Review Article

Artificial intelligence and machine learning in neurosurgery: A review of diagnostic significance and treatment planning efficiency

Rani G. Ahmad

Department of Radiology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

This review analyzes the significance of artificial intelligence (AI) and deep learning (DL) approaches used in radiology in neurosurgery patients and compares AI applications with human models to determine the applicability of AI in disease diagnosis, decision-making, and outcome prediction. A systematic review was conducted from 1997 to 2020 from the PubMed (MEDLINE) database. The search strategy adhered to guidelines outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. The keywords used for the literature search included "Deep learning," "Neurosurgery," "Artificial Intelligence," "Brain," "Magnetic resonance imaging-MRI Brain," and "Machine learning." The studies focusing on the significance of DL and comparing AI applications with radiologists or clinical experts to enhance diagnostic protocols were included, whereas non-English articles, animal studies, articles lacking full text, and publications such as commentaries, technical notes, abstracts, editorials, opinions, and letters were excluded. A total of 24 articles were included in the review. The P value was observed in 44 out of 63 outcome measures (70%), out of which in 26 out of 63 outturn measures, artificial application subset machine learning (ML) has a significant edge over clinical diagnosis ($P < 0.05$). The review highlights the potential impact of AI-driven advancements in clinical radiology on enhancing treatment plans for neurosurgery patients, emphasizing the benefits of early intervention, cost reduction, time-saving approaches, and judicious health-care resource utilization. The study's limitations include potential constraints in identifying relevant literature due to the selected search scope and inclusion criteria, not including studies published outside the specified timeframe and database, and a small number of included studies. Consequently, there is a risk of overlooking innovative methodologies or ground-breaking studies contributing to a more comprehensive understanding of AI applications in neurosurgery. Furthermore, the exclusion of certain publication types, such as commentaries, and conference papers may limit the diversity of different perspectives. However, the study highlights the potential of ML in neurosurgery and the importance of addressing variability in study design, patient populations, and outcome measures in future research to enhance the applicability of AI-driven approaches in clinical practice. It is imperative to recognize and address these challenges to understand the opportunities and limitations inherent in the integration of AI in neurosurgical practice.

Keywords: Artificial intelligence, electrocorticography, machine learning, neurosurgery

Address for correspondence: Dr. Rani G. Ahmad, Department of Radiology, King Abdulaziz University, P. O. Box: 80215, Jeddah 21589, Saudi Arabia.
E-mail: rahmad@kau.edu.sa

Submitted: 05-Dec-2022 **Revised:** 22-Apr-2024 **Acceptance:** 10-May-2024 **Published:** 05-Aug-2024

Access this article online	
Quick Response Code:	Website: https://journals.lww.com/wajr
	DOI: 10.4103/wajr.wajr_32_22

How to cite this article: Ahmad RG. Artificial intelligence and machine learning in neurosurgery: A review of diagnostic significance and treatment planning efficiency. West Afr J Radiol 2023;30:29-40.

INTRODUCTION

Artificial intelligence (AI) is a broader discipline and fast-growing field that enables machines to mimic human cognitive behavior. These machines require human intelligence, computer algorithms, and recognizable rule-based systems to display properties of intelligence through driving knowledge from data.[1] Machine learning (ML), as a discipline of AI, is a branch of data science involving statistical model application to data using computers and enabling algorithms (computer programs) to learn associations of predictive power from existing data without explicit programming to forecast new data points.[2] The integration of AI and ML holds significant promise in neurosurgery, where precision and timely decision-making are dominant.

The main objective of this systematic review is to compare ML available algorithms performance in neurosurgery patients in comparison to clinical experts to gain insight into recent advancements in AI approaches to further strengthen neurosurgical patients' perioperative care decision-making. The implications of such advancements extend far beyond solely the technical aspects, rather they bear the potential to create profound enhancements in patient outcomes, reductions in mortality rates, and advancements in neurosurgical practices. Through a comprehensive examination of ML algorithms, this review aims to explain the extent to which AI-driven approaches can enhance and refine perioperative care decision-making processes in neurosurgery. By delineating the strengths and limitations of AI applications in comparison to conventional clinical expertise, this investigation would provide critical insights into the transformative potential of AI in enhancing the quality, efficiency, and efficacy of neurosurgical patient care.

LITERATURE REVIEW

ML algorithms can be further divided into supervised, reinforcement, and unsupervised learning algorithms. Supervised learning involves computer program training to associate data input and output through the output of interest analysis defined by the supervisor (ground truth) and label input data with required output. In unsupervised learning, data do not require explicit labeling and, based on the underlying distribution model, produces data representation.[3] Recent machine learning (ML) algorithms, specifically supervised learning algorithms, include artificial neural networks (ANNs), support vector machines (SVMs), decision trees, K-nearest neighbors, linear discriminant analysis (LDA), and Naïve Bayes. In contrast, fuzzy C-means (FCM) is an unsupervised learning algorithm. Deep learning (DL), a subclass of ML, utilizes deep neural networks with many hidden layers. DL methods enable the use of multiple layers to process large amounts of data, allowing machines to discover the necessary representations for tasks such as classification and detection. It is made possible due to recent computational advancements and led to fundamental advancements in ML.[4,5] The relationship of AI with its subsets: ML, DL, and neural networks is shown in Figure 1.

Since ML requires a large data set without explicit programming, in medical research and clinical neurosurgical care, ML still requires comprehensive validation before implementation. Neurosurgery research and clinical practice are ideal for ML model application as complex therapeutic and diagnostic modalities generate a huge amount of data with a rich assortment which is ideal for AI tools, especially ML models to improve neurosurgical care through improved, précised, and efficient perioperative predictive analysis through integration of all patient-relevant factors including extraction of deep features such as genomic data histological or radiological images, etc., in a way which is more complicated and complex for clinician to integrate risk factors and outcome predictors into single prognosis.[6-8]

METHODS

Search approach

This systematic review adheres to the guidelines outlined by the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" to comprehensively identify

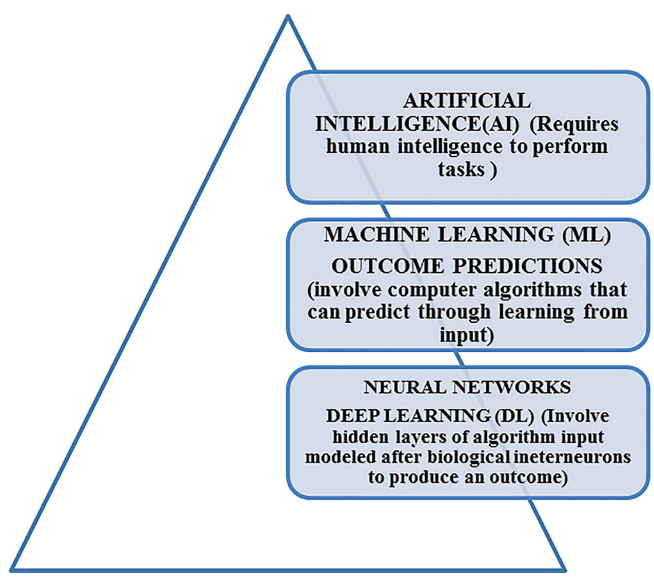

Figure 1: Relationship of artificial intelligence with its subsets machine learning, neural networks, and deep learning

relevant studies about the integration of DL models into neurosurgery and their comparison with the expertise of radiologists or clinical professionals to improve diagnostic protocols.

The search strategy primarily involves querying the PubMed and MEDLINE databases. The research question guiding the literature search has been refined to focus on assessing "The significance and impact of artificial intelligence in neurosurgical patient care, specifically examining deep learning models, and comparing their effectiveness with clinical experts." The search strategy employs a comprehensive set of terms relevant to the study domain, including "Deep learning," "Neurosurgery," "Artificial Intelligence," "Brain," "Magnetic resonance imaging-MRI Brain," and "Machine learning."

Inclusion and exclusion criteria

To ensure the inclusion of pertinent studies, eligibility criteria have been meticulously established. The search encompasses articles published between 1997 and 2020, with a strict language restriction to English. Studies that were included focused exclusively on neurosurgical patients, specifically targeting those that compare AI applications using machine learning (ML) with clinical expert practices. Neurosurgical patients were categorized as individuals eligible for neurosurgical treatment at any stage of their illness. There are no predefined limitations regarding disease diagnosis, screening, prognosis, treatment, or outcome.

Exclusion criteria encompass articles in languages other than English, animal studies, conference papers, books/book chapters, and those lacking full-text availability. In addition, the search strategy is augmented by manually screening the reference lists of potentially eligible articles. The duplicated studies were removed and the titles/abstracts were reviewed to identify papers that are pertinent to the present study's topic. Table 1 outlines the PICO elements guiding the search strategy. It helps clarify the specific aspects considered in the search strategy, guiding the selection of relevant studies.

Table 1: Population, Intervention, Comparator, and Outcome elements for search strategy

Conditions	Qualifications
Main question	Significance and impact of AI in neurosurgery patients and its comparison with clinicians to understand current best practices
Population	Diagnostics images obtained from neurosurgery human subjects
Intervention	AI algorithms found a diagnostic model
Comparator	Brain MRI, CT, clinical examination by clinicians, etc.
Outcome	Outcome measures of proposed AI model (ML model) in terms of AUC, sensitivity, accuracy, specificity, PPV, *P*-value, validation method

AI – Artificial intelligence, MRI – Magnetic resonance imaging, CT – Computerized tomography, ML – Machine learning, AUC – Area under curve, PPV – Positive predictive value

Assessing eligibility

In assessing the eligibility of studies for inclusion in this systematic review, stringent criteria were established to ensure both relevance and quality. First, all clinical studies providing data on the significance and impact of AI in neurosurgery patients were considered eligible for inclusion. These studies were required to directly address the application of AI in neurosurgical patient care, encompassing areas such as diagnostic accuracy, treatment planning, or outcome prediction. In addition, to maintain consistency and accuracy in evaluation, only articles written in English were deemed suitable for inclusion, reflecting the authors' proficiency in the language. This criterion aimed to facilitate thorough assessment and interpretation of study content. Conversely, certain types of publications, including commentaries, technical notes, abstracts, editorials, opinions, and letters, were excluded from consideration. Such formats typically lack original research findings or provide insufficient data for systematic review purposes. Furthermore, research involving biomechanical assessments on animals and *in vitro* studies were excluded, as they did not align with the specific focus on clinical applications of AI in neurosurgery patients. Through the application of these inclusion and exclusion criteria, the systematic review sought to ensure the selection of only the most relevant and high-quality studies, thereby enhancing the validity and reliability of the review's findings.

Study selection and data collection

The quality analysis of selected studies was ranked by two assessors individualistically, and in case of any disparity, it was fixed through dialog. The data for each included study consisted of the following details author and year of publication, output and input features, outcome measures for machine model and clinician model (natural intelligence), *P* value, and validation methods. Due to data heterogeneity, quantitative synthesis was considered inappropriate. Instead, a qualitative assessment of outcome risk of bias was conducted using narrative analysis.

Initially, titles of articles were manually screened, and those relevant to the research topic were considered for further evaluation. Subsequently, if the abstracts corresponded with the study's focus, the full texts of the articles were retrieved for thorough examination. Articles lacking full-text availability were excluded from the analysis at this stage. In addition, a manual screening of bibliographies was conducted to identify additional relevant studies. All articles identified through these procedures underwent comprehensive evaluation, with their eligibility for inclusion being deliberated among the researchers to ensure consensus.

Data items

Two assessors independently conducted a quality analysis of the selected studies, resolving any discrepancies through dialog and consensus building to ensure the reliability and validity of the quality assessment. Data extraction from each included study was thorough, capturing key details such as the author and year of publication, output and input features, outcome measures for both machine and clinician models, *P* value, and validation methods used. This comprehensive approach aimed to provide a detailed understanding of each study's methodology and findings, facilitating a robust analysis in line with the review's objectives. Given the heterogeneity of the collected data, quantitative synthesis was deemed inappropriate, and instead, a qualitative assessment of the studies' outcomes and risk of bias was conducted through narrative synthesis. The study selection process began with the manual screening of article titles, followed by the examination of abstracts corresponding with the study's focus. Full-text retrieval was conducted for articles meeting the inclusion criteria, while those lacking full-text availability were excluded. In addition, a manual screening of bibliographies was performed to identify any additional relevant studies. Throughout the selection process, all articles underwent a comprehensive evaluation to determine their eligibility for inclusion, with the research team deliberating to ensure consensus and minimize bias.

Data analysis

The included studies' applicability and risks of biases were evaluated using the "Quality Assessment of Diagnostic Accuracy Studies"-2 tool. This evaluation assessed the risk of bias in four domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed for risk of bias and applicability, with judgments categorized as "low," "unclear," or "high" risk. Disagreements between assessors were resolved through discussion.

RESULTS

In this systematic review, a total of 6652 citations were retrieved during the initial search from both PubMed (MEDLINE) 1452 citations and from additional database 5200 citations. However, after detailed screening, 1100 nonduplicate citations were identified. Five thousand four hundred and sixty-five articles were excluded based on their abstract and titles, resulting in 87 articles being studied for their full-text details, 53 articles were excluded after full-text screening, and 10 articles were excluded during data extraction as no comparison with clinical experts was found. Finally, a total of 24 articles were included for the review to assess the significance of AI applications in neurosurgical patients' diagnosis, prognosis, and preoperative preparation for treatment. The flow diagram of the search strategy and inclusion criteria for this review study is illustrated in Figure 2.

Considering the statistical measures, the most frequent measure used was accuracy 12 (50% of the studies), followed by area under curve (AUC) 10 (41.7%), sensitivity 8 (33%), and specificity 12 (20.8%), respectively, both for the ML and clinical model [Figure 3]. The *P* value was observed in 44 outturn measures including accuracy, sensitivity, specificity, AUC, positive predictive value (PPV), high-grade glioma, false discovery rate, negative predictive value, digital span forward [DSF], F-measure, speed, dice similarity coefficient [DSC], and low-grade glioma) out of overall 63 outcome measures (70%), out of which in 26 out of 63 outturn measures, artificial application subset ML has a significant edge over clinical diagnosis ($P < 0.05$) [Figures 4 and 5].

Table 2 shows a description of the machine models used to evaluate the significance of AI in clinical radiology. The table evaluated recruited studies for diagnostic input tools used, their ML model used, clinical expert outcome, the outcome of the ML model, validation methodology, statistical variables, and diagnostic or grading criteria. Out of 24 studies, 12 studies have emphasized the diagnostic capability of AI applications using magnetic resonance imaging (MRI) brain as a primary diagnostic parameter in comparison to clinical experts. Out of these 12 studies, 4 studies[9-12] focused on tumor diagnostic cataloging among the pediatric population through differentiation of "posterior fossa tumor." The major input characteristics applied to these were brain MRI along with age and gender in Table 2.

DISCUSSION

The objective of this systemic review was to evaluate the significance and impact of AI analogous enactment "ML models" in neurosurgery patient pollution to help attain better treatment plans. The use of AI applications including ML and "CNNs" if provided with adequate teaching datasets could result in self-explanatory patient beneficiary performances along with expertise from clinical

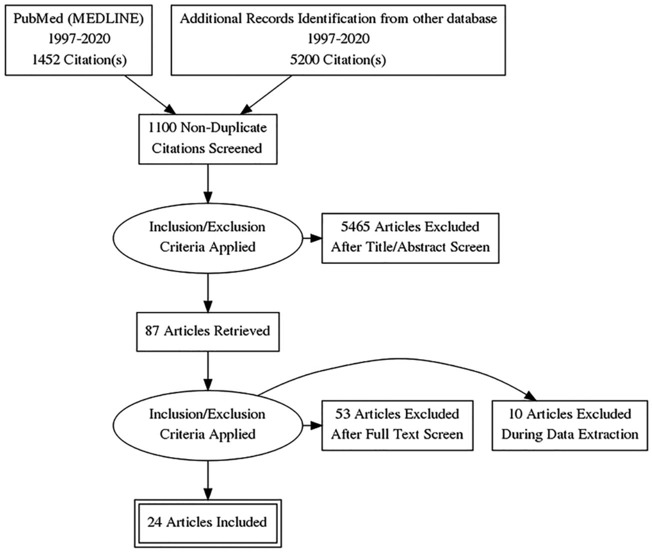

Figure 2: Flow diagram for the searches and inclusion criteria in the study

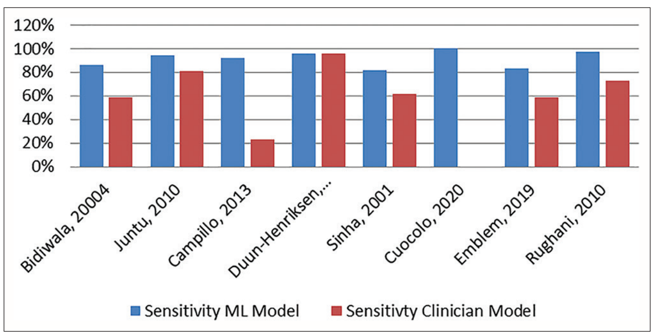

Figure 4: Sensitivity comparison among machine learning and clinician model

experts. This ultimately improves diagnostic accuracy which ultimately aids in correct decisions and treatment plans for patients.

AI possesses the capability to enhance the results for patients by enhancing the skills of neurosurgeons, thereby advancing the accuracy of diagnoses and predictions, and refining the choices made during surgical operations.[13] By integrating AI into various treatments, whether they involve direct intervention or not, neurosurgeons can offer optimal care to those under their treatment.

ML techniques have found extensive use in analyzing MRI data for glioma studies, proving highly beneficial.[14] The ML model used in four of the studies was ANN. Major outcome measures monitored were AUC, sensitivity, specificity, accuracy, and PPV. Out of the four mentioned studies, two by Kitajima *et al.*[9] and Yamashita *et al.*[10] showed that the AUC values were almost similar in both the ML models (AUC: 0.99, 0.95) and clinician models (AUC: 0.91, 0.9). This indicates that the AUC results were comparable between the ML and clinician models. However, sensitivity, specificity, PPV, and accuracy were found to be improved in the ML models compared to the clinician models. Therefore, we concluded that the ML model performed significantly better in terms of accuracy ($P < 0.001$) and no significant difference was found among both models in terms of sensitivity ($P = 0.074$), specificity ($P = 77$), and PPV ($P = 17$) better in posterior fossa tumor differentiation in pediatrics.[11,12]

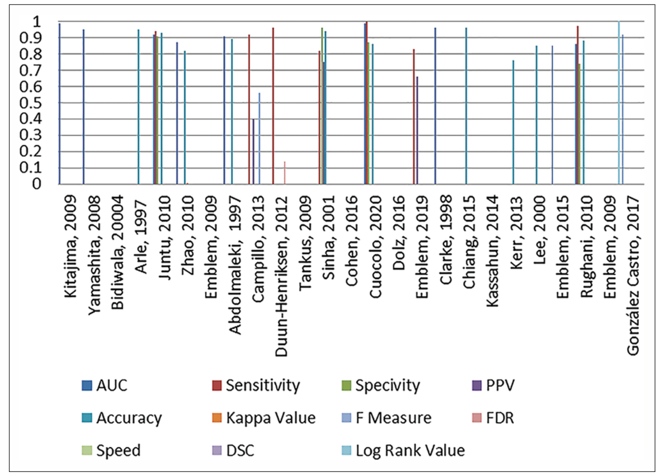

Figure 3: Outcome measures for machine learning model

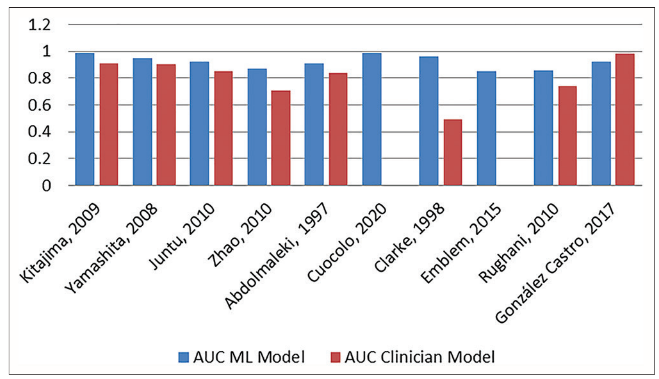

Figure 5: Area under curve comparison among machine learning and clinician model

ML technology effectively forecasts outcomes and aids in clinical decision-making within the field of neurosurgery.[15] Humans and machines can collaborate effectively to leverage the latest advancements in AI technology to elevate the standard of health-care provision across various stages, including image acquisition, processing, and interpretation, as well as patient allocation for surgeries, intraoperative procedures, postoperative monitoring, and enhancing access to top-tier health-care services.[16] Moreover, the application of AI can be extended to address neuromuscular and neurodegenerative disorders, such as Parkinson's disease, currently managed through medication

Table 2: Description of artificial intelligence applications (machine learning model) performance in comparison to clinicians, validation technique, machine learning model for diagnosis, preoperative plan, segmentation, localization, and outcome measure

Author	Year	Outcome	Input characteristics	ML models used	Outcome measures	ML model	Outcome in clinician model	Sensitivity	Specificity	*P*	Validation technique	Diagnosis/grading criteria
					Diagnostic tumor classification							
Kitajima *et al*.	2009	Distinguishable Rathke's cleft pituitary adenoma, craniopharyngioma	Age, MRI	ANN	AUC	0.99	0.91	NA	NA	NA	LOOCV	Histological diagnosis
Yamashita *et al*.	2008	Distinguishable glioma Grade II–V, lymphoma, brain metastases, malignant	Age, history of brain tumor, MRI	ANN	AUC	0.95	0.9	NA	NA	NA	LOOCV	Histological diagnosis
Bidiwala *et al*.	2004	Differentiate pediatric posterior fossa tumors: Medulloblastoma, cerebellar astrocytoma, ependymoma	Age, gender, symptoms, signs, CT, MRI	ANN	Sensitivity, specificity, PPV	73%–86%, 86%–93%, 73%–86%	57%–59%, 82%–83%, 62%–63%			0.074, 77, 17	CV (NOS)	Histological diagnosis
Arle *et al*.	1997	Differentiate pediatric posterior fossa tumors: Astrocytoma, PNET, ependymoma/other	Age, gender, MRI, MRS	ANN	Accuracy	95%	73%			<0.001	5-FCV	Histological diagnosis
					Tumor grading							
Juntu *et al*.	2010	Differentiate between benign and MRI malignant soft-tissue tumors including neural tumors	MRI	SVM, ANN	Accuracy, sensitivity, specificity, AUC	93%, 94%, 91%, 0.92	90% 81% 92% 0.85			0.61 0.009 1.00 NA	10-FCV	Histological diagnosis
Zhao *et al*.	2010	Classify glioma into Grades I–IV	Age, MRI	SVM	Accuracy overall Accuracy LGG Accuracy HGG Kappa value AUC	82% 82% 85% 0.68 0.870	65% 62% 66% 0.47 0.71			0.001 09 0.008 NA 0.004	5-FCV	Histological diagnosis
Emblem *et al*.	2009	Classify glioma into Grade I–IV	MRI	FCM	AUC	NA	NA			0.56–0.97		Histological diagnosis
Abdolmaleki *et al*.	1997	Differentiate between low and high-grade astrocytomas	MRI	ANN	Accuracy AUC *r*	89% 0.91 0.87	80% 0.84 0.56			0.003 <0.001 NA		Histological diagnosis
					Other applications							
Campillo *et al*.	2013	Detection of surgical site infection	Free text of HER	NA	Sensitivity PPV F-measure	92% 40% 0.56	23% 100% 0.38			<0.001 <0.001		Patients identification through experts, ICD-10 code, or NLP
Duun-Henriksen *et al*.	2012	Automated seizure detection in epilepsy patients	iEEG	SVM	Sensitivity FDR	96% 0.14	96% 0.18			1.00 1.00		NA
Tankus *et al*.	2009	Classify spike clusters in epilepsy	iEEG	LDA	Accuracy	91%–92%	38%–69%			<0.001	LOOCV	Synthetic database
Sinha *et al*.	2001	Predict the presence of CT abnormalities and DSF in pediatric TBI patients	Age, gender, symptoms, signs, history of trauma	ANN	Accuracy Sensitivity Specificity PPV NPV LR+	94% 82% 96% 75% 97% 21	92% 62% 96% 72% 95% 17.3			<0.05 <0.001 1.00 0.73 0.25 NA		Imaging CT

Contd...

Table 2: Contd...

Author	Year	Outcome	Input characteristics	ML models used	Outcome measures	ML model	Outcome in clinician model	Sensitivity	Specificity	P	Validation technique	Diagnosis/grading criteria
Other applications												
					LRAUC DP Sensitivity DSF	0.18 0.93 2.75 79%	0.39 0.90 2.10 54%			NA NA NA <0.001		
Prospective planning												
Cohen *et al*.	2016	Identify surgical candidates among pediatric epilepsy patients using NLP	Free text of HER	SVM, NB	F-measure	0.77–0.82	0.71			<0.001	10-FCV	Clinical outcome
Cuocolo *et al*.	2020	Pituitary macroadenoma	MRI and transsphenoidal surgery		Accuracy Sensitivity Specificity AUC	86% 100% 87% 0.99				NA		
Segmentation												
Dolz *et al*.	2016	Segmentation of brain stem in trigeminal neuralgia, brain metastases, brainstem cavernoma patients	MRI	SVM, DL	DSC pVD Speed	0.88–0.92 3.1%–7.2% 36–40 s	0.84–0.90 19%–39% 20.2 min			<0.001 <0.001 <0.001		
Emblem *et al*.	2009	Tumor segmentation in glioma patients Grade I–IV	MRI	FCM	Sensitivity LGG Sensitivity HGG PPV LGG PPV HGG	83% 69% 66% 73%	59% 57% 89% 87%			<0.001 0.005 <0.001 004		
Clarke *et al*.	1998	Tumor segmentation in glioma patients Grade III–IV	MRI	FCM	AUC	0.96	0.28–0.49			NA		
Localizing epileptogenic zone												
Chiang *et al*.	2015	Differentiate L-TLE and R-TLE	fMRI	QDA	Accuracy	96%	67%			0.05	LOOCV	VEEG
Kassahun *et al*.	2014	Differentiate TLE and E-TLE	Symptoms, genetics	SVM, GS, DT	Accuracy	66%–78%	56%–78%			0.41	CV (NOS)	After the operation disease remission
Kerr *et al*.	2013	Differentiate L-TLE, R-TLE, and NES	FDG-PET	ANN	Accuracy	76%	80%			0.53	LOOCV	Physical and neuro examination, VEEG
Lee *et al*.	2000	Localization of epileptogenic zone	FDG-PET	LDA, ANN	Accuracy	85%	81%			0.43		Two physicians census
Outcome prediction												
Emblem *et al*.	2015	Predict survival in glioma Patients grade II–IV	MRI	SVM	AUC 6 months AUC 1 year AUC 2 years AUC 3 years	0.794 0.762 0.806 0.851	0.50–0.66 0.50–0.66 0.50–0.66 0.50–0.66			<0.01 <0.01 <0.01 <0.01	10-FCV	Survival
Rughani *et al*.	2010	Predict in-hospital survival in TBI patients	Age, gender, vital parameters	ANN	Accuracy Sensitivity Specificity AUC	88% 97% 74% 0.86	72% 73% 75% 0.74			<0.001 <0.001 40 <0.001		Hospital survival
Emblem *et al*.	2009	Predict survival in glioma patients Grade I–IV	MRI	FCM	Log-rank value	14.4	10.6–12.8			NA		Survival

Contd...

Table 2: Contd...

Author	Year	Outcome	Input characteristics	ML models used	Outcome measures	ML model	Outcome in clinician model	Sensitivity	Specificity	P	Validation technique	Diagnosis/grading criteria
					Outcome prediction							
González Castro *et al*.	2017	SVM classifier binary classifier of the burden of enlarged PVS as low or high	MRI		AUC	0.9265	0.9813			NA	A stratified 5-FCV repeating ten times	Hospital survival

SVM – Support vector machine, PVS – Perivascular spaces, MRI – Magnetic resonance imaging, TBI – Traumatic brain injury, DSF – Digital span forward, CT – Computerized tomography, NLP – Natural language processing, L-TLE – Left temporal lobe epilepsy, R-TLE – Right temporal lobe epilepsy, E-TLE – Extratemporal temporal lobe epilepsy, MRS – Magnetic resonance spectroscopy, iEEG – Intracranial electroencephalography, fMRI – Functional magnetic resonance imaging, PET – Positron emission tomography, ANN – Artificial neural network, ML – Machine learning, FCM – Fuzzy C-mean, NA – Not available, LDA – Linear discriminant analysis, NB – Naïve Bayes, DL – Deep learning, QDA – Quadratic discriminant analysis, DT – Decision tree, AUC – Area under curve, PPV – Positive predictive value, LGG – Low-grade glioma, HGG – High-grade glioma, NPV – Negative predictive value, DSC – Dice similarity coefficient, pVD – Percentage volume difference, FCV – Fold cross-validation, VEEG – Video-electroencephalography, ICD-10 – International Classification of Diseases-10, NOS – Newcastle-ottawa scale, LR – Likelihood ratio, FDR – False discovery rate, LRAUC – Logistic regression area under the curve, DP – Data processing, GS – Gold tandard, LOOCV – Leave-one-out cross-validation

and deep brain stimulation.[17] In addition, AI can contribute to advancing our understanding of molecular cell biology, including areas like the subcellular trafficking of cargoes in individual neurons.[18,19]

Among all the included studies, 4 studies were categorized under tumor grading.[20-23] In these four studies, MRI techniques were used for classification, and ML models were evaluated against clinicians. Two studies showed improved outcomes for the ML models: (i) One study using an SVM model demonstrated a significant correlation with the clinician model in terms of accuracy ($P = 0.001$) and Kappa ($P = 0.004$),[21] and (ii) another study using an ANN model showed statistically significant results for both accuracy ($P = 0.003$) and AUC ($P = 0.001$).[23] The other two studies showed a nonsignificant correlation between the ML models and clinician models when using SVM and ANN models, with $P = 1.00$ for specificity and $P = 0.009$ for sensitivity,[20] and $P = 0.56–0.97$[22] for various applications. Additionally, three studies did not use radiological diagnostic tools but instead used ML based on 'intracranial electroencephalography' waves, which reflected improved accuracy in distinguishing epileptic patients' single and multiunit spikes. In the same way, Sinha *et al*.[24] predicted "computerized tomography" anomalies through ANN technique in pediatric "traumatic brain injury" (TBI) patients. Hence, the study found a significant correlation between the improved diagnostic measures through AI application in comparison to the clinician or radiologist model in terms of accuracy ($P < 0.05$), sensitivity ($P < 0.001$), and DSF ($P < 0.001$).

In the prospective strategy development for treatment, 9 studies were selected, out of these 2 studies stated the selection of surgical patients among epileptic patients and pituitary macroadenoma patients simultaneously.[25,26] Natural language processing (NLP), a technique used to develop machine learning-based predictive models through written text processing, was employed to identify surgical site infections using electronic health data, demonstrating significant predictive value ($P < 0.001$). Similarly, MRI is used by radiologists to interpret the accuracy of the ML model which showed 93% accuracy and AUC (0.99) which aided in surgical design planning for treatment. Three studies emphasized neurosurgery planning through segmentation to mine tumor three-dimensional shape from MRI. Manual segmentation effectively about ML-oriented MRI segmentation was weighed with 2 studies assessing glioma and 1 assessing brainstem segmentation. The results demonstrated a significant impact of ML applications ($P < 0.001$) in terms of Dice similarity coefficient (DSC), percentage volume difference, and speed. Median segmentation time with the ML model was 36–40 seconds, compared to manual segmentation which took 20.2 minutes.[27] This indicates that the ML model was notably more efficient and accurate. Correspondingly, both segmentation studies with glioma showed potentially significant sensitivity for ML in contrast to clinical experts in neurosurgery.[28,29]

The localization of the epileptic zone was investigated in four studies[30-33] using functional MRI, fluorodeoxyglucose-positron emission tomography, and other input features. The studies utilized quadratic discriminant analysis, support vector machines (SVM), artificial neural networks (ANN), and linear discriminant analysis (LDA) as machine learning models, respectively. There was no potentially significant alteration observed based on signs and symptoms in "temporal lobe epilepsy (TLE)" from extratemporal lobe epilepsy. A momentous high accuracy was demonstrated by the ML model based on MRI in the differential diagnosis of both right and left-sided temporal lobe epilepsy (TLE).

Among four studies that examined outcome forecast, two studies[34,35] through MRI brain have evaluated predicted survival in glioma patients. The ML model used was SVM and FCM. One of these studies showed significantly improved AUC with the ML model when evaluated against the clinician model. Based on clinical presentation, another study predicted TBI patients in hospital survival. The ML model showed superiority in terms of AUC, sensitivity, and accuracy, while specificity was also found to be equivalent in this study.[35] Another study used SVM to estimate the burden of "perivascular space" enlargement in patients to predict hospital survival and outcomes.[36] Although this study reflected better AUC for neuroradiologists in comparison to AI applications, still this study was unable to prespecified presentation keys and was categorized as reporting bias due to unclear risk of discerning reportage.

In the study, ML models were observed; in addition to ML model input, diagnostic characteristics including MRI with or without other characteristics were also evaluated for diagnosis and prognosis through both clinician and ML models. Similarly, Haug's[37] study also stated that in e-medical records, the use of ML to attain AI is centrally related to the extraction of predictive information of multifaceted health-care data through the ML model and its effective prognostic algorithm revolution. Precise patient outcome calculation can also help in primary preventive interference and assigning more effective health-care reserves to identify high-risk patients precisely.[37]

In the study, the implication of the ML model in neurosurgical patients is recurrently used for radiological data examination by mostly "ANN" means along with other neurosurgical applications. ANN-based supervised learning in the study was used to handle complex relationships between input and output. Emblem *et al*.[38] emphasized and explained the usage of each voxel as a single input piece and the abstraction of information extent through the ML model is enormously high making it more speedy and precise in comparison to human efforts which take more time.[38] Therefore, radiological and clinical data analysis by ML for diagnostic, segmentation, and outcome predictions served as one of the first ML applications that were correlated to actual clinical practices.

Although the study results demonstrated significant improvements in diagnosis, preoperative surgical decisions, and outcome predictions using the ML model, suggesting it could be more effective and time-saving compared to clinician diagnoses. However, it was clarified that despite the high accuracy of ML models in analysis, they ultimately enhance decision-making for clinicians and radiologists by providing more accurate and precise medical condition images of patients. This correlation between humans and machines also saves practitioners time in diagnosis and segmentation.[36]

A seamlessly integrated AI component within the imaging workflow would increase efficiency, reduce errors, and achieve objectives with minimal manual input by providing trained radiologists with prescreened images and identified features. Therefore, substantial efforts and policies are being put forward to facilitate technological advances related to AI in medical imaging. Almost all image-based radiology tasks are contingent upon the quantification and assessment of radiographic characteristics from images. These characteristics can be important for the clinical task at hand, that is, for the detection, characterization, or monitoring of diseases. The application of logic and statistical pattern recognition to problems in medicine has been proposed since the early 1960s.[34,35] As computers became more prevalent in the 1980s, the AI-powered automation of many clinical tasks has shifted radiology from a perceptual subjective craft to a quantitatively computable domain.[37,39] The rate at which AI is evolving radiology is parallel to that in other application areas and is proportional to the rapid growth of data and computational power. Health-care providers produce and apprehend huge amounts of data including tremendously valued signals and information through AI and ML technology at a far better pace outstanding what "traditional" methods of analysis for large quantitative data sets. ML has emerged as a recent approach that excels in integrating, investigating, and predicting outcomes based on large, heterogeneous datasets (cf. health informatics). Applications of DL in healthcare range from one-dimensional biosignal analysis[40] to predicting and assessing medical events such as seizures[41] or cardiac arrests,[42] computer-aided detection,[43] and improving diagnostic accuracy.[44] These advancements aid in survival analysis, facilitate clinical decision-making,[45] contribute to drug discovery,[46] and pharmacogenomics, aid in therapy selection,[47] enhance operational efficiency,[48] enable stratified care delivery,[49] and facilitate examination of electronic health records.[50]

Beyond AI's application in medical imaging, its integration holds immense potential across various domains within health care. As AI continues to advance, its impact on medical practices and patient care is becoming increasingly profound. AI can enhance surgeons' abilities across all stages of neurosurgery, including preoperative planning, intraoperative guidance, and postoperative monitoring.[51] A recent study by Kozel *et al*.[52] reported that ChatGPT-4 achieved an 85% accuracy rate for diagnoses and a 75% accuracy rate for treatment plans, whereas ChatGPT-3.5 had rates of 65% and 10%, respectively. Another review study revealed that ML methods have demonstrated their effectiveness in various aspects of neurosurgery, including identifying tumors, predicting surgical outcomes, forecasting seizure outcomes, anticipating aneurysms, and beyond, highlighting the extensive influence and potential of ML in enhancing patient care and outcomes within neurosurgical practice.[53]

The adoption of robotics in neurosurgery is on the rise, alongside the integration of AI in neurointensive care units for data analysis and patient management. In addition, AI holds the potential to predict patient outcomes. Various AI applications have been introduced in neurosurgery, with further advancements anticipated in the coming years.[54] In upcoming years, AI algorithms are likely to play a more significant role in clinical research, assessing the effectiveness of clinical and surgical procedures, and performing analyses in health economics.[55] It is anticipated that AI and ML play a prominent role in spinal care by developing algorithms to assist in decision-making regarding complex spinal conditions. However, integrating these technologies into clinical practice presents challenges, including ensuring data quality, overcoming integration obstacles, addressing data security concerns, and navigating ethical considerations.[56]

While the potential of AI in medicine and neurosurgery is promising, numerous hurdles must be overcome before its impact becomes evident in neurosurgical practice. These challenges range from ensuring patient privacy to securing access to reliable datasets and addressing the risk of surgeons overly relying on AI.[57] AI in neurosurgery seems to be heading toward a patient-centered model, focusing on aiding with clinical tasks and assisting in patient diagnosis and preoperative assessment.

The progression of AI presents opportunities to merge data-driven disciplines such as genomics with surgical practices, paving the way for tailored treatments and refined approaches to public health initiatives.[57] Upcoming health-care professionals must stay abreast of health-care advancements and adeptly integrate them into their practices to yield enhanced results.

One of the challenges regarding the implementation of ML models in clinical setup is that algorithm-driven mechanisms are complex and difficult to interpret and therefore referred to as the "black box technique." On the other hand, if we check conventional statistical methodologies. Another challenge is that a large amount of completely categorized data are required for the generation of an ML model. Therefore in the research setup, AI application's performance could be exceed expectations due to data high quality.[58]

To address these challenges, more studies are needed to explore further the connection between machines and humans, benefiting both clinicians and patients through influential analysis using ML and AI applications. This includes assessing outcomes after surgery planned with the aid of ML model segmentation, among other approaches. ML is the finest method for data integration and heterogeneous data. Radiological and clinical data are predictively correlated with clinical practices.

Study limitations

It is important to highlight the limitations of this study, although it has provided valuable insights into the comparative efficacy of ML algorithms and clinical expertise in neurosurgical patient care. First, the scope of the literature search may be constrained by the selected timeframe and database, potentially overlooking relevant studies published outside these parameters. In addition, the relatively small number of studies included in this review may limit the depth of the analysis, affecting the generalizability of the findings.

Second, the quality and heterogeneity of the included studies may introduce variability and bias into the analysis. Variations in study design, patient populations, and outcome measures could limit the comparability and generalizability of findings across studies. While ML algorithms offer promising insights, their performance may be influenced by the quality and quantity of available data. Variability in data sources, data preprocessing methods, and feature selection techniques could affect algorithmic performance and generalizability.

CONCLUSION

The present review study has shed light on the potential impact of ANN in neurosurgery, highlighting their ability to save time, enhance diagnosis, facilitate segmentation, aid in data interpretation, and improve prediction outcomes. Notably, the investigation reveals that ANN applications offer tangible benefits such as time-saving measures, heightened diagnostic accuracy, streamlined data segmentation and interpretation, and enhanced predictive capabilities. These findings suggest a profound shift in neurosurgical practice toward more efficient and effective patient care strategies.

Specifically, the study highlights how the integration of ANN technologies can expedite diagnostic processes, enabling clinicians to promptly identify and address neurological conditions. This accelerated diagnosis holds significant implications for patient outcomes, as timely interventions can mitigate the progression of diseases and improve overall treatment efficacy. In addition, the study highlights the role of ANN in refining decision-making processes by providing clinicians with valuable insights gleaned from sophisticated data analysis. Moreover, the review identifies promising avenues for future research within the field of AI and neurosurgery. By further refining and validating ANN algorithms, researchers can unlock new opportunities for enhancing diagnostic precision and treatment efficacy. Future investigations may also explore the integration of ANN technologies into larger patient cohorts and diverse neurosurgical procedures, thereby expanding the scope of their applicability in clinical settings.

Finally, the study also highlighted the importance of addressing critical challenges such as data quality, integration, security, and ethical considerations. By proactively addressing these obstacles, researchers can ensure the responsible and effective implementation of ANN technologies in neurosurgical practice, thereby maximizing their potential to revolutionize patient care and clinical outcomes.

Acknowledgment

The author is thankful to all the associated personnel, who contributed to this study by any means.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging applications of artificial intelligence in neuro-oncology. Radiology 2019;290:607-18.
2. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018;8:020303.
3. Yao AD, Cheng DL, Pan I, Kitamura F. Deep learning in neuroradiology: A systematic review of current algorithms and approaches for the new wave of imaging technology. Radiol Artif Intell 2020;2:e190026.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
5. Nucci CG, De Bonis P, Mangiola A, Santini P, Sciandrone M, Risi A, *et al.* Intracranial pressure wave morphological classification: Automated analysis and clinical validation. Acta Neurochir (Wien) 2016;158:581-8.
6. Swinburne NC, Schefflein J, Sakai Y, Oermann EK, Titano JJ, Chen I, *et al.* Machine learning for semi-automated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. Ann Transl Med 2019;7:232.
7. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, *et al.* Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med 2018;24:1337-41.
8. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
9. Kitajima M, Hirai T, Katsuragawa S, Okuda T, Fukuoka H, Sasao A, *et al.* Differentiation of common large sellar-suprasellar masses effect of artificial neural network on radiologists' diagnosis performance. Acad Radiol 2009;16:313-20.
10. Yamashita K, Yoshiura T, Arimura H, Mihara F, Noguchi T, Hiwatashi A, *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.
11. Bidiwala S, Pittman T. Neural network classification of pediatric posterior fossa tumors using clinical and imaging data. Pediatr Neurosurg 2004;40:8-15.
12. Arle JE, Perrine K, Devinsky O, Doyle WK. Neural network analysis of preoperative variables and outcome in epilepsy surgery. J Neurosurg 1999;90:998-1004.
13. Awuah WA, Adebusoye FT, Wellington J, David L, Salam A, Weng Yee AL, *et al.* Recent outcomes and challenges of artificial intelligence, machine learning, and deep learning in neurosurgery. World Neurosurg×2024;23:100301.
14. Buchlak QD, Esmaili N, Leveque JC, Bennett C, Farrokhi F, Piccardi M. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review. J Clin Neurosci 2021;89:177-98.
15. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, *et al.* Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review. Neurosurg Rev 2020;43:1235-53.
16. Mofatteh M. Neurosurgery and artificial intelligence. AIMS Neurosci 2021;8:477-95.
17. Groiss SJ, Wojtecki L, Südmeyer M, Schnitzler A. Deep brain stimulation in Parkinson's disease. Ther Adv Neurol Disord 2009;2:20-8.
18. Pilaz LJ, Liu J, Joshi K, Tsunekawa Y, Musso CM, D'Arcy BR, *et al.* Subcellular mRNA localization and local translation of Arhgap11a in radial glial progenitors regulates cortical development. Neuron

2023;111:839-56.e5.
19. Mofatteh M. Neurodegeneration and axonal mRNA transportation. Am J Neurodegener Dis 2021;10:1-12.
20. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. J Magn Reson Imaging 2010;31:680-9.
21. Zhao ZX, Lan K, Xiao JH, Zhang Y, Xu P, Jia L, *et al*. A new method to classify pathologic grades of astrocytomas based on magnetic resonance imaging appearances. Neurol India 2010;58:685-90.
22. Emblem KE, Nedregaard B, Hald JK, Nome T, Due Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: Brain tumor segmentation using knowledge-based fuzzy clustering. J Magn Reson Imaging 2009;30:1-10.
23. Abdolmaleki P, Mihara F, Masuda K, Buadu LD. Neural networks analysis of astrocytic gliomas from MRI appearances. Cancer Lett 1997;118:69-78.
24. Sinha M, Kennedy CS, Ramundo ML. Artificial neural network predicts CT scan abnormalities in pediatric patients with closed head injury. J Trauma 2001;50:308-12.
25. Cohen KB, Glass B, Greiner HM, Holland Bouley K, Standridge S, Arya R, *et al*. Methodological issues in predicting pediatric epilepsy surgery candidates through natural language processing and machine learning. Biomed Inform Insights 2016;8:11-8.
26. Cuocolo R, Ugga L, Solari D, Corvino S, D'Amico A, Russo D, *et al*. Prediction of pituitary adenoma surgical consistency: Radiomic data mining and machine learning on T2-weighted MRI. Neuroradiology 2020;62:1649-56.
27. Dolz J, Betrouni N, Quidet M, Kharroubi D, Leroy HA, Reyns N, *et al*. Stacking denoising auto-encoders in a deep network to segment the brainstem on MRI in brain cancer patients: A clinical study. Comput Med Imaging Graph 2016;52:8-18.
28. Hashido T, Saito S, Ishida T. A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas. Sci Rep 2020;10:6121.
29. Clarke LP, Velthuizen RP, Clark M, Gaviria J, Hall L, Goldgof D, *et al*. MRI measurement of brain tumor response: Comparison of visual metric and automatic segmentation. Magn Reson Imaging 1998;16:271-9.
30. Chiang S, Levin HS, Haneef Z. Computer-automated focus lateralization of temporal lobe epilepsy using fMRI. J Magn Reson Imaging 2015;41:1689-94.
31. Kassahun Y, Perrone R, De Momi E, Berghöfer E, Tassi L, Canevini MP, *et al*. Automatic classification of epilepsy types using ontology-based and genetics-based machine learning. Artif Intell Med 2014;61:79-88.
32. Kerr WT, Nguyen ST, Cho AY, Lau EP, Silverman DH, Douglas PK, *et al*. Computer-aided diagnosis and localization of lateralized temporal lobe epilepsy using interictal FDG-PET. Front Neurol 2013;4:31.
33. Lee JS, Lee DS, Kim SK, Lee SK, Chung JK, Lee MC, *et al*. Localization of epileptogenic zones in F-18 FDG brain PET of patients with temporal lobe epilepsy using artificial neural network. IEEE Trans Med Imaging 2000;19:347-55.
34. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 1959;130:9-21.
35. Lodwick GS, Keats TE, Dorst JP. The coding of roentgen images for computer analysis as applied to lung cancer. Radiology 1963;81:185-200.
36. Rughani AI, Dumont TM, Lu Z, Bongard J, Horgan MA, Penar PL, *et al*. Use of an artificial neural network to predict head injury outcome. J Neurosurg 2010;113:585-90.
37. Haug PJ. Uses of diagnostic expert systems in clinical care. Proc Annu Symp Comput Appl Med Care 1993;1993:379-83.
38. Emblem KE, Pinho MC, Zöllner FG, Due Tonnessen P, Hald JK, Schad LR, *et al*. A generic support vector machine model for preoperative glioma survival associations. Radiology 2015;275:228-34.
39. Ambinder EP. A history of the shift toward full computerization of medicine. J Oncol Pract 2005;1:54-6.
40. Ganapathy N, Swaminathan R, Deserno TM. Deep learning on 1-D biosignals: A taxonomy-based survey. Yearb Med Inform 2018;27:98-109.
41. Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction – Ready for a new era. Nat Rev Neurol 2018;14:618-30.
42. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018;7:e008678.
43. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, *et al*. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35:1285-98.
44. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, *et al*. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172:1122-31.e9.
45. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
46. Jiménez J, Škalič M, Martínez Rosell G, De Fabritiis G. K (DEEP): Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model 2018;58:287-96.
47. Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn Feuer A, Dinov ID, *et al*. Deep learning in pharmacogenomics: From gene regulation to patient stratification. Pharmacogenomics 2018;19:629-50.
48. Jiang S, Chin KS, Tsui KL. A universal deep learning approach for modeling the flow of patients under different severities. Comput Methods Programs Biomed 2018;154:191-203.
49. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, *et al*. Identifying Distinct Subgroups of ICU patients: A machine learning approach. Crit Care Med 2017;45:1607-15.
50. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, *et al*. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18.
51. Shariff S, Kantawala B, Mkrtchyan A, Mirzaei F, Grigoryan V. Artificial intelligence in neurosurgery: Pre, intra and post operation applications. J Surg 2023;8:1895.
52. Kozel G, Gurses ME, Gecici NN, Gökalp E, Bahadir S, Merenzon MA, *et al*. Chat-GPT on brain tumors: An examination of artificial intelligence/machine learning's ability to provide diagnoses and treatment plans for example neuro-oncology cases. Clin Neurol Neurosurg 2024;239:108238.
53. Tangsrivimol JA, Schonfeld E, Zhang M, Veeravagu A, Smith TR, Härtl R, *et al*. Artificial intelligence in neurosurgery: A state-of-the-art review from past to future. Diagnostics (Basel) 2023;13:2429.
54. Noh SH, Cho PG, Kim KN, Kim SH, Shin DA. Artificial intelligence for neurosurgery: Current state and future directions. J Korean Neurosurg Soc 2023;66:113-20.
55. Charles YP, Lamas V, Ntilikina Y. Artificial intelligence and treatment algorithms in spine surgery. Orthop Traumatol Surg Res 2023;109:103456.
56. Yagi M, Yamanouchi K, Fujita N, Funao H, Ebata S. Revolutionizing spinal care: Current applications and future directions of artificial intelligence and machine learning. J Clin Med 2023;12:4188.
57. Iqbal J, Jahangir K, Mashkoor Y, Sultana N, Mehmood D, Ashraf M, *et al*. The future of artificial intelligence in neurosurgery: A narrative review. Surg Neurol Int 2022;13:536.
58. González Castro V, Valdés Hernández MD, Chappell FM, Armitage PA, Makin S, Wardlaw JM. Reliability of an automatic classifier for brain enlarged perivascular spaces burden and comparison with human performance. Clin Sci (Lond) 2017;131:1465-81.