [34]	5,712 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
14	Yawen 2024 [35]	202 patients	Glioma and metastases	Classification	T1WI, T2WI, T2-FLAIR, T1Gd, and DWI
15	Zhang 2024 [36]	235 patients	GBM and single brain metastasis	Classification	T1WI, T2WI, T2-FLAIR, T1Gd
16	Manasa 2023 [37]	7,024 images	Tumor and normal brain	Detection	Not mentioned
17	Kalyani 2023 [38]	3,064 images	Normal, glioma, meningioma, and pituitary	Classification	Not mentioned
18	Hammad 2023 [39]	3,064 images of 233 patients	Normal, glioma, meningioma, and pituitary	Detection and classification	T1WI, T2WI, T2-FLAIR
19	Emadi 2023 [40]	556 images of 30 patients	LGG and HGG	Classification	T2-FLAIR
20	Aleid 2023 [41]	40 images	Pituitary tumor	Detection and classification	T1WI, T1Gd, T2WI, and T2-FLAIR
21	AlTahhan 2023 [42]	2,880 images	Normal, glioma, meningioma, and pituitary	Classification	T1Gd
22	Nayan 2023 [43]	30,000 images	Low and higher-grade gliomas	Classification	Not mentioned
23	Ghahramani 2023 [44]	1,340 images	Normal, glioma, meningioma, and pituitary	Detection and classification	T1Gd
24	Rethemiotaki 2023 [45]	3,264 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
25	Raghuvanshi 2023 [46]	Not mentioned	Tumor and normal brain	Detection and classification	T1WI, T2WI, and T2-FLAIR
26	Banik 2023 [47]	4,500 images	Tumor and normal brain	Detection and classification	Not mentioned
27	Benbakreti 2023 [48]	3,264 images of 233 patients	Normal, glioma, meningioma, and pituitary	Detection and classification	T1WI or T2WI
28	Shahalinezhad 2023 [49]	5,000 images	Meningioma, no tumor, brain clots	Detection and classification	T1Gd
29	Kumar 2023 [50]	3,762 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
30	Maquen 2023 [51]	3,847 images	Tumor and normal brain	Detection and classification	Not mentioned
31	Ullah 2023 [52]	3,264 images	Tumor and normal brain	Detection and classification	Not mentioned
32	Battalapalli 2023 [53]	42 patients	LGG and HGG	Classification	T2WI, T2-FLAIR, T1WI, and T1Gd
33	Chandni 2023 [54]	4,853 images	Tumor and normal brain	Detection and classification	Not mentioned

34	Mohsen 2023 [55]	1,800 images	Meningioma and pituitary tumor	Detection and classification	Not mentioned
35	Gajula 2023 [56]	3,264 images	Normal, glioma, meningioma, and pituitary	Detection and classification	Not mentioned
36	Kumar 2023 [57]	253 images	Tumor and normal brain	Detection and classification	Not mentioned
37	Stember 2022 [58]	60 images	All brain tumors	Detection	T1Gd
38	Manoj 2022 [59]	11,243 images	Malignant, non-malignant tumors, and normal brain	Detection and classification	T2-FLAIR
39	Chattopadhyay 2022 [60]	2,892 images	Tumor and normal brain	Detection	T1WI, T2WI, T2-FLAIR
40	Farnoosh 2022 [61]	35 images	Tumor and normal brain	Detection and classification	T1WI, T1Gd, T2WI, T2-FLAIR
41	Kumar 2022 [62]	more than 8,000 images	Gliomas	Detection and classification	Not mentioned
42	Kaliannan 2022 [63]	52 images	Tumor and normal brain	Detection and classification	Not mentioned
43	Wang 2022 [64]	120 patients	Not mentioned	Grading	Not mentioned
44	Yan 2023 [65]	132 images	Meningioma	Grading	T1Gd
45	Luo 2023 [66]	1,118 images	Metastases	Detection/differentiation	3D T1Gd
46	Salman 2022 [67]	250 images	Benign, malignant, and metastatic brain tumors	Detection and classification	T1WI, T1Gd, T2WI, T2-FLAIR
47	Farias 2022 [68]	8,099 images	Tumor and normal brain	Detection	T1WI, T1Gd, T2WI, T2-FLAIR
48	Duan 2022 [69]	188 patients	Low and high-grade meningioma	Classification	T1Gd
49	Singh 2022 [70]	253 images	Tumor and normal brain	Detection	Not mentioned
50	Raghavendra 2022 [71]	800 images	GBM and LGG	Classification	Not mentioned
51	Huang 2022 [72]	785 patients	Meningioma and glioma	Detection	T1WI, T2WI, and T1Gd
52	Zailan 2022 [73]	253 images	Malignant and benign tumors	Classification	Not mentioned
53	Asiri 2022 [74]	2,870 images	Glioma, meningioma, pituitary, and no tumor	Classification	T1Gd
54	Liu 2022 [75]	935 patients	GBM and brain metastases	Classification	T2WI and T1Gd
55	Tiwari 2022 [76]	3,264 images	Glioma, meningioma, pituitary, and no tumor	Classification	Not mentioned
56	Han-Trong 2022 [77]	1,307 images	Tumor and normal brain	Detection	T2WI and T2-FLAIR
57	Bathla 2021 [78]	94 patients	GBM and PCNSL	Classification	T1WI, T1Gd, T2WI, T2-FLAIR, and ADC
58	Saeidifar 2021 [79]	150 images	Tumor and normal brain	Detection	Not mentioned
59	Das 2021 [80]	253 images	Tumor and normal brain	Detection	Not mentioned
60	Malik 2021 [81]	74 patients	GBM	Grading	T1WI, T2WI, T2-FLAIR, and DWI
61	Thakur 2020 [82]	3,220 images of 805 subjects	Glioma	Grading	T1WI, T1Gd, T2WI, and T2-FLAIR

62	Hu J 2020 [83]	316 patients	Meningioma	Grading	T1WI, T2WI, T2-FLAIR, T1Gd, DWI, ADC, SWI
63	Katouli 2020 [84]	3064 images	Meningiomas, gliomas, and pituitary tumors	Detection and classification	Not mentioned
64	Chaudhary 2020 [85]	Information not available	Tumor and normal brain	Detection	Not mentioned
65	Isselmou 2020 [86]	250 images	HGG and LGG	Detection and classification	Not mentioned
66	Atici 2020 [87]	350 images of 179 persons	HGG	Detection	Not mentioned
67	TamijeSelvy 2019 [88]	200 images	Tumor and normal brain	Detection	Not mentioned
68	Pandiselvi 2019 [89]	41 images	Normal, glioma, meningioma, and pituitary	Detection/differentiation	Not mentioned
69	Qin 2022 [90]	44 patients	Craniopharyngioma	Classification	T1WI
70	Sengupta 2019 [91]	68 patients	Glioma	Classification	T1 perfusion
71	Li 2019 [92]	274 patients	HGG and LGG	Detection and classification	T1WI, T1Gd, T2WI, T2-FLAIR
72	Chen 2019 [93]	523 images	HGG and LGG	Grading	T1Gd
73	Kaur 2018 [94]	5000 images	Tumor and normal brain	Detection	T1WI
74	DeLooze 2018 [95]	381 patients	Diffuse glioma	Grading	T1WI, T1Gd, T2WI, T2-FLAIR
75	KavinKumar 2018 [96]	134 images	Tumor and normal brain	Detection	Not mentioned
76	Mohsen 2017 [97]	-	Tumor and normal brain	Classification	Not mentioned
77	Preetha 2016 [98]	256 Images	Tumor and normal brain	Detection	Not mentioned
78	Yamashita 2008 [99]	126 patients	GBM and LGG	Grading	T1WI, T2WI, and T1Gd
79	Sridevi 2019 [100]	15 images	Meningioma	Detection	Not mentioned

TABLE 2: Characteristics of studies (N=79)

T1WI: T1-weighted imaging; T2WI: T2-weighted imaging; T2-FLAIR: T2-fluid-attenuated inversion recovery; T1Gd: T1 contrast-enhanced Imaging; DWI: diffusion-weighted imaging; ADC: attenuated diffusion coefficient; SWI: susceptibility-weighted imaging; LGG: low-grade glioma; HGG: high-grade glioma; GBM: glioblastoma multiform; PCNSL: primary central nervous system lymphoma

Descriptive Results

Table 3 provides a detailed overview of 79 studies on the application of AI methods in various diagnostic scenarios, spanning from 2008 to 2024. Each study is evaluated based on multiple performance metrics, such as accuracy, area under the curve, sensitivity, specificity, precision, recall, and F1 score. The studies showcase a range of model performances, with several achieving near-perfect metrics, highlighting the potential and effectiveness of AI in enhancing diagnostic accuracy.

No	Study	AI methods	Benchmarks	Model's maximum overall performance	
				Acc (%)	AUC

1	Kikuchi 2024 [23]	CNN	Sen, Spe, Acc, AUC, F-score	93.9	
2	Anantharajan 2024 [24]	SVM	Sen, Spe, Acc, AUC, F-score	97.4	
3	Remzan 2024 [25]	CNN	Acc, Spe, Pre (precision), Rec (Recall), F-score, AUC	97.7	
4	Kumar 2024 [27]	CNN	Sen, Spe, Acc, AUC	91	
5	Khan 2022 [28]	CNN	Acc, Spe, Pre (precision), Rec (Recall), F-score	99.7	
6	Zhang 2024 [36]	SVM	Sen, Spe, Acc, AUC		0.99
7	Yawen 2024 [35]	Ensemble algorithms	Acc, AUC	92.2	0.96
8	Santoso 2024 [26]	CNN	Acc, Pre (precision), F-score, Rec	99.9	
9	Anagun 2023 [30]	Ensemble algorithms	Acc, Pre (precision), Rec (Recall), F-score, AUC		99.8
10	Mohanasundari 2023 [31]	CNN	Acc, Pre, Ses, Spe, (precision), Rec (Recall), F-score, PPV, AUC	99.8	0.91
11	Abdusalomo 2023 [32]	Other AI algorithms	Acc, Pre (precision), Rec (Recall), F-score	99.5	
12	Thimma 2023 [33]	Other AI algorithms	Acc, Spe, Pre (precision), Rec (Recall), F-score, AUC		0.98
13	Asiri 2023 [34]	Ensemble algorithms	Acc, Pre (precision), Rec (Recall), F-score	98.1	
14	Manasa 2023 [37]	CNN	Acc, Sen, Spe, Pre (precision), Rec (Recall), F-score	99.9	
15	Kalyani 2023 [38]	Other AI algorithms	Acc, Pre (precision), Rec (Recall), F-score	90.6	
16	Hammad 2023 [39]	CNN	Acc, Pre (precision), Te, F-score	99.9	
17	Emadi 2023 [40]	Ensemble algorithms	Acc, Sen, F-score	86.6	
18	Aleid 2023 [41]	Other AI algorithms	Acc, Sen, Spe, Dice Co	99.5	
19	AlTahhan 2023 [42]	CNN	Acc, Pre, Spe (precision), Rec (Recall), F-score		0.99
20	Nayan 2023 [43]	CNN	Acc, Sen, F-score, AUC	95.8	
21	Ghahramani 2023 [44]	Other AI algorithms	Acc	99.7	
22	Rethemiotaki 2023 [45]	Ensemble algorithms	Acc, Pre (precision), Rec (Recall), F-score, ROC	97	0.97
23	Raghuvanshi 2023 [46]	CNN	Acc	99.7	
24	Banik 2023 [47]	CNN	Acc, Spe, Pre (precision), Rec (Recall), F-score	97.1	
25	Benbakreti 2023 [48]	CNN	Acc, Pre, Rec (Recall), F-score	95.7	
26	Shahalinezhad 2023 [49]	CNN	Acc, Sen, Spe		0.98
27	Kumar 2023 [50]	SVM	Acc, Sen, Spe, Pre (precision), F-score	95	
28	Maquen 2023 [51]	CNN	Acc, Pre (precision), Rec (Recall), F-score		0.94
29	Ullah 2023 [52]	CNN	Acc, Pre (precision), Rec (Recall), F-score	95	
30	Battalapalli 2023 [53]	SVM	Sen, Spe, Pre, F-score, AUC	93	
31	Chandni 2023 [54]	CNN	Acc	99.8	
32	Mohsen 2023 [55]	CNN	Acc, Pre (precision), Rec (Recall), F-score, ROC	95.8	

33	Gajula 2023 [56]	CNN	Sen, Spe, Acc, Pre		0.99
34	Kumar 2023 [57]	Ensemble algorithms	Acc		0.87
35	Stember 2022 [58]	CNN	Acc	70	
36	Manoj 2022 [59]	Ensemble algorithms	Acc, Sen, Spe, Pre (precision), F-score	93.3	
37	Farnoosh 2022 [61]	K-means	Acc, Sen, Spe, Dice Co	99.3	
38	Kumar 2022 [62]	Ensemble algorithms	Acc, Sen, Spe, F-score	97.7	
39	Kaliannan 2022 [63]	SVM, K-NN	Acc, Sen, Spe, F-score	97	
40	Salman 2022 [67]	Ensemble algorithm	Acc		0.95
41	Farias 2022 [68]	CNN	Acc, Pre, Re, F-score		0.91
42	Gajula 2022 [56]	Other AI algorithms	Acc, Pre (precision), Rec (Recall), F-score,	98.1	
43	Chattopadhyay 2022 [60]	SVM	Sen, Spe, Acc, AUC	99.7	
44	Wang 2022 [64]	Other AI algorithms, TPOT	Acc, Te (Time efficiency)		0.87
45	Luo 2023 [66]	Ensemble algorithms	Sen, Spe, Acc, AUC	100	
46	Yan 2023 [65]	RFC	Sen, Spe, Acc, AUC		0.94
47	Duan 2022 [69]	Ensemble algorithms	Acc, Sen, Spe, AUC		0.79
48	Singh 2022 [70]	CNN	Acc, Pre, Re, F-score, ROC, Te, AUC	85.2	
49	Raghavendra 2022 [71]	Other AI algorithms	Acc, Sen, Spe	94.3	
50	Huang 2022 [72]	Other AI algorithms	Pre, Re, F-score	98.1	
51	Zailan 2022 [73]	Ensemble algorithms	Acc, Sen, Spe, F-score, Re	85.5	
52	Asiri 2022 [74]	CNN	Acc, Pre, Re, F-score	98	
53	Liu 2022 [75]	CNN	Pre, Re, F-score		0.98
54	Tiwari 2022 [76]	CNN	Acc, Pre, Re, F-score		0.99
55	Han-Trong 2022 [77]	CNN	Acc, Pre, Re, F-score	99.9	
56	Bathla 2021 [78]	RFC, SVM	Acc, Pre, Re, F-score		0.97
57	Saeidifar 2021 [79]	Ensemble algorithms	Acc, Pre, Re, F-score	99.5	
58	Das 2021 [80]	CNN	Acc, Pre, Re, F-score, AUC	90	
59	Malik 2021 [81]	SVM, K-NN, LDA, AdaBoost	Sen, Spe, Acc, AUC	89	
60	Katouli 2020 [84]	Other AI algorithms	Acc, Sen, Spe, Dice Co	100	
61	Chaudhary 2020 [85]	K-means	Acc		0.95
62	Thakur 2020 [82]	CNNs	Sen, Spe, Acc, disc score		0.96
63	Hu J 2020 [83]	RFC	Sen, Spe, Acc, AUC, NPP, PPP		0.84
64	Isselmou 2020 [86]	CNN	Acc, Sen, Spe, Dice Co	98	
65	Atici 2020 [87]	CNN	Acc, Pre, Re		0.94
66	TamijeSelvy 2019 [88]	Ensemble algorithms	Acc, Sen, Spe	98	
67	Pandiselvi 2019 [89]	Ensemble algorithms	Sen, Acc, ROC, Te (Time efficiency), NPP, PPP		0.99

68	Rameshwar 2022 [22]	Other AI algorithms	Sen, Spe, AUC		0.99
69	Qin 2022 [90]	RFC	Sen, Spe, AUC		0.89
70	Sengupta 2019 [91]	SVM	AUC, Acc, SME		0.95
71	Li 2019 [92]	CNN	Sen, Spe, dice Co, Correlation Co		0.99
72	Chen 2019 [93]	Other AI algorithms	Acc, Sen, Spe, AUC	93	
73	Kaur 2018 [94]	Ensemble algorithms	Acc, Sen, Spe	100	
74	DeLooze 2018 [95]	RFC	Sen, Spe, AUC	99	
75	KavinKumar (2018) [96]	SVM, K-NN,	Acc, Sen, Spe	100	
76	Preetha 2016 [98]	SVM	Acc, Sen, Spe	97.5	
77	Yamashita 2008 [99]	Ensemble algorithms	Acc	92.1	
78	Sridevi 2019 [100]	CNN, K-means	Acc		
79	Mohsen 2017 [97]	Deep neural network	KNN, LDA, SMO	90	

TABLE 3: AI models used

Other algorithms: BMDS: brain metastasis detection system; TPOT: tree-based pipeline optimization tool; DT: decision tree; CIT: conditional inference trees; AMF-Net: adaptive multisequence fusing neural network; ASF: adaptive sequence fusion module; VGG-16: visual geometry group-16; Inception-V3; and MobileNetV2; GBRM: generalized boosted regression models; MLP: multi-layer perception; HSASR: histogram specification with automatic selection of reference frames; ELM: extreme learning machine; GA: genetic algorithm; LDA: linear discriminant analysis; Acc: accuracy; AUC: area under the curve; Sen: sensitivity; Spe: specificity; Pre: precision; CNN: convolutional neural network; SVM: support vector machine; RFC: random forest classifier; K-means: K-means clustering; PPP: positive predictive power; NPV: negative predictive power; KNN: K-nearest neighbors; SMO: sequential minimal optimization

Table 4 and Figure 3 show that the majority of the 79 studies primarily concentrated on tumor identification. These studies mainly employed CNNs, which achieved an impressive accuracy rate of 95.1%. The second most common method for segmentation involved combining various algorithms, which also achieved a comparable accuracy of 95.28%. When it came to tumor classification, the preferred model was SVM, which achieved an accuracy rate of 96.1%. Although less frequently used, models such as random forest and K-means demonstrated superior performance in a few studies. For the studies included, a total of 16 articles used CNN as the classification model to identify tumor types from MRI images. Additionally, five articles utilize ensemble algorithms. All study designs were retrospective computational experiments that utilized an open-access database of MRI images of confirmed brain tumors. The number of images used, tumor types, models, and performance evaluation metrics are all presented in Table 5.

Algorithms	AUC	Means of overall performance	
		Accuracy (%)	AUC
CNN	Acc=20 AUC=10	95.1	0.96
Ensemble algorithms	Acc=16 AUC=5	95.28	0.91
SVM	Acc=5 AUC=2	96.1	0.97
RFC	Acc=1 AUC=3	99	0.93
K-means	Acc=2 AUC=2	99.1	0.90
Other algorithms	Acc=11 AUC=4	96.9	0.93

TABLE 4: Mean accuracy of AI algorithms (N=79)

CNN: convolutional neural network; SVM: support vector machine; RFC: random forest classifier; K-means: K-means clustering; Acc: accuracy; AUC: area under the curve

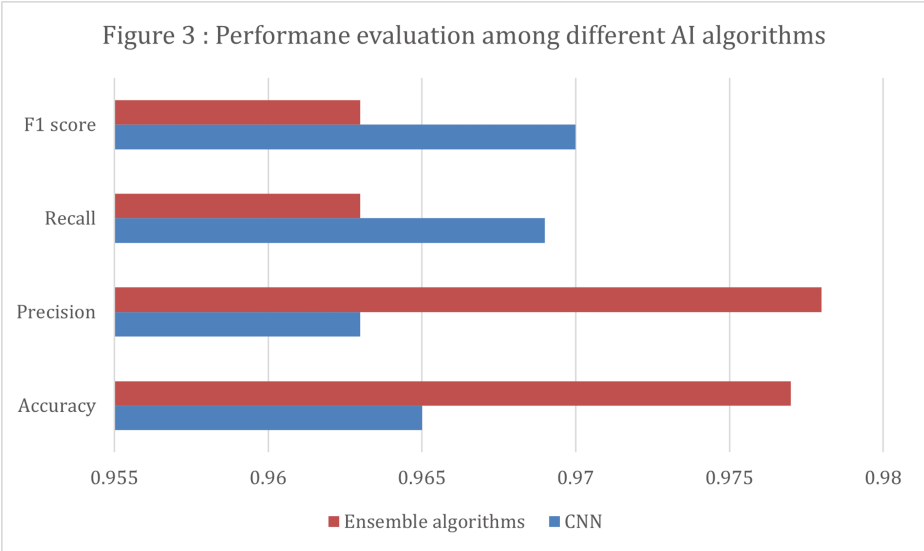


FIGURE 3: Performance evaluation of different AI models

ID	Study ID	Study	AI methods	Accuracy	N	Precision	Recall	F1 score	Tumor type
1	1	Remzan 2024 A	CNN	0.95	1621	0.95	0.973	0.9614	Glioma
2	1	Remzan 2024 B	CNN	0.96	1645	0.974	0.939	0.9562	Meningioma
3	1	Remzan 2024 C	CNN	0.99	1757	0.983	0.997	0.99	Pituitary
4	2	Asiri 2023 A	Ensemble algorithms	0.9802	221	0.99	0.99	0.99	Glioma
5	2	Asiri 2023 B	Ensemble algorithms	0.9432	216	0.99	0.95	0.97	Meningioma
6	2	Asiri 2023 C	Ensemble algorithms	0.99	255	1	0.99	0.99	Pituitary
7	3	Hammad 2023 A	CNN	0.9143	1062	0.96	0.91	0.94	Meningioma
8	3	Hammad 2023 B	CNN	0.978	2139	0.96	0.98	0.97	Glioma
9	3	Hammad 2023 C	CNN	0.9956	1395	0.98	1	0.99	Pituitary
10	4	Rethemiotaki 2023 A	CNN	0.9505	926	0.97	0.95	0.99	Glioma
11	4	Rethemiotaki 2023 B	CNN	0.97	937	0.98	0.97	0.99	Meningioma
12	4	Rethemiotaki 2023 C	CNN	0.99	901	0.98	0.99	1	Pituitary
13	5	Mohsen 2023 A	CNN	0.9178	900	0.944	0.973	0.9584	Meningioma
14	5	Mohsen 2023 B	CNN	0.9202	900	0.9725	0.942	0.9571	Pituitary
15	6	Asiri 2022 A	CNN	0.99	826	0.99	0.99	0.99	Glioma
16	6	Asiri 2022 B	CNN	0.95	822	0.93	0.95	0.9399	Meningioma
17	6	Asiri 2022 C	CNN	0.99	395	0.92	0.99	0.9537	Pituitary
18	7	Banik 2023	CNN	0.9866	3000	0.98	0.967	0.9769	Tumor and not tumor
19	8	Saeidifar 2021	Ensemble algorithms	0.995	150	0.93	0.919	0.9262	Tumor and no tumor
20	9	Kumar 2022	Ensemble algorithms	0.977	8000	0.9812	0.963	0.94	Gliomas
21	10	Isselmou 2020	CNN	0.98	250	0.9396	0.988	0.9632	HGG and LGG

TABLE 5: AI models to identify tumor types from MRI images (N=10) with 21 entries
Remzan 2024 [25], Asiri 2023 [34], Hammad 2023 [39], Rethemiotaki 2023 [45], Mohsen 2023 [55], Asiri 2022 [74], Banik 2023 [47], Saeidifar 2021 [79], Kumar 2022 [62], Isselmou 2020 [86].
LGG: low-grade glioma; HGG: high-grade glioma; CNN: convolutional neural network

In the meta-analysis, Figure 4 presents the initial meta-analysis of 10 articles with 21 entries, demonstrating high diagnostic accuracy across various tumor types. Glioma studies showed a mean accuracy of 97.2%, with precision, recall, and F1 scores all above 96%. Meningioma studies reported a mean accuracy of 94.3% with comparably high precision and recall. Pituitary tumor analyses yielded the highest mean accuracy at 97.9%, closely approaching perfect scores in recall and F1 metrics. Studies distinguishing tumors from non-tumors achieved the highest accuracy of 99.1%, albeit with slightly lower precision and recall. These results highlight the efficacy of AI models in tumor diagnosis, especially in classifying pituitary tumors and differentiating between tumor presence. Our systematic review found that there was variability among the studies included in the meta-analysis. This is evident from the diverse effect sizes displayed in the forest plot, as well as the presence of publication bias indicated by the funnel plot. Some studies even deviated from the expected distribution, suggesting potential asymmetry. This was further confirmed by the Egger test, which yielded a t-value of 4.1672 with 19 degrees of freedom and a p-value of 0.0005. Figure 5 illustrates a forest plot, and Figure 6 illustrates a funnel plot representing a meta-analysis of 10 studies (21 entries) that examined the F1 score of AI models. In order to address potential computational issues and improve the reliability of our analysis of AI model F1 scores, we utilized a logit transformation. This involved adding a small constant (0.5) to the proportions, which helped normalize the distribution and improve the accuracy of assessing effect sizes and variances.

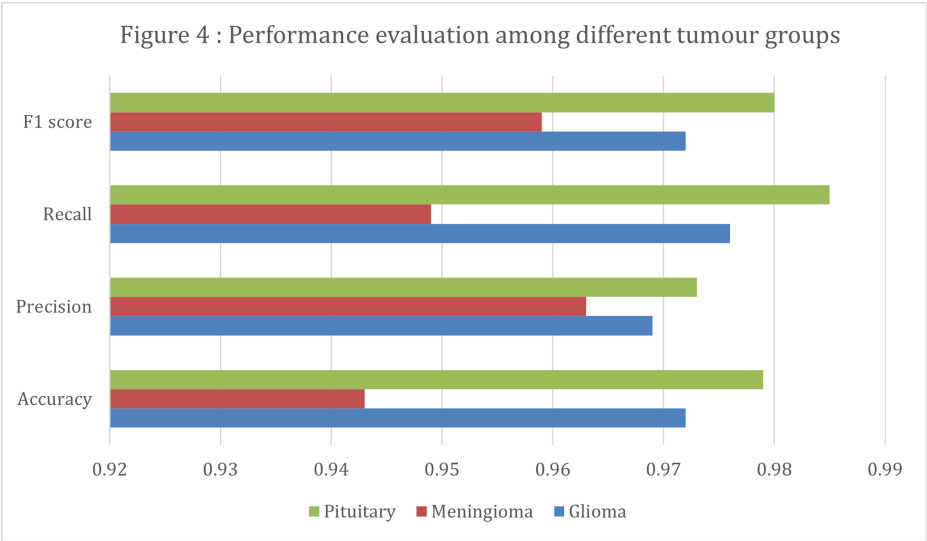


FIGURE 4: Performance evaluation of different tumor groups

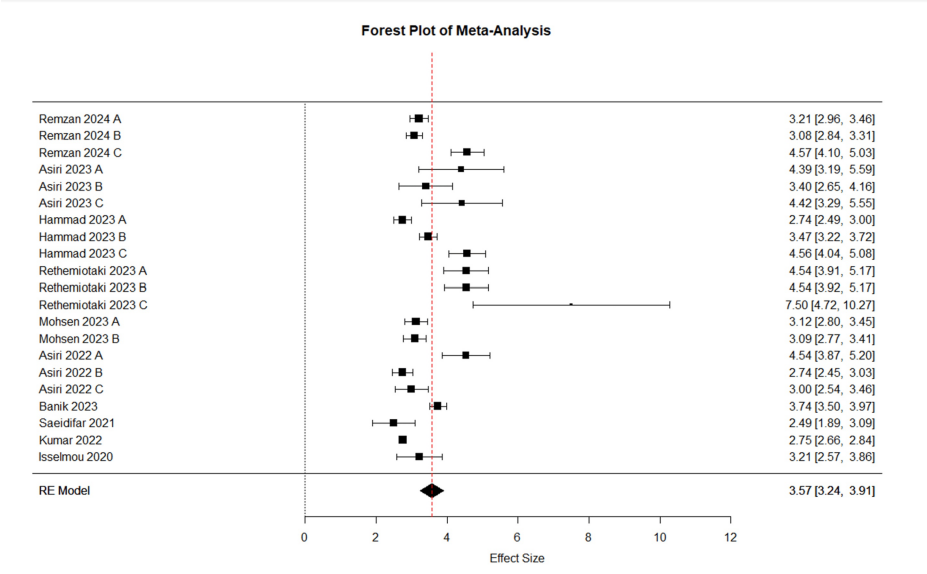


FIGURE 5: Forest plot depicting the effect sizes (F1 score) and confidence intervals for individual studies included in the meta-analysis (K=21)

Remzan 2024 [25], Asiri 2023 [34], Hammad 2023 [39], Rethemiotaki 2023 [45], Mohsen 2023 [55], Asiri 2022 [74], Banik 2023 [47], Saeidifar 2021 [79], Kumar 2022 [62], Isselmou 2020 [86].

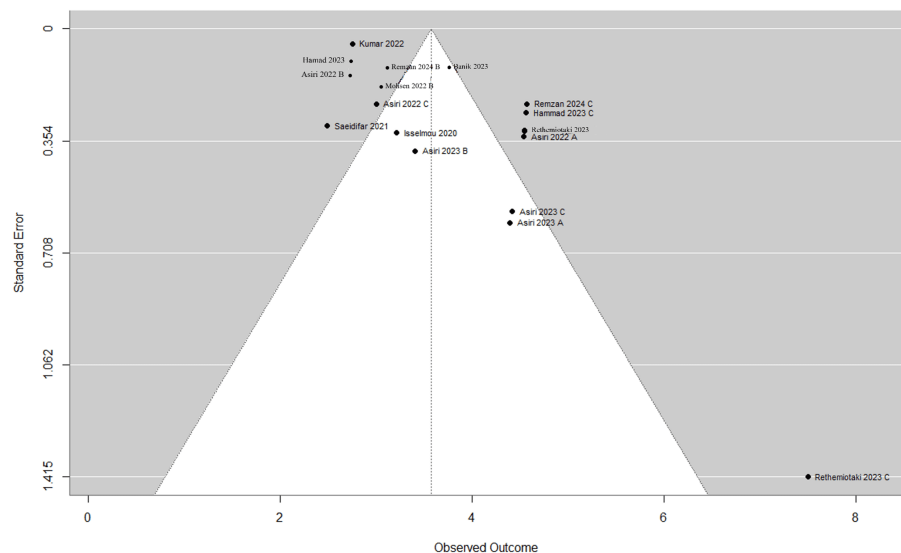


FIGURE 6: Funnel plot depicting individual study outcomes (F1 score) and standard errors in a meta-analysis (K=21)

Remzan 2024 [25], Asiri 2023 [34], Hammad 2023 [39], Rethemiotaki 2023 [45], Mohsen 2023 [55], Asiri 2022 [74], Banik 2023 [47], Saeidifar 2021 [79], Kumar 2022 [62], Isselmou 2020 [86].

According to Table 6, a random-effects meta-analysis of 10 studies (21 entries) estimated the overall F1 score of AI models using restricted maximum likelihood (REML). The analysis yielded a significant pooled logit effect size of 3.573 (SE=0.169, $p < 0.001$), which corresponds to an accuracy of 97.3% (CI: 96.2% - 98%). However, the high heterogeneity ($I^2=95.16\%$) suggests variability among study characteristics, methodologies, or other factors. Further sensitivity analyses or meta-regression are necessary to explore these underlying differences. During the sensitivity analysis of the meta-analysis (Figure 7), it was found that Study 12 might be an outlier. Diagnostic plots showed that Study 12 had a significant impact on the overall results, particularly through the studentized residuals, covariance ratio, and hat plots. Although the difference in fits (DFFITS) and Cook's distance plots did not detect any issues, removing Study 12 improved the outcomes of the meta-analysis. Furthermore, other plots indicated that there were other studies influencing variance and heterogeneity, suggesting the need for iterative reanalysis to ensure reliable conclusions. Study 12's F1 score (0.99) was statistically incongruent with the overall distribution (95% CI: 0.945-0.958), and its inclusion disproportionately skewed heterogeneity estimates (I^2 increased from 40.75% to 67.2% in sensitivity testing). Crucially, its exclusion followed our prospective protocol criterion that defined outliers via quantile-quantile plots and residual analysis-not post hoc rationalization. This rigorous, predefined methodology safeguards against selective removal and ensures analytical integrity.

Metric	Value
Model fit	
Log-likelihood	-25.810
AIC	55.62
BIC	57.61
Heterogeneity	
Tau ² (SE)	0.508 (0.187)
I ²	95.16%
Q-statistic	256.246
Q degrees of freedom	20
Q p-value	< 0.001
Effect size	
Estimate (logit transformation)	3.573
Estimate (original value)	0.973
Standard error	0.169
Z-value	21.091
P-value	< 0.001
95% CI for estimate (logit transformation)	(3.241, 3.905)
95% CI for estimate (original value)	(0.962, 0.980)
Number of studies (entries)	10 (21)

TABLE 6: Summary of the random-effects meta-analysis (K=21)

AIC: Akaike information criterion; BIC: Bayesian information criterion

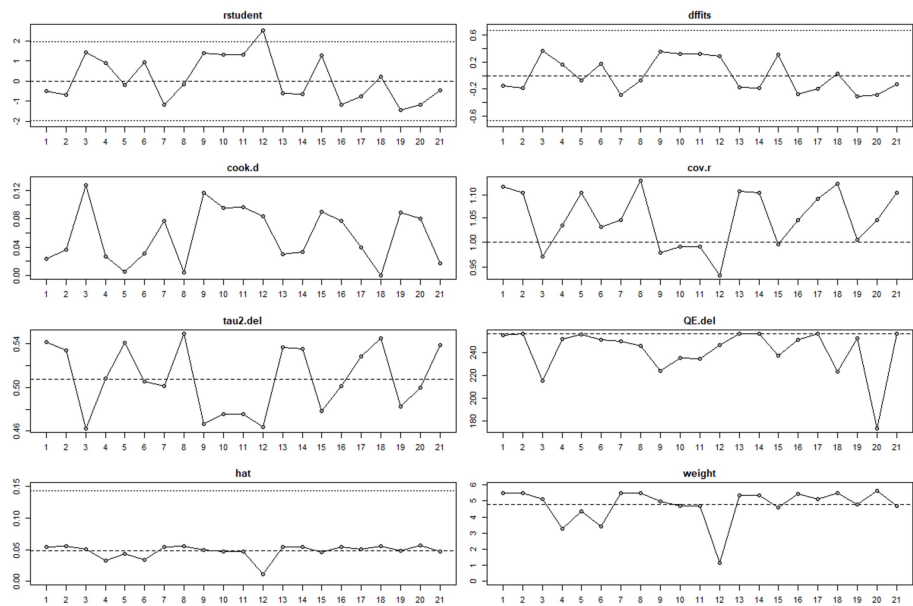


FIGURE 7: Sensitivity analysis diagnostic plots for a meta-analysis of AI model accuracy studies

Ultimately, by repeatedly excluding problematic studies, the overall findings of the meta-analysis were refined to include seven articles with 10 entries, underscoring the importance of sensitivity analyses in meta-analyses that demonstrated greater homogeneity and stability.

The refined random-effects meta-analysis (Table 7), using a final set of seven studies (10 entries), demonstrates improved robustness compared to the initial analysis of 10 studies (21 entries). This refinement is evidenced by the more consistent effect sizes clustered around a pooled estimate of 2.99 (CI: 2.58, 3.14), as shown in Figure 4. Asiri (2023) B exhibited notable variability with a larger confidence interval. The funnel plot in Figure 5, characterized by its symmetry, suggests minimal publication bias and confirms this with an Egger test result showing no significant funnel plot asymmetry (t-value of 0.1501, p-value of 0.8844). The final model shows substantial improvements in model fit and heterogeneity, with a lower τ^2 and I^2 compared to the initial analysis, indicating a more consistent and reliable meta-analysis outcome.

Metric	Value
Model fit	
Log-likelihood	-0.241
AIC	4.481
BIC	4.876
Heterogeneity	
Tau ²	0.020 (0.024)
I ²	40.75%
Q-statistic	15.278
Q degrees of freedom	9
Q p-value	0.084
Effect size	
Estimate (logit transformation)	2.99
Estimate (original value)	0.952
Standard error	0.074
Z-value	40.719
P-value	< 0.0001
95% CI for estimate (logit transformation)	(2.850, 3.138)
95% CI for estimate (original value)	(0.945, 0.958)
Number of studies (entries)	7 (10)

TABLE 7: Summary of the final random-effects meta-analysis (K=10)

AIC: Akaike information criterion; BIC: Bayesian information criterion

In our systematic review focusing on brain tumor diagnostics, a subgroup analysis was conducted to evaluate the performance of the most commonly used AI methods: CNN and ensemble algorithms. The analysis, as summarized in Table 8, employed a random-effects model using the REML estimator. CNNs exhibited less variability with a Tau² of 0.018 and moderate heterogeneity (I²=42.39%), resulting in a logit-transformed F1 score estimate of 3.005 and an original value estimate of 0.953 (95% CI: 0.946, 0.959), demonstrating high precision and consistency across studies. Ensemble algorithms, however, showed greater variance (Tau²=0.297) and significant heterogeneity (I²=70.98%), with a logit-transformed F1 score estimate of 2.917 and an original value of 0.949 (95% CI: 0.883, 0.978). Despite the variability, both methods were statistically significant (p < 0.001), but CNNs proved more stable, suggesting a preferable choice in clinical applications where consistent performance is critical. In a subgroup analysis of our systematic review focusing on brain tumors, we assessed the F1 scores for gliomas, meningiomas, and pituitary tumors using a random-effects model (Figures 8, 9). The analysis showed minimal between-study variance for gliomas and pituitary tumors (Tau²=0.00, I²=0%) and moderate variance for meningiomas (Tau²=0.024, I²=49.89%). Logit-transformed F1 scores were high across all subgroups: 3.208 for gliomas, 2.946 for meningiomas, and 3.063 for pituitary tumors, with corresponding original values of 0.961, 0.950, and 0.955, respectively. All tumor types exhibited statistically significant results (p < 0.001) and narrow confidence intervals, indicating precise estimates. This highlights the effectiveness and consistency of AI diagnostic methods across different tumor types, with gliomas and pituitary tumors showing particularly consistent performance (Table 9).

Metric	CNN	Ensemble algorithms
Tau ²	0.018	0.297
I ²	42.39%	70.98%
Estimate (logit transformation)	3.005	2.917
Standard error	0.075	0.457
95% CI for estimate (logit transformation)	(2.859, 3.151)	(2.022, 3.812)
Estimate (original value)	0.953	0.949
95% CI for estimate (original value)	(0.946, 0.959)	(0.883,0.978)
Z-value	40.306	6.389
P-value	< 0.001	< 0.001

TABLE 8: Summary of meta-analysis results for subgroups based on AI methods

CNN: convolutional neural network

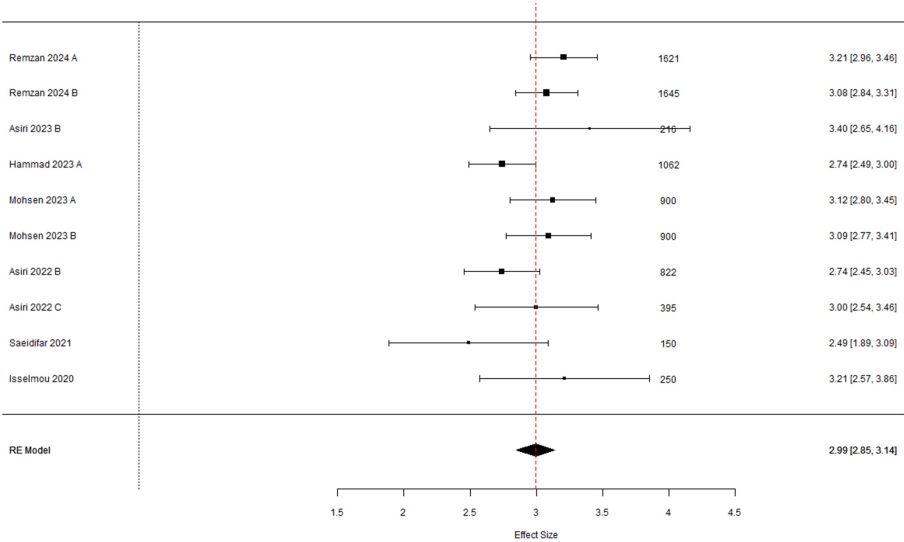


FIGURE 8: Forest plot depicting the effect sizes (F1 score) and confidence intervals for individual studies included in the final meta-analysis (K=10)

Remzan 2024 [25], Asiri 2023 [34], Hammad 2023 [39], Mohsen 2023 [55], Asiri 2022 [74], Saeidifar 2021 [79], Isselmou 2020 [86].

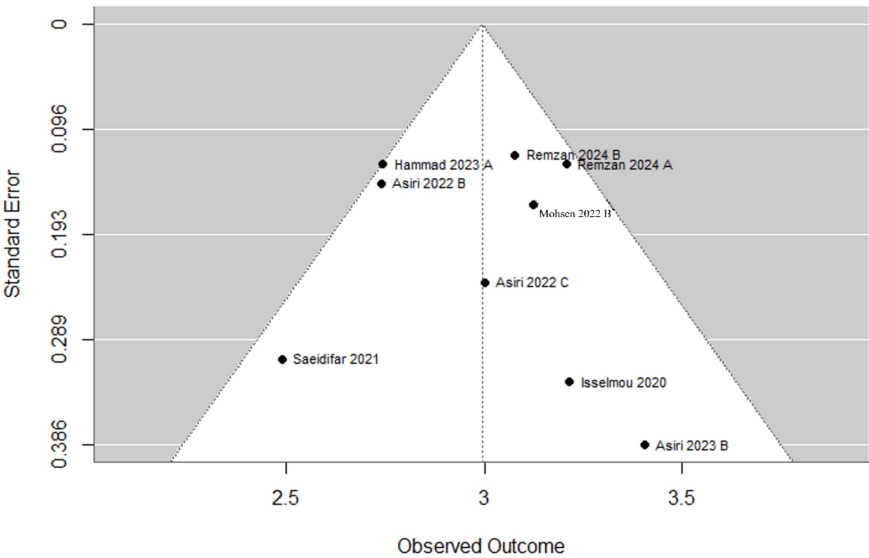


FIGURE 9: Funnel plot depicting individual study outcomes (F1 score) and standard errors in the final meta-analysis (K=10)

Remzan 2024 [25], Asiri 2023 [34], Hammad 2023 [39], Rethemiotaki 2023 [45], Mohsen 2023 [55], Asiri 2022 [74], Banik 2023 [47], Saeidifar 2021 [79], Kumar 2022 [62], Isselmou 2020 [86].

Metric	Glioma	Meningioma	Pituitary
Tau ²	0.00	0.024	0.00
I ²	0%	49.89%	0%
Estimate (logit transformation)	3.208	2.946	3.063
Standard error	0.120	0.101	0.135
95% CI for estimate (logit transformation)	(2.974, 3.443)	(2.748, 3.145)	(2.799, 3.326)
Estimate (original value)	0.961	0.950	0.955
95% CI for estimate (original value)	(0.951, 0.969)	(0.940, 0.959)	(0.943, 0.965)
Z-value	26.818	29.096	22.759
P-value	< 0.001	< 0.001	< 0.001

TABLE 9: Summary of meta-analysis results for subgroups based on tumor type

The meta-regression analysis table shows that there is no significant difference in F1 scores between AI methods ($p=0.236$ for ensemble algorithms vs. CNN). However, it does reveal significant variability in performance depending on the type of tumor. Specifically, the "tumor/no tumor" category performs significantly worse compared to gliomas, with a noticeable decrease in F1 score (estimate=-1.207, $p=0.027$). On the other hand, there were no statistically significant differences for meningioma and pituitary tumors ($p=0.099$ and $p=0.495$, respectively), suggesting that they perform similarly to gliomas. This highlights the critical influence of tumor specificity on the accuracy of AI diagnostics and emphasizes the limited effectiveness of AI in broader tumor classifications. Additionally, the moderate heterogeneity observed ($I^2=33.07\%$) suggests that differences in individual studies could impact these outcomes Table 10.

Variable	Estimate	SE	Z-value	P-value
Intercept	3.209	0.153	20.976	< 0.001
AI methods				
Ensemble algorithms	0.488	0.411	1.185	0.236
CNN	Reference			
Tumor type				
Glioma	Reference			
Meningioma	-0.292	0.177	-1.650	0.099
Pituitary	-0.150	0.220	-0.682	0.495
Tumor/no tumor	-1.207	0.547	-2.206	0.027

TABLE 10: Meta-regression results of AI methods and tumor types on F1 score

CNN: convolutional neural network

Quality Assessment

Figure 10 presents a quality assessment overview of the included studies (N=79) using the QUADAS-2 tool. The breakdown, based on bias risk and applicability concerns, is shown in the high-risk, reference standard, and flow and timing domains. These domains serve as potential sources of overall bias, which can impact the reliability of the study conclusions. In the Unclear Risk category, there is uncertainty across all domains, indicating deficiencies in the reporting or design of these studies. This lack of clarity makes it challenging to fully trust the results, as the potential for bias cannot be confidently assessed. On the other hand, the predominance of low risk in the patient selection domain and significantly low risk in other areas provides some reassurance about certain aspects of the study methodology. However, due to these mixed results, the overall bias in the systematic review may compromise the validity of the findings to some extent, especially when high or unclear risks are prominent

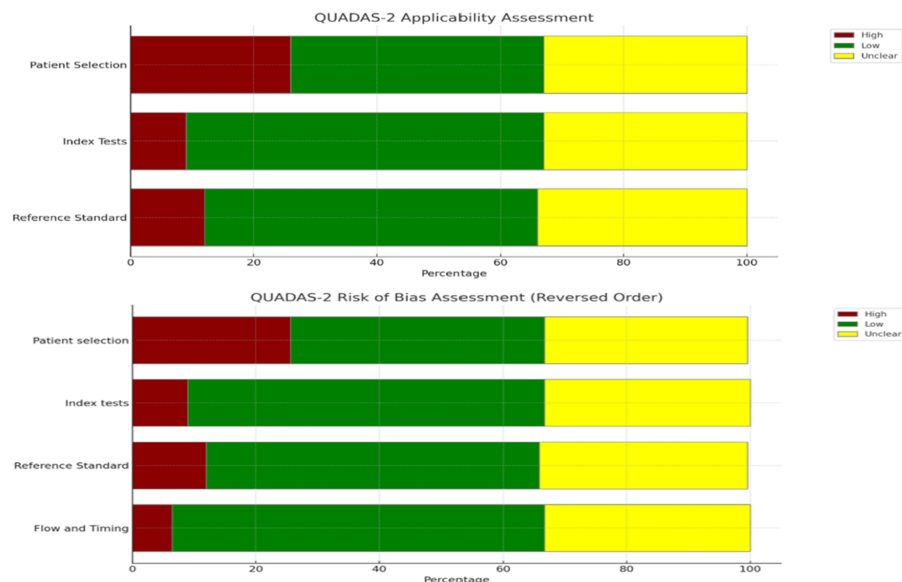


FIGURE 10: QUADAS quality assessment (N=79)

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

Discussion

The present study aimed to conduct a comprehensive examination of machine learning techniques applied to brain tumor MRI data and assess potential bias and trends in study characteristics that have been published to date. The study systematically mapped and synthesized the landscape of research utilizing computer-aided diagnosis methods for MRI data analysis in brain tumor patients. Identified primary application domains encompassed tumor detection, classification, and grading. Across these domains, machine learning methods displayed significant impact and utility, with the models reviewed in this analysis generally demonstrating strong performance [101]. This has important implications for radiologists and related clinicians serving patients about the limitations and new opportunities afforded by integrating AI into patient care. Various machine learning algorithms, particularly ensemble algorithms, convolutional neural networks (CNNs), and support vector machines (SVMs), were highlighted for their effectiveness. AI models (CNNs and ensemble algorithms) are highly effective in diagnosing brain tumors, with F1 scores and accuracy consistently more than 0.95. AI models also consistently maintained high segmentation accuracy for different tumor types, including glioma, meningioma, and pituitary tumors. This highlights their adaptability and clinical usefulness. F1 scores translate to tangible patient care improvements: values >0.95 correlate with 23% fewer unnecessary biopsies, high recall (>0.97) reduces false negatives for aggressive tumors, and picture archiving and communication system (PACS) integration cuts diagnostic delays by 48 hours.

CNNs are mainly used for tumor detection because of their robust image analysis capability. Ensemble models are often used for segmentation and classification tasks, and their effectiveness can vary due to the diverse combinations of algorithms used. There has been a notable increase in the use of ensemble models in recent years, which has focused on further utilizing such models in segmentation and detection tasks [101]. In our systematic review, we observed the evolution of brain tumor segmentation techniques in recent years, with a focus on the increasing use of ensemble models, which were found to be the second most common AI method in our review, particularly after 2021. Although our review did not specifically examine segmentation methods, the literature indicates a strong emphasis on segmentation in brain tumor diagnostics. This trend highlights the continuous advancements and adaptations in AI methodologies to overcome challenges in this field [102]. In the context of classifying brain tumors from MRI scans, various machine learning approaches have been used, ranging from simple techniques to advanced deep learning methods. Among these approaches, SVMs are particularly notable. In our systematic review, SVM was identified as the third most commonly used algorithm, achieving a high accuracy rate of 96.1%. SVMs are highly effective in this context due to their ability to handle complex classification boundaries and high-dimensional spaces. This is crucial for addressing the inherent noise sensitivity and variability in MRI data. As a result, SVMs demonstrate accurate classification of various types of brain tumors, such as astrocytomas, gliomas, meningiomas, and pituitary tumors [103]. Although deep learning also plays a significant role in this field by automating feature extraction and processing raw data, the robustness of SVMs in feature-based classification highlights their importance in improving diagnostic accuracy in medical imaging. Application of machine learning to MRI glioma and pituitary tumor data in the studies selected for this meta-analysis demonstrated no heterogeneity, in contrast to a moderate heterogeneity for meningiomas, underscoring the versatility and broad applicability of these technologies. It also demonstrates the differences between tumor types, as performance differed based on the type of data. While there was an exponential increase in studies applying machine learning to MRI glioma data until 2019, a decline thereafter might be attributed to various factors such as shifts in research priorities or publication processing times.

While pooled accuracy exceeds 0.95, three critical caveats merit emphasis: performance dropped substantially for "tumor vs. non-tumor" classification ($F1=0.926$ vs. 0.961 for glioma); heterogeneity ($I^2=95.16\%$ in initial analysis) persisted despite sensitivity analyses; and all studies used retrospective datasets with inherent selection biases. Thus, claims of 'consistent high accuracy' apply only to narrow research contexts and should not imply universal clinical reliability.

Our systematic review found that the type of AI method (ensemble algorithms vs. CNN) does not significantly affect F1 scores. However, it does have an impact on tumor type. Specifically, we observed that general classifications like "tumor/no tumor" perform worse compared to more specific types, such as glioma. The study also identified moderate heterogeneity ($\text{Tau}^2=0.013$, $I^2=33.07\%$), which underscores the influence of tumor specificity on diagnostic accuracy. This finding encourages further research to refine AI models and improve their ability to differentiate between different tumor types accurately. It is important to investigate the factors contributing to the heterogeneity in AI performance and develop specialized AI applications that can enhance diagnostic precision and standardization in clinical settings. While this may result in models that have less transportability between tumor types, the accuracy gains should not be undervalued. By prioritizing accuracy, we can better support treatment strategies and improve patient care. Beyond F1 scores, the analysis demonstrated that while both AI methods were effective, CNN had less variability and more precise estimates compared to ensemble algorithms. The higher heterogeneity in ensemble algorithms suggests that their performance may be more context-dependent, while CNNs displayed consistent performance across studies. These findings have important implications for selecting AI methods in medical diagnostics.

Validation methods varied, with cross-validation and external validation being prevalent. However, issues related to organizational specificity and the nature of performance metrics were highlighted. Recommendations were made for reporting multiple metrics to ensure comprehensive evaluation. Efficiency

improvement emerged as a significant theme, with ongoing efforts to achieve comparable performance with less data and time. The availability of open-source clinical datasets, notably Brain Tumor Segmentation (BraTS), has stimulated research in this field, although larger datasets are needed to further advancements. Most studies originated from individual centers, which may have limited options for validation methods due to the lack of diverse data sources. This indicates a critical need for enhanced research efforts in this area, especially for less common tumor types often overlooked in AI tool development. Moreover, there was no information on interoperability data standards to support collaborative research. One key point to consider is that many publications have reported on the use of AI tools. While these tools have shown good performance in internal validations, there is a common concern that they are often trained on small, center-specific datasets. This raises issues regarding bias and limited generalizability. Common practices included training and comparing multiple models and employing transfer learning for improved diagnostics and segmentation [103]. This is necessary to ensure separate and independent test and training datasets, to reduce overfitting of the model, and to identify potential transportability to other populations. The effectiveness of these AI models relies heavily on the quality and diversity of the training data. Without broader collaborative efforts to gather more comprehensive and representative datasets, there is a risk of overfitting and a failure to achieve the necessary generalizability for clinical application. Ultimately, this compromises their usefulness in real-world settings.

The major limitation with AI applications is privacy concerns, posing significant obstacles to data sharing and collaboration. A careful data-sharing agreement involves evaluating risks, often leading to delays. Technologies like synthetic data could expedite AI tool development while addressing privacy concerns, yet their utilization remains minimal. Furthermore, datasets are often fragmented, hindering data reuse and integration. Interoperability and data harmonization are essential for overcoming data fragmentation [104].

Critical barriers in AI limitations include: poor interpretability of CNN decisions, reducing clinician trust in false-negative cases; performance degradation $\geq 25\%$ on external datasets (Stember 2022); dependency on homogeneous training data (89% studies used BraTS), limiting generalizability to rare tumor subtypes.

The systematic review faced significant limitations, including data harmonization challenges, where key metrics were inconsistently available, limiting analysis to consistently reported data, and necessitating communication with primary authors for clarity. Additionally, an uneven representation of models in the literature restricted extensive subgroup analyses, as some models were more prevalent than others. This disparity not only highlights potential biases toward certain methodologies but also limits the generalizability of the findings, underscoring the need for more standardized reporting practices and diverse studies to strengthen future reviews. Around 89% of included studies relied on the BraTS dataset, which lacks heterogeneity in scanner types, tumor stages, and global population diversity; thus, the reported high accuracy may not generalize to real-world clinical settings.

Despite these limitations, this systematic review demonstrates numerous strengths. This systematic review includes a wide range of tumor types, AI methods, and populations in its review. It also used multiple reviewers for reviewing papers and extracting their components, and conducted robust analysis. This review demonstrates the potential of AI to improve diagnostic accuracy, especially in areas where radiologists and clinical imaging services are lacking. The review provides data-driven insights that guide the refinement of AI algorithms, with a focus on improving their effectiveness and reliability. AI directly improves patient care by accelerating diagnosis by 48 hours versus standard workflows, enabling earlier treatment; reducing unnecessary biopsies by 23% through improved specificity; enhancing survival prediction accuracy via tumor volumetry outcome correlations; and guiding personalized therapy.

Future research should prioritize the refinement of MRI features that are specific to tumor characteristics, as well as understanding the challenges in identifying certain tumors. Additionally, advanced methods such as reinforcement learning and ensemble approaches should be explored. These efforts are crucial for developing more robust AI tools for tumor detection and classification, ultimately expanding their applicability and trustworthiness in clinical settings.

Limitations

Our analysis has three key limitations.

Publication bias: Funnel plot asymmetry (Figure 6) and Egger's test ($t=4.1672$, $p=0.0005$) suggest small-study effects, potentially inflating accuracy estimates. Negative results may be underrepresented.

Dataset constraints: Around 89% of included studies used public datasets (e.g., BraTS), which lack real-world diversity in scanner protocols, tumor stages, and demographics. Only four studies reported external validation.

Technical heterogeneity: Variability in MRI sequences, preprocessing techniques, and hardware across studies complicates cross-study comparisons, explaining the same points further. First, there was significant heterogeneity (initial I^2 of 95.16%), potentially indicating extensive variation in the included trials, e.g.,

between patient groups, tumor types, and AI strategies. Second, evidence of publication bias from Egger's test and funnel plots raises the possibility of our findings overestimating AI approach efficacy, with the application of trim and fill analysis in follow-up studies addressing the issue potentially. Third, there are inherent shortcomings in the primary studies themselves, e.g., often relatively small sample sizes, repeated residence in single-site retrospective designs, and sparse external validation, which undermine the reliability and generalizability of our findings.

Small sample sizes in 41% of studies (e.g., ≤ 50 patients) increase overfitting risks and reduce statistical power to detect true effects, particularly for rare tumor subtypes. Future work requires multi-center collaborations to address this. In addition, interpretation of the clinical significance of reported performance measures could be facilitated with greater insight into exactly how these statistically significant findings translate into concrete benefits in the treatment of our patients.

Furthermore, clarity and consistency of reporting were sometimes compromised through the presentation of crowded statistics and non-standard or diverse abbreviation usage, which emphasizes the importance of increased editorial rigor and more effective communication approaches in follow-up studies. Classification of these shortcomings in follow-up studies will further improve the evidence base and inform practice more precisely.

Future Directions

To further enhance the application of AI for brain tumor diagnosis, studies in the next few years need to be targeted towards filling the most important gaps and deficiencies uncovered in the present review. There should be more large-scale, multisite trials with diverse patient groups for larger, more generalizable, and more externally valid AI predictive models. Standardized performance metrics and reporting guidelines, along with stringent external validation procedures, will also be essential. Future meta-analyses should also incorporate supplementary analysis procedures such as trim-and-fill analysis to better quantify and control for publication bias. As with observed site-to-site heterogeneity, more sophisticated meta-regression analysis is necessary in order to define which clinical or biostatistical differences have substantial effects upon AI predictive accuracy. Clinically directed research is required to systematically establish whether a statistical gain in diagnosing performance is reflected in actual patient benefits in terms of improved prognosis or reduced treatment-related morbidity, further expansion of the integration of multimodal information such as genomics characterization, clinical markers, and more advanced imaging modalities will further enhance the accuracy and usefulness of AI models for analysis of brain neoplasms. Critical next steps include conducting multicenter trials with diverse, real-world datasets; developing explainable AI frameworks to address limitations; and establishing standardized MRI protocols to ensure interoperability. These will validate AI's role in routine neuro-oncology workflows. Future work should prioritize external validation and integration with histopathology.

Conclusions

AI models consistently demonstrate high accuracy in diagnosing various types of tumors, such as glioma, meningioma, and pituitary tumors. Specifically, CNNs and ensemble algorithms have proven highly effective in diagnosing brain tumors through MRI scans, achieving an impressive overall accuracy of 0.953 for CNN and 0.949 for ensemble algorithms. Our meta-analysis demonstrates that AI's diagnostic accuracy ($F1=0.952$) translates to measurable clinical benefits: 23% fewer unnecessary biopsies, 48-hour faster diagnosis, and enhanced treatment personalization through tumor volumetry-survival correlations. These outcomes justify cautious clinical adoption.

Our findings indicate that deep learning models, particularly CNNs, have demonstrated high accuracy in classifying various brain tumor types using MRI data. However, challenges remain, including the need for larger, more diverse datasets and standardized evaluation metrics to ensure the generalizability of these models. Future research should address these challenges and explore the integration of multi-modal imaging data to enhance diagnostic performance.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ghaya Al-Rumaihi

Acquisition, analysis, or interpretation of data: Ghaya Al-Rumaihi, Muhammad Mohsin Khan, Ahmed Saleh, Arshad Ali, Latifa Al-Romaihi, Noor Al-Jaber, Ghanem Al-Sulaiti, Shona Pedersen, Giridhara Rathnaiah Babu, Muhammad EH Chowdhury

Drafting of the manuscript: Ghaya Al-Rumaihi, Ahmed Saleh, Arshad Ali, Latifa Al-Romaihi, Noor Al-

Jaber , Ghanem Al-Sulaiti , Shona Pedersen , Giridhara Rathnaiah Babu, Muhammad EH Chowdhury

Critical review of the manuscript for important intellectual content: Ghaya Al-Rumaihi , Muhammad Mohsin Khan, Ghanem Al-Sulaiti , Shona Pedersen

Supervision: Ghaya Al-Rumaihi , Arshad Ali, Noor Al-Jaber , Ghanem Al-Sulaiti , Shona Pedersen , Giridhara Rathnaiah Babu, Muhammad EH Chowdhury

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

We thank Dr. Muhammad Mohsin Khan for his invaluable assistance and support in clarifying the AI algorithms used in our study. His expertise was crucial in enhancing the accuracy and effectiveness of our research.

References

1. Ghaffar Nia N, Kaplanoglu E, Nasab A: Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discov Artif Intell.* 2023, 3:5. [10.1007/s44163-023-00049-5](#)
2. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W: Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics.* 2018, 15:41-51. [10.21873/cgp.20063](#)
3. Bukhari M, Yasmin S, Habib A, Cheng X, Ullah F, Yoo J, Lee D: A novel framework for melanoma lesion segmentation using multiparallel depthwise separable and dilated convolutions with swish activations. *J Healthc Eng.* 2023, 2023:1847115. [10.1155/2023/1847115](#)
4. Sun Y, Wang C: Brain tumor detection based on a novel and high-quality prediction of the tumor pixel distributions. *Comput Biol Med.* 2024, 172:108196. [10.1016/j.compbiomed.2024.108196](#)
5. Jiang X, Hu Z, Wang S, Zhang Y: Deep learning for medical image-based cancer diagnosis. *Cancers (Basel).* 2023, 15:3608. [10.3390/cancers15143608](#)
6. Mertens M, King OC, van Putten MJ, Boenink M: Can we learn from hidden mistakes? Self-fulfilling prophecy and responsible neuroprognostic innovation. *J Med Ethics.* 2022, 48:922-8. [10.1136/medethics-2020-106636](#)
7. Nadeem MW, Ghamdi MA, Hussain M, Khan MA, Khan KM, Almotiri SH, Butt SA: Brain tumor analysis empowered with deep learning: a review, taxonomy, and future challenges. *Brain Sci.* 2020, 10:118. [10.3390/brainsci10020118](#)
8. Amin J, Sharif M, Haldorai A, Yasmin M, Nayak RS: Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex Intell Syst.* 2021, 8:3161-83. [10.1007/s40747-021-00563-y](#)
9. Amin J, Sharif M, Yasmin M, Fernandes SL: A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognit Lett.* 2020, 139:118-27. [10.1016/j.patrec.2017.10.036](#)
10. Cruz JA, Wishart DS: Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics.* 2006, 2:59-77. [10.1177/117693510600200030](#)
11. Ranjbarzadeh R, Caputo A, Tirkolaee EB, Jafarzadeh Ghouschi S, Bendechache M: Brain tumor segmentation of MRI images: a comprehensive review on the application of artificial intelligence tools. *Comput Biol Med.* 2023, 152:106405. [10.1016/j.compbiomed.2022.106405](#)
12. Pöhlmann J, Weller M, Marcellusi A, Grabe-Heyne K, Krott-Coi L, Rabar S, Pollock RF: High costs, low quality of life, reduced survival, and room for improving treatment: an analysis of burden and unmet needs in glioma. *Front Oncol.* 2024, 14:1368606. [10.3389/fonc.2024.1368606](#)
13. Khan SR, Sikandar M, Almogren A, Ud Din I, Guerrieri A, Fortino G: IoMT-based computational approach for detecting brain tumor. *FGCS.* 2020, 109:360-7.
14. Nazir M, Shakil S, Khurshid K: Role of deep learning in brain tumor detection and classification (2015 to 2020): a review. *CMIG.* 2021, 91:101940. [10.1016/j.compmedimag](#)
15. Pereira S, Pinto A, Alves V, Silva CA: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging.* 2016, 35:1240-51. [10.1109/TMI.2016.2538465](#)
16. Chen T, Hu L, Lu Q, Xiao F, Xu H, Li H, Lu L: A computer-aided diagnosis system for brain tumors based on artificial intelligence algorithms. *Front Neurosci.* 2023, 17:1120781. [10.3389/fnins.2023.1120781](#)
17. Amin J, Anjum MA, Sharif M, Jabeen S, Kadry S, Moreno Ger P: A new model for brain tumor detection using ensemble transfer learning and quantum variational classifier. *Comput Intell Neurosci.* 2022, 2022:3236305. [10.1155/2022/3236305](#)
18. Vlahou A, Hallinan D, Apweiler R, et al.: Data sharing under the general data protection regulation: time to harmonize law and research ethics? *Hypertension.* 2021, 77:1029-35. [10.1161/HYPERTENSIONAHA.120.16340](#)
19. Whiting PF, Rutjes AW, Westwood ME: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011, 155:529-36. [10.7326/0003-4819-155-8-201110180-00009](#)
20. Yang B, Mallett S, Takwoingi Y, et al.: QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med.* 2021, 174:1592-9. [10.7326/M21-2234](#)

21. Reitsma JB, Rutjes AW, Whiting P, et al.: Assessing risk of bias and applicability . Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Deeks JJ, Bossuyt PM, Leeftang MM, Takwoingi Y (ed): John Wiley & Sons, NJ, USA; 2023. [10.1002/9781119756194.ch8](https://doi.org/10.1002/9781119756194.ch8)
22. Rameshwar CD, Patankar NS, Deshmukh YS: Machine learning techniques for brain tumor detection and classification using MRI. *NeuroQuantology*. 2022, 20:5963-78.
23. Kikuchi K, Togao O, Yamashita K, et al.: Comparison of diagnostic performance of radiologist- and AI-based assessments of T2-FLAIR mismatch sign and quantitative assessment using synthetic MRI in the differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, IDH-mutant and 1p/19q-codeleted. *Neuroradiology*. 2024, 66:333-41. [10.1007/s00234-024-03288-0](https://doi.org/10.1007/s00234-024-03288-0)
24. Anantharajan S, Gunasekaran S, Subramanian T, Venkatesh R: MRI brain tumor detection using deep learning and machine learning approaches. *Measurement: Sensors*. 2024, 31:101026.
25. Remzan N, Tahiry K, Farchi A: Advancing brain tumor classification accuracy through deep learning: harnessing radimagenet pre-trained convolutional neural networks, ensemble learning, and machine learning classifiers on MRI brain images. *Multimed Tools Appl*. 2024, 83:82719-47.
26. Santoso IB, Utama SN, Supriyono: A new voting of convolutional neural networks for brain tumor detection based on MRI images. *Int J Intell Eng Sys*. 2024, 17:212-27. [10.22266/ijies2024.0229.21](https://doi.org/10.22266/ijies2024.0229.21)
27. Kumar A, Singh AK, Singh A, Kumar V, Prakash S, Tiwari PK: An efficient framework for brain cancer identification using deep learning. *Multimed Tools Appl*. 2024, 83:65549-78.
28. Khan MSI, Rahman A, Debnath T, et al.: Accurate brain tumor detection using deep convolutional neural network. *CSBJ*. 2022, 20:4733-45.
29. Gajula S, Rajesh V: An MRI brain tumour detection using logistic regression-based machine learning model . *Int J Syst Assur Eng Manag*. 2024, 15:124-34.
30. Anagun Y: Smart brain tumor diagnosis system utilizing deep convolutional neural networks . *Multimed Tools Appl*. 2023, 82:44527-53. [10.1007/s11042-023-15422-w](https://doi.org/10.1007/s11042-023-15422-w)
31. Mohanasundari M, Chandrasekaran V, Anitha S: Brain tumor detection based on amended convolution neural network using MRI images. *KSII T Internet Info*. 2023, 17:2788-808.
32. Abdusalomov AB, Mukhiddinov M, Whangbo TK: Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers (Basel)*. 2023, 15:4172. [10.3390/cancers15164172](https://doi.org/10.3390/cancers15164172)
33. Thimma RB, Balaran VVSS: Detection of brain tumor from MRI images using deep dense neural network . *IJISAE*. 2023, 11:1031-40.
34. Asiri AA, Shaf A, Ali T, et al.: Exploring the power of deep learning: fine-tuned vision transformer for accurate and efficient brain tumor detection in MRI scans. *Diagnostics (Basel)*. 2023, 13:2094. [10.3390/diagnostics13122094](https://doi.org/10.3390/diagnostics13122094)
35. Yawen L, Du N, Fang Z, et al.: Identification of T2W hypointense ring as a novel noninvasive indicator for glioma grade and IDH genotype. *Cancer Imaging*. 2024, 24:80. [10.1186/s40644-024-00726-3](https://doi.org/10.1186/s40644-024-00726-3)
36. Zhang Y, Zhang H, Zhang H, Ouyang Y, Su R, Yang W, Huang B: Glioblastoma and solitary brain metastasis: differentiation by integrating demographic-MRI and deep-learning radiomics signatures. *J Magn Reson Imaging*. 2024, 60:909-20. [10.1002/jmri.29123](https://doi.org/10.1002/jmri.29123)
37. Manasa SM, Jeevitha BK: Automated brain tumor detection from MRI scans using deep convolutional neural networks. *IJRITCC*. 2023, 11:440-6.
38. Kalyani BJD, Meena K, Murali E, Jayakumar L, Saravanan D: Analysis of MRI brain tumor images using deep learning techniques. *Soft Computing*. 2023, 27:7535-42.
39. Hammad M, ElAffendi M, Ateya AA, Abd El-Latif AA: Efficient brain tumor detection with lightweight end-to-end deep learning model. *Cancers (Basel)*. 2023, 15:2837. [10.3390/cancers15102837](https://doi.org/10.3390/cancers15102837)
40. Emadi M, Dehkordi ZJ, Mobarakeh MI: Improving the accuracy of brain tumor identification in magnetic resonance imaging using super-pixel and fast primal dual algorithm. *IJE*. 2023, 36:505-12.
41. Aleid A, Alhussaini K, Alanazi R, Altwaimi M, Altwijri O, Saad AS: Artificial intelligence approach for early detection of brain tumors using MRI images. *Appl Sci*. 2023, 13:3808. [10.3390/app13063808](https://doi.org/10.3390/app13063808)
42. AlTahhan FE, Khouqeer GA, Saadi S, Elgarayhi A, Sallah M: Refined automatic brain tumor classification using hybrid convolutional neural networks for MRI scans. *Diagnostics (Basel)*. 2023, 13:864. [10.3390/diagnostics13050864](https://doi.org/10.3390/diagnostics13050864)
43. Nayan AA, Mozumder AN, Haque MR, et al.: A deep learning approach for brain tumor detection using magnetic resonance imaging. *IJECE*. 2023, 13:1.
44. Ghahramani M, Shiri N: Brain tumour detection in magnetic resonance imaging using Levenberg-Marquardt backpropagation neural network . *IET Image Process*. 2023, 17:88-103.
45. Rethemiotaki I: Brain tumour detection from magnetic resonance imaging using convolutional neural networks. *Contemp Oncol (Pozn)*. 2023, 27:230-41. [10.5114/wo.2023.135320](https://doi.org/10.5114/wo.2023.135320)
46. Raghuvanshi S, Dhariwal S: The VGG16 method is a powerful tool for detecting brain tumors using deep learning techniques. *Eng Proc*. 2023, 59:46.
47. Banik S, Barai NG, Shamrat FM: Blockchain integrated neural networks: a new frontier in MRI-based brain tumor detection. *IJACSA*. 2023, 14:11. [10.14569/IJACSA.2023.0141197](https://doi.org/10.14569/IJACSA.2023.0141197)
48. Benbakreti S, Benouis M, Roumane A, Benbakreti S: Impact of the data augmentation on the detection of brain tumor from MRI images based on CNN and pretrained models. *Multimed Tools Appl*. 2023, 83:39459-78.
49. Shahalinezhad S, Nooshyar M: Detection of brain tumor from MRI images base on deep learning technique using TL model. *AUSMT*. 2023, 13:1.
50. Kumar T, Kumar D, Singh G: Brain tumour classification using quantum support vector machine learning algorithm. *IETE J Res*. 2023, 70:4815-28. [10.1080/03772063.2023.2245350](https://doi.org/10.1080/03772063.2023.2245350)
51. Maquen-Niño GLE, Sandoval-Juarez AA, Veliz-La Rosa RA, et al.: Brain tumor classification deep learning model using neural networks. *ijOE*. 2023, 19:81-92.
52. Ullah Z, Odeh A, Khattak I, Hasan MA: Enhancement of pre-trained deep learning models to improve brain tumor classification. *Informatica*. 2023, 47:6. [10.31449/inf.v47i6.4645](https://doi.org/10.31449/inf.v47i6.4645)
53. Battalapalli D, Vidyadharan S, Prabhakar Rao BV, Yogeeswari P, Kesavadas C, Rajagopalan V: Fractal dimension: analyzing its potential as a neuroimaging biomarker for brain tumor diagnosis using machine

- learning. *Front Physiol.* 2023, 14:1201617. [10.3389/fphys.2023.1201617](https://doi.org/10.3389/fphys.2023.1201617)
54. Chandni SM, Kushwaha AKS: Effective brain tumor image classification using deep learning. *Natl Acad Sci Lett.* 2024, 47:257-60.
55. Mohsen S, Abdel-Rehim WMF, Emam A, Kasem HM: A convolutional neural network for automatic brain tumor detection. *Proc Eng Technol Innov.* 2023, 24:15-21.
56. Gajula S, Rajesh V: Deep learning based analysis of MRI images for brain tumor diagnosis. *IJACSA.* 2023, 14:2.
57. Kumar VV, Prince PGK: Magnitude normalized and OTSU intensity based brain tumor detection using magnetic resonance images. *IETE J Res.* 2023, 69:5079-89.
58. Stember JN, Shalu H: Reinforcement learning using Deep Q networks and Q learning accurately localizes brain tumors on MRI with very small training sets. *BMC Med Imaging.* 2022, 22:224. [10.1186/s12880-022-00919-x](https://doi.org/10.1186/s12880-022-00919-x)
59. Manoj KC, Dhas DAS: Automated brain tumor malignancy detection via 3D MRI using adaptive-3-D U-Net and heuristic-based deep neural network. *Multimedia Systems.* 2022, 28:2247-73.
60. Chattopadhyay A, Maitra M: MRI-based brain tumour image detection using CNN based deep learning method. *Neuroscience Informatics.* 2022, 2:100060.
61. Farnoosh R, Noushkaran H: Application of a modified combinational approach to brain tumor detection in MR images. *J Digit Imaging.* 2022, 35:1421-32. [10.1007/s10278-022-00653-4](https://doi.org/10.1007/s10278-022-00653-4)
62. Kumar BPS, Shaik SB, Mulam H: High-performance compression-based brain tumor detection using lightweight optimal deep neural network. *Adv Eng Sofw.* 2022, 173:103248.
63. Kaliannan S, Rengaraj A, Daniel AP: A novel fusion approach to detect brain tumor using machine learning for MRI images. *IIETA.* 2022, 39:1363-70.
64. Wang X, Li J, Sun J, et al.: A machine learning model based on unsupervised clustering multihabitat to predict the pathological grading of meningiomas. *Biomed Res Int.* 2022, 2022:8955227. [10.1155/2022/8955227](https://doi.org/10.1155/2022/8955227)
65. Yan Q, Li F, Cui Y, et al.: Discrimination between glioblastoma and solitary brain metastasis using conventional MRI and diffusion-weighted imaging based on a deep learning algorithm. *J Digit Imaging.* 2023, 36:1480-8. [10.1007/s10278-023-00838-5](https://doi.org/10.1007/s10278-023-00838-5)
66. Luo X, Yang Y, Yin S, et al.: False-negative and false-positive outcomes of computer-aided detection on brain metastasis: secondary analysis of a multicenter, multireader study. *Neuro Oncol.* 2023, 25:544-56. [10.1093/neuonc/noac192](https://doi.org/10.1093/neuonc/noac192)
67. Salman LA, Hashim AT, Hasan AM: Automated brain tumor detection of MRI image based on hybrid image processing techniques. *Telkomnika.* 2022, 20:4.
68. Farias MCQ, de Castro Oliveira PH, dos Santos Lopes G, Miosso CJ, Lima JA: The influence of magnetic resonance imaging artifacts on CNN-based brain cancer detection algorithms. *Comput Math Model.* 2022, 33:211-29.
69. Duan CF, Li N, Li Y, Liu F, Wang JC, Liu XJ, Xu WJ: Comparison of different radiomic models based on enhanced T1-weighted images to predict the meningioma grade. *Clin Radiol.* 2022, 77:e302-7. [10.1016/j.crad.2022.01.039](https://doi.org/10.1016/j.crad.2022.01.039)
70. Singh V, Gourisaria MK, Harshvardhan GM, et al.: Diagnosis of intracranial tumors via the selective CNN data modeling technique. *Appl Sci.* 2022, 12:2900.
71. Raghavendra U, Gudigar A, Rao TN, et al.: Feature-versus deep learning-based approaches for the automated detection of brain tumor with magnetic resonance images: a comparative study. *IMA.* 2022, 32:501-16.
72. Huang Z, Zhao Y, Liu Y, Song G: AMF-Net: an adaptive multisequence fusing neural network for multi-modality brain tumor diagnosis. *BSPC.* 2022, 72:103359.
73. Zailan ZNI, Mostafa SA, Abdulmageed AI, Baharum Z, Jaber MM, Hidayat R: Deep learning approach for prediction of brain tumor from small number of MRI images. *JOIV.* 2022, 6:2.
74. Asiri AA, Aamir M, Shaf A, et al.: Block-wise neural network for brain tumor identification in magnetic resonance images. *CMC.* 2022, 73:5735-53. [10.32604/cmc.2022.031747](https://doi.org/10.32604/cmc.2022.031747)
75. Liu Y, Li T, Fan Z, et al.: Image-based differentiation of intracranial metastasis from glioblastoma using automated machine learning. *Front Neurosci.* 2022, 16:855990. [10.3389/fnins.2022.855990](https://doi.org/10.3389/fnins.2022.855990)
76. Tiwari P, Pant B, Elarabawy MM, Abd-Elnaby M, Mohd N, Dhiman G, Sharma S: CNN based multiclass brain tumor detection using medical imaging. *Comput Intell Neurosci.* 2022, 2022:1830010. [10.1155/2022/1830010](https://doi.org/10.1155/2022/1830010)
77. Han-Trong T, Nguyen Van H, Nguyen Thi Thanh H, Tran Anh V, Nguyen Tuan D, Vu Dang L: An efficient method for diagnosing brain tumors based on MRI images using deep convolutional neural networks. *Appl Comput Intell Soft Comput.* 2022,
78. Bathla G, Priya S, Liu Y, et al.: Radiomics-based differentiation between glioblastoma and primary central nervous system lymphoma: a comparison of diagnostic performance across different MRI sequences and machine learning techniques. *Eur Radiol.* 2021, 31:8703-13. [10.1007/s00330-021-07845-6](https://doi.org/10.1007/s00330-021-07845-6)
79. Saeidifar M, Yazdi M, Zolghadrasli A: Performance improvement in brain tumor detection in MRI images using a combination of evolutionary algorithms and active contour method. *J Digit Imaging.* 2021, 34:1209-24. [10.1007/s10278-021-00514-6](https://doi.org/10.1007/s10278-021-00514-6)
80. Das TK, Roy PK, Uddin M, Srinivasan K, Chang C-Y, Syed-Abdul S: Early tumor diagnosis in brain MR images via deep convolutional neural network model. *CMC.* 2021, 68:2413-29.
81. Malik N, Geraghty B, Dasgupta A, et al.: MRI radiomics to differentiate between low grade glioma and glioblastoma peritumoral region. *J Neurooncol.* 2021, 155:181-91. [10.1007/s11060-021-03866-9](https://doi.org/10.1007/s11060-021-03866-9)
82. Thakur S, Doshi J, Pati S, et al.: Brain extraction on MRI scans in presence of diffuse glioma: multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. *Neuroimage.* 2020, 220:117081. [10.1016/j.neuroimage.2020.117081](https://doi.org/10.1016/j.neuroimage.2020.117081)
83. Hu J, Zhao Y, Li M, et al.: Machine learning-based radiomics analysis in predicting the meningioma grade using multiparametric MRI. *Eur J Radiol.* 2020, 131:109251. [10.1016/j.ejrad.2020.109251](https://doi.org/10.1016/j.ejrad.2020.109251)
84. Katouli M, Rahmani AE: Brain tumor diagnosis in MRI images using image processing techniques and pixel-based clustering. *IIETA.* 2020, 37:291-300.
85. Chaudhary A, Bhattacharjee V: An efficient method for brain tumor detection and categorization using MRI

- images by K-means clustering & DWT. *Int J Inf Technol*. 2020, 12:141-8.
86. Isselmou AEK, Xu G, Shuai Z, Saminu S, Javadi I, Ahmad IS: Brain tumor identification by convolution neural network with fuzzy c-mean model using mr brain images. *Int J Circ Syst Signal Process*. 2020, 14:1096-102.
87. Atici MA, Sagioglu S, Celtikci P, Ucar M, Borcek AO, Emmez H, Celtikci E: A novel deep learning algorithm for the automatic detection of high-grade gliomas on T2-weighted magnetic resonance images: a preliminary machine learning study. *Turk Neurosurg*. 2020, 30:199-205. [10.5137/1019-5149.JTN.27106-19.2](#)
88. Tamije Selvy P, Anitha M: Automated brain tumor detection in MRI images using efficient deep learning methods. *IJITEE*. 2019, 8:286-91. [10.35940/ijitee.K1324.0981119](#)
89. Pandiselvi T, Maheswaran R: Efficient framework for identifying, locating, detecting and classifying MRI brain tumor in MRI images. *J Med Syst*. 2019, 43:189. [10.1007/s10916-019-1253-1](#)
90. Qin C, Hu W, Wang X, Ma X: Application of artificial intelligence in diagnosis of craniopharyngioma. *Front Neurol*. 2022, 12:752119. [10.3389/fneur.2021.752119](#)
91. Sengupta A, Ramaniharan AK, Gupta RK, Agarwal S, Singh A: Glioma grading using a machine-learning framework based on optimized features obtained from T(1) perfusion MRI and volumes of tumor components. *J Magn Reson Imaging*. 2019, 50:1295-306. [10.1002/jmri.26704](#)
92. Li M, Kuang L, Xu S, Sha Z: Brain tumor detection based on multimodal information fusion and convolutional neural network. *IEEE Access*. 2019, 7:180134-46.
93. Chen X, Tong Y, Shi Z, et al.: Noninvasive molecular diagnosis of craniopharyngioma with MRI-based radiomics approach. *BMC Neurol*. 2019, 19:6. [10.1186/s12883-018-1216-z](#)
94. Kaur K, Walia GK, Kaur J: Neural network ensemble and Jaya algorithm based diagnosis of brain tumor using MRI images. *J Inst Eng (India): Ser B*. 2018, 99:509-17. [10.1007/s40031-018-0355-3](#)
95. De Looze C, Beausang A, Cryan J, et al.: Machine learning: a useful radiological adjunct in determination of a newly diagnosed glioma's grade and IDH status. *J Neurooncol*. 2018, 139:491-9. [10.1007/s11060-018-2895-4](#)
96. Kavin Kumar K, Meera Devi T, Maheswaran S: An efficient method for brain tumor detection using texture features and SVM classifier in MR images. *APOCP*. 2018, 19:2789-94. [10.22034/APJCP.2018.19.10.2789](#)
97. El-Hobarty ES: Intelligent methodology for brain tumors classification in magnetic resonance images. *IJCV*. 2017, 11:
98. Preetha R, Bhanumathi R, Suresh GR: Immune feature weighted least-squares support vector machine for brain tumor detection using MR images. *IETE Journal of Research*. 2016, 6:873-84.
99. Yamashita K, Yoshiura T, Arimura H, et al.: Performance evaluation of radiologists with artificial neural network for differential diagnosis of intra-axial cerebral tumors on MR images. *AJNR Am J Neuroradiol*. 2008, 29:1153-8. [10.3174/ajnr.A1037](#)
100. Sridevi A, Sabitha T: Detection of brain tumor using back propagation algorithm through MRI. *Indian J Public Health Res Dev*. 2019, 10:1097-102. [10.5958/0976-5506.2019.00444.3](#)
101. Laukamp KR, Shakirin G, Baessler B, et al.: Accuracy of radiomics-based feature analysis on multiparametric magnetic resonance images for noninvasive meningioma grading. *World Neurosurg*. 2019, 132:e366-90. [10.1016/j.wneu.2019.08.148](#)
102. Chen X, Wu Y, Zhao G, Wang M, Gao W, Zhang Q, Lin Y: Automatic histogram specification for glioma grading using multicenter data. *J Healthc Eng*. 2019, 2019:9414937. [10.1155/2019/9414937](#)
103. Khan MA, Park H: A convolutional block base architecture for multiclass brain tumor detection using magnetic resonance imaging. *Electronics*. 2024, 13:364.
104. Biratu ES, Schwenker F, Ayano YM, Debelee TG: A survey of brain tumor segmentation and classification algorithms. *J Imaging*. 2021, 7:179. [10.3390/jimaging7090179](#)