Aus der Klinik und Poliklinik für Neuroradiologie
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Evaluation einer KI-Software zur Klassifikation intrakranieller Tumoren in der Magnetresonanztomographie

Inauguraldissertation
zur Erlangung des Doktorgrades der
Medizin
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Mariya Kondova
aus Petritsch, Bulgarien

Mainz, 2025

Es gilt das Urheberrecht.

Wissenschaftlicher Vorstand:	Univ.-Prof. Dr. Hansjörg Schild
1. Gutachter:	
2. Gutachter:	
Tag der Promotion:	12. August 2025

Für meine Eltern

und

meinen Onkel Valeri Tanev

Inhaltsverzeichnis

1. Zusammenfassung ... 1
2. Publikation ... 5
3. Literaturverzeichnis ... 13
4. Danksagung ... 15
5. Tabellarischer Lebenslauf ... 17

1. Zusammenfassung

In den vergangenen Jahren wurde weltweit ein Anstieg der Inzidenz von Hirntumoren in allen Altersgruppen dokumentiert (1, 2). Die Bildgebung intrakranieller Raumforderungen ist daher für die Diagnostik, Therapieplanung, Vorhersage des Outcomes und das Follow-up der betroffenen Patienten essenziell. Insbesondere die Differenzierung zwischen malignen und benignen Tumoren sowie zwischen primären und metastatischen Hirntumoren ist für die Festlegung einer optimalen Therapie unerlässlich. So soll beispielwiese bei Verdacht auf ein ZNS-Lymphom eine Kortikosteroidtherapie vor der Biopsie vermieden werden, da diese die histopathologische Diagnose verfälschen kann (3).

Die Magnetresonanztomographie (MRT) ist die Methode der Wahl für die Beurteilung intrakranieller Pathologien, da sie einen exzellenten Weichteilkontrast bietet, der dem der CT überlegen ist, und dabei ohne ionisierende Strahlung und daher ohne Risiko der Entwicklung von strahleninduzierten Tumorerkrankungen auskommt (4, 5). Aufgrund phänotypischer Überlappungen vieler Tumorentitäten, des seltenen Auftretens einiger Tumoren und in Abhängigkeit von der Erfahrung des Untersuchers stellt die korrekte Klassifikation dieser Raumforderungen oftmals eine Herausforderung dar.

Der aktuelle Arbeitsablauf sieht die konventionelle Befundung durch einen Arzt vor. In Anbetracht der zunehmenden Anzahl radiologischer Untersuchungen erscheint jedoch eine Optimierung der Arbeitsabläufe unerlässlich (6, 7). In den letzten Jahren hat sich der Einsatz von künstlicher Intelligenz (KI) in der Medizin zunehmend etabliert. Die Entwicklung und Validierung zahlreicher KI-Softwares zeigen, dass mit Hilfe dieser Anwendungen der klinische Workflow verbessert werden kann. Dazu gehören z.B. Softwares zu Bildrekonstruktion, automatisierter Segmentierung und Klassifizierung unterschiedlicher Pathologien, zur Auswertung von Röntgen-Thorax-Aufnahmen (8, 9), Detektion intrakranieller Blutungen und ischämischer Hirninfarkte in nativer CT-Bildgebung (9, 10), sowie zu Detektion akuter Ischämien mittels diffusionsgewichteter MRT (10, 11). Da im Falle der intrakraniellen Raumforderungen eine schnelle und akkurate Diagnose wesentlich die Therapie beeinflusst und zu einer Erhöhung der Patientenüberlebensrate beiträgt, könnte die Verwendung einer dafür spezialisierten KI-Software einen direkten Nutzen für die Verbesserung der Patientenversorgung bedeuten (12). Einige KI-basierte Anwendungen sind bereits in der Lage, Bilddaten automatisch zu segmentieren und Pathologien sicher zu erkennen. Die genaue Analyse und Klassifikation intrakranieller Tumoren ist allerdings komplex, sodass sich hier Deep-learning-basierte Techniken anbieten.

Ziel dieser Studie war die Evaluation und Optimierung eines Deep Learning-basierten Algorithmus zur Detektion und Klassifikation maligner und benigner intrakranieller

Raumforderungen anhand von MRT-Datensätzen sowie die Bewertung seiner Anwendbarkeit im klinischen Alltag. Durch qualitativen Vergleich der Ergebnisse der konventionellen mit der KI-basierten Befundung sollte der klinische Mehrwert und der Nutzen für den Patienten evaluiert werden.

Ein von der National Medical Products Administration (NMPA) China zugelassenes Deep-Learning-Tool zur automatisierten Tumordetektion, -segmentierung und -klassifikation (BioMind; Beijing, China) wurde eingesetzt. Dieses besteht aus drei Convolutional Neural Network (CNN)-Modellen: i) einem Modell zur Läsionssegmentierung, ii) einem atlasbasierten Segmentierungsmodell und iii) einem Modell zur Pathologieklassifizierung. Zur Analyse wurden T2-gewichtete und kontrastmittelgestützte T1-gewichtete Bilder herangezogen.

In dieser retrospektiven monozentrischen Studie wurden 138 Patienten (65 Frauen und 73 Männer) mit einem Durchschnittsalter von 35±26 Jahren eingeschlossen. Bei 97 Patienten wurde ein intrakranieller Tumor diagnostiziert, während bei 41 Patienten keine intrakranielle Neoplasie vorlag. Zu den Einschlusskriterien zählte das Vorhandensein eines bei 1,5T oder 3,0T akquirierten MRT-Datensatzes, der die folgenden Sequenzen umfasste: T2w axial, native T1w axial und T1w nach intravenöser Kontrastmittelgabe. Die Schichtdicke der Bildgebung lag dabei zwischen 3 und 6 mm. Patienten mit vorausgegangenen Gehirnoperationen wurden exkludiert.

Die Bildanalyse erfolgte durch zwei Radiologen (Radiologe 1 (R1): 5 Jahre Erfahrung, Radiologe 2 (R2): 10 Jahre Erfahrung in der Befundung zerebraler MRT-Untersuchungen) sowie durch die Deep Learning (DL)-basierte Software. Zunächst führten die Radiologen eine qualitative Auswertung der Bilddaten durch und klassifizierten die jeweilige Pathologie. Daraufhin erfolgte die automatische Bildverarbeitung mittels des DL-Algorithmus. Die KI-Software sollte die Pathologie detektieren, lokalisieren und segmentieren. Auf Grundlage eines vortrainierten Netzwerks wurde dann eine automatische Klassifikation der erkannten Pathologie durchgeführt. Im letzten Schritt wurden die Ergebnisse der MRT-Auswertungen der KI-Software mit denen der beiden Radiologen verglichen und hinsichtlich der Genauigkeit, Sensitivität und Spezifität bei der Detektion zerebraler Raumforderungen ausgewertet. Als Goldstandard dienten die histopathologischen Ergebnisse.

Von den 97 Patienten mit intrakraniellem Tumor wurden 91 Fälle vom DL-basierten Algorithmus korrekt erkannt, was einer Sensitivität von 93,81 % [87,02–97,70] entsprach. Allerdings diagnostizierte der Algorithmus bei 15 von 41 Patienten ohne intrakranielle Raumforderung fälschlicherweise eine solche Pathologie, was zu einer Spezifität von 63,41 % [46,94–77,88] führte. Im Vergleich dazu zeigten die beiden Radiologen eine signifikant höhere Sensitivität und Spezifität von jeweils 100 % [96,27–100,00] bzw. 100 % [91,40–100,00].

Darüber hinaus wurde die Genauigkeit der von der KI und den Radiologen gestellten Diagnose untersucht. Betrachtete man nur die Hauptdiagnose des KI-Modells, betrug die Genauigkeit 0,63 (95 %-KI: 0,54–0,71). Berücksichtigte man jedoch auch die beiden Differenzialdiagnosen, verbesserte sich die Genauigkeit auf 0,74 (95 %-KI: 0,66–0,81). Im Vergleich dazu zeigten die Radiologen eine signifikant höhere diagnostische Genauigkeit mit 0,89 (95 %-KI: 0,82–0,94) für R1 und 0,92 (95 %-KI: 0,86–0,96) für R2. Wenn alle von den Radiologen angegebenen Differenzialdiagnosen berücksichtigt wurden, stieg die Genauigkeit weiter auf 0,93 (95 %-KI: 0,88–0,97) für R1 bzw. 0,98 (95 %-KI: 0,94–0,99) für R2, was im Vergleich zum KI-Modell einen signifikanten Unterschied darstellte (R1: $p < 0{,}001$; R2: $p < 0{,}001$).

Zusammenfassend war der in der zugrundeliegenden Studie eingesetzte KI-Algorithmus in der Lage, intrakranielle Tumoren eigenständig zu identifizieren und zu segmentieren und erzielte dabei zufriedenstellende Ergebnisse bei der Diagnosestellung. Obwohl die KI derzeit noch den erfahrenen Neuroradiologen deutlich unterlegen ist, wird der Algorithmus kontinuierlich weiterentwickelt und stellt einen bedeutenden Schritt in Richtung eines durch künstliche Intelligenz unterstützten Radiologie-Workflows dar.

AKADÉMIAI KIADÓ

IMAGING

ORIGINAL ARTICLE

Deep learning-based detection and classification of intracranial tumors on magnetic resonance imaging

MARIYA KONDOVA[1*], OLIVER KORCZYNSKI[1], MATTHIAS MÜLLER-ESCHNER[1], WINSON CHAN[2], ANTOINE SANNER[1], AHMED OTHMAN[1], MARC A. BROCKMANN[1] and SEBASTIAN ALTMANN[1]

[1] Department of Neuroradiology, University Medical Center Mainz, Johannes Gutenberg University, Langenbeckstr. 1, 55131 Mainz, Germany

[2] BioMind, Zhongguancun Medical Engineering Center, 10 Anxiang Road, 8th floor, Zhongguan, China

Received: April 6, 2024 • Revised manuscript received: July 29, 2024 • Accepted: August 17, 2024

ABSTRACT

Background: We evaluated the capability of an AI application to independently detect, segment and classify intracranial tumors in MRI.

Methods: In this retrospective single-centre study, 138 patients (65 female and 73 male) with a mean age of 35 ± 26y were included. 97 were diagnosed with an intracranial neoplasm, while 41 exhibited no intracranial pathology. Inclusion criteria were a 1.5 or 3.0T MRI dataset with the following sequences: T2 axial, T1 axial pre- and post-contrast with a slice thickness between 3 and 6 mm and no previous brain surgery. Image analysis was performed by two human readers (R1 = 5 years and R2 = 10 years of experience in brain MRI) and a deep learning (DL)-based AI model. Sensitivity, specificity and accuracy of the AI model and the human readers to detect and correctly classify brain tumors were measured. Histological results served as the gold standard.

Results: The AI model reached a sensitivity of 93.81% [87.02–97.70] and a specificity of 63.41% [46.94–77.88], while human readers reached 100% [96.27–100.00] and 100% [91.40–100.00], respectively. Human readers provided a significantly higher accuracy rate with R1 0.93 (95% CI: 0.88–0.97) and R2 0.98 (95% CI: 0.94, 0.99) vs. 0.74 (95% CI: 0.66–0.81) for the AI model (*P*-value <0.001).

Conclusion: The underlying DL-based AI algorithm can independently identify and segment intracranial tumors while providing satisfactory results for establishing the correct diagnosis. Despite its current inferiority compared to experienced radiologists, it still experiences ongoing development and it is a step towards developing an artificial intelligence-augmented radiology workflow.

KEYWORDS

brain tumor, artificial intelligence, automated classification, MRI

IMAGING 16 (2024) 2, 73–80
DOI: 10.1556/1647.2024.00232

© 2024 The Author(s)

*Corresponding author.
Tel.: +49 6131 17 7139.
E-mail: Mariya.Kondova@unimedizin-mainz.de

AKJournals

Introduction

In recent years, an increase in the incidence of brain tumors has been reported globally across all age groups [1, 2]. Therefore, imaging of intracranial tumors is crucial for establishing the correct diagnosis and for therapy planning and monitoring of the disease. Especially differentiation between malignant and benign tumors and between primary and metastatic brain tumors is essential to determine further therapeutic steps: E.g. if there is a suspicion of CNS lymphoma corticosteroid therapy should be avoided prior to biopsy because it may impede the histopathological diagnosis [3].

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

MRI is the method of choice for evaluating intracranial pathologies, demonstrating excellent soft tissue contrast superior to CT, with no ionizing radiation, and therefore, no risk of developing radiation-induced cancer [4, 5]. There are various sequences, techniques, and suggested protocols for tumor detection. Currently, recommended brain tumor MRI protocols include the following sequences: 2D fluid-attenuated inversion recovery (FLAIR) axial, susceptibility-weighted imaging (SWI), 2D DWI axial, 2D T1-weighted imaging (T1WI) and 3D T1 postcontrast [6]. Nevertheless, the correct classification of brain tumors remains challenging, especially for residents and doctors with limited experience in the field of neuroradiology. This is not only because of the phenotypic overlap of different tumor entities but also due to the rare occurrence of some tumor types.

In recent years, deep learning (DL)-based methods have been introduced in radiology with regard to a wide range of tasks, e.g., image reconstruction, automated segmentation and classification, but also identification of abnormal chest radiographs [7, 8], detection of intracranial hemorrhage and stroke on non-contrast head CT [8, 9], and acute stroke on diffusion-weighted MRI [8, 10], aiming to improve the clinical workflow.

We hypothesized that we could generate quick and readily available, high-quality reports with quantitative image information by applying a novel AI application for tumor detection, segmentation, quantitative volumetry, automated image interpretation and reporting. We also proposed that this deep learning-based algorithm may prove to be a valuable tool in our daily routine, potentially speeding up workflow while simultaneously improving the quality of medical reports.

Materials and methods

This retrospective study was approved by the Ethics Committee of the Rhineland Palatinate Chamber of Physicians, and written informed consent was waived.

Patient cohort

During the inclusion period of this retrospective single-centre study between October 2022 and July 2023, 361 patients underwent a clinically indicated brain MRI because of suspicion of intracranial tumor or as a follow-up due to already confirmed neoplasm. 138 patients who had undergone a clinically indicated MRI study were included, 97 were diagnosed with an intracranial tumor and 41 exhibited no intracranial pathology. The inclusion criteria were as follows: (i) an MRI study obtained on either 1.5T or 3T, (ii) available T2-weighted images (T2WI), T1-weighted pre-contrast (T1W) and T1-weighted post-contrast image (T1CWI) in the axial plane with a slice thickness of 3–6 mm, (iii) no previous brain surgery and (iv) tumor types used to train the AI (Artificial Intelligence) algorithm. 223 patients were excluded due to: (i) no obtained T2WI (n = 57), (ii) no T1CWI (n = 86), (iii) neither T2WI nor T1W pre- and post-contrast (n = 37) and (iv) due to brain surgery with resulting anatomic distortion and no preoperative imaging available (n = 43). The inclusion/exclusion process is presented in Fig. 1.

BioMind setup and software architecture

A deep learning-based tool approved by NMPA China (BioMind, Beijing, China) was used for automated tumor segmentation and classification, comprising three deep convolutional neural Network (CNN) models: i) a lesion segmentation model, ii) an atlas-based segmentation and iii) a classification model. The lesion segmentation model consists of two components: One trained to detect brain anomalies using MRI T2-weighted images (T2WI) and the other trained on T1-weighted post-contrast image (T1CWI). A final heat map is generated by overlaying the segmentation results obtained from T2WI and T1CWI. The heat map is a probabilistic distribution for image pixels to determine if a pixel is tumor-positive. The higher the probability, the “hotter” it is in visualization. The region of interest (ROI) is loosely cropped across all MRI sequences and concatenated with the heat map as multiple input channels. As a next step, the classification model analyzes the selected region to further distinguish different types of tumors with a confidence score, suggesting the two most probable diagnoses. The atlas-based algorithm, on the other hand, provides information about the spatial position of different structures

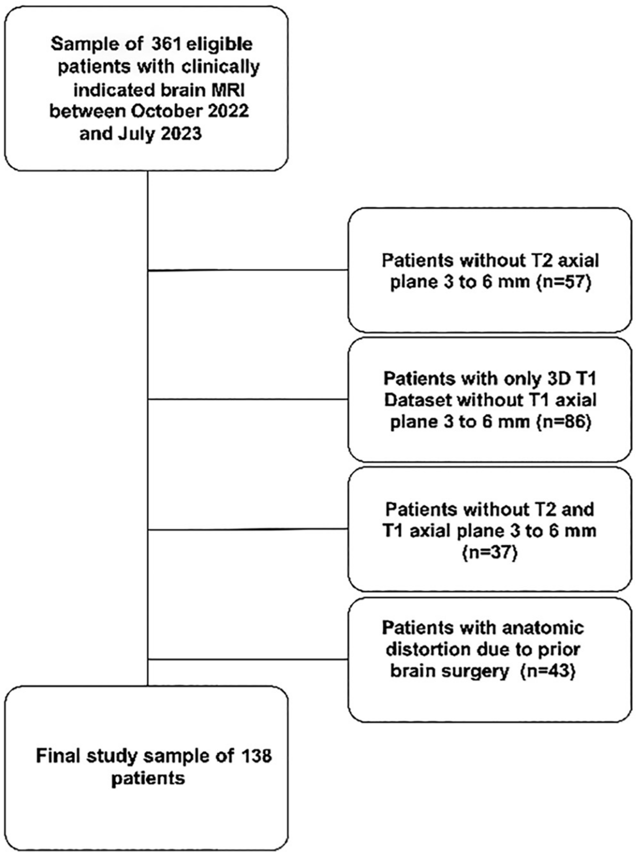

Fig. 1. Flow chart of the inclusion/exclusion process

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

[11] by mapping at least 12 major anatomic brain structures such as prefrontal lobes, temporal lobes, cerebellum, brainstem, pituitary region, etc., using T2WI as an input. The detected abnormality is fused with the segmented brain atlas afterwards to eliminate unlikely tumor types based on their regular distribution pattern. This filtering via brain atlas segmentation helps improve tumor differentiation accuracy. For example, if the region of interest coincides with the pituitary region's segmentation mask, the differential diagnosis will unlikely include medulloblastoma, vestibular schwannoma or hemangioblastoma. Hence, their likelihood score is assigned as zero by default. Eventually, the tumor type with the highest likelihood score is displayed as output, and the two most favorable diagnoses are given. For a visual representation of the AI algorithm analysis process, refer to Fig. 2.

Neuronal network

U-Net [12] has been the basis of neural network architecture for most medical image segmentation, including the tumor segmentation and the brain-atlas models used in our study. Meanwhile, the tumor classification model is an ensemble of multiple 3D CNNs with an attention mechanism.

In our applied algorithm, SE-blocks [13, 14] were incorporated for their ability to enhance feature representation power by improving channel dependencies and residual blocks [15, 16] to tackle the vanishing gradient problem in substantially deep neural network training.

Furthermore, this segmentation model uses stacks of three slices of 2D images as input (2.5D). It is done by having neighbouring images sandwich the main image in the middle layer, and the neural network's task is to produce a segmentation mask for the main image in the middle layer. Compared to pure 2D input, the neighbouring slice gives extra information on the construction and connectivity of brain tissue and lesions. The advantage of 2.5D, in contrast to complete 3D datasets, is that it can adapt MRI scans with different slice thicknesses and is relatively lightweight in model parameters and computations. Each convolutional layer has a residual block, or SE-block added parallel to the existing U-Net convolutional block.

The classification model is an ensemble [17] of multiple 3D CNNs with an attention mechanism [18]. T2WI, T1CWI and segmentation heat map are resampled before feeding into the model as multichannel input. Each image is resized to 256 × 256xD (D is the depth of the image), and the cropped image is fixed at 128 × 128 × 8. The output block is several fully connected layers, with the final output being an array of 24 elements representing 24 types of tumors found in our training dataset.

Image analysis

Image analysis was performed by the AI and two human readers, of which one was a board-certified radiologist (R1) with five years of experience in brain MRI reading, and the other one was a board-certified radiologist and neuroradiologist (R2) with ten years of experience in reading brain MRI scans. Both readers were blinded to the patient's history; however, they had access to the patient's age. The arrangement with the most pertinent MRI sequences was saved in the PACS Workstation (SECTRA, Linköping, Sweden) for the human readers. In contrast to the AI, however, all of the acquired sequences (e.g., 3D Datasets or diffusion-weighted imaging) were available for image interpretation to the readers, who were required to name the two most favorable diagnoses. The reading time was not limited. The accuracy, sensitivity and specificity were measured for the human readers. The same parameters were assessed also for the DL algorithm, and the results were compared. The ground truth was the histopathology obtained through biopsy or tumor resection.

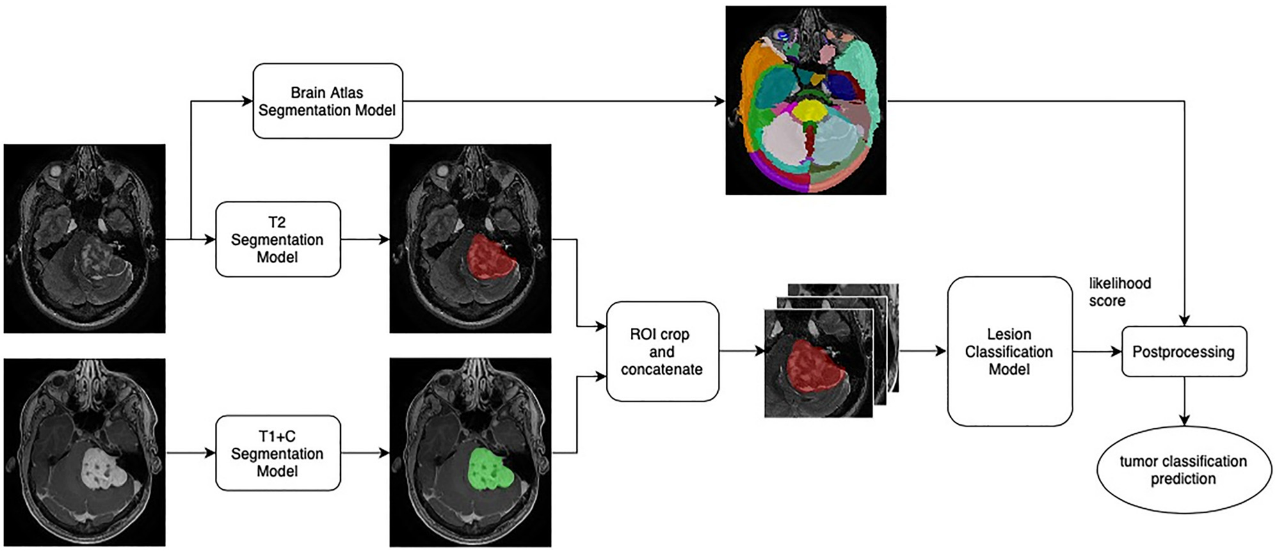

Fig. 2. Diagram illustrating the DL-based algorithm's analysis process

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

Statistical analysis

Statistical analysis was performed using R-Statistics (Version 4.1.3). Continuous variables were reported as mean ± standard deviation if normally distributed and as median/interquartile range for a non-normal distribution. Model performance was quantified by assessing the accuracy using Pearson's Chi-squared test. Furthermore, we reported sensitivity and specificity, including their 95% Confidence Intervals (95% CI), for each reader and the AI model and compared the results using McNemar's test. The multi-rater Fleiss' Kappa (χ) was assessed for the inter-rater agreement. The level of agreement was defined as follows: poor, $\chi < 0.21$; fair, $\chi = 0.21–0.40$; moderate, $\chi = 0.41–0.60$; substantial, $\chi = 0.61–0.80$; almost perfect, $\chi = 0.81–1.0$. P-values less than 0.05 were considered statically significant [12].

Results

Patient cohort

138 patients (65 female and 73 male) with a mean age of 35 ± 26y (age range between 0 and 83) were included. 97 were diagnosed with a benign or malignant intracranial tumor, while 41 exhibited no intracranial pathology. Among the tumor entities were glioma ($n = 52$), metastasis ($n = 13$), meningioma ($n = 11$), medulloblastoma ($n = 11$) and other less common tumors ($n = 10$) such as craniopharyngioma ($n = 4$), germinoma ($n = 2$), dysembryoplastic neuroepithelial tumor ($n = 1$), rosette-forming glioneuronal tumor ($n = 1$), dermoid cyst ($n = 1$), primitive neuroectodermal tumor of the CNS ($n = 1$). Patient characteristics are listed in Table 1.

AI model and human reader performance

Of 97 patients with an intracranial tumor on imaging, 91 were detected by the DL-based algorithm resulting in a sensitivity of 93.81% [87.02–97.70]. An example is illustrated in Fig. 3. However, 15 out of 41 patients with no intracranial tumor on imaging were falsely diagnosed with an intracranial pathology by the AI algorithm, which resulted in a specificity of 63.41% [46.94–77.88] (Fig. 4). The human readers demonstrated a higher sensitivity and specificity of 100% [96.27–100.00] and 100% [91.40–100.00], respectively.

Furthermore, we investigated the accuracy of the AI model and the human readers. When considering only the principal diagnosis provided by the AI model, the accuracy was 0.63 (95% CI: 0.54–0.71). However, when both differential diagnoses were considered, the accuracy improved to 0.74 (95% CI: 0.66–0.81). The human readers demonstrated an accuracy of 0.89 (95% CI: 0.82–0.94) (R1) and 0.92 (95% CI: 0.86–0.96) (R2) only for the most favorable diagnosis. The accuracy rates improved to 0.93 (95% CI: 0.88–0.97) and 0.98 (95% CI: 0.94, 0.99), respectively, when also the differential diagnosis was considered, resulting in a significantly higher accuracy rate compared to the AI model (R1 $P < 0.001$; R2 $P < 0.001$) (Fig. 5).

Table 1. Patient characteristics

Characteristics	Value (n = 138)
*Age (years)**	35 ± 26 (0–83)
Sex	
Male	73 (52.9%)
Female	65 (47.1%)
Intracranial neoplasm	97 (70.3%)
Glioma	53 (8.2%)
Metastasis	13 (13.4%)
Meningioma	11 (11.3%)
Medulloblastoma	11 (11.3%)
Others	10 (10.3%)
No clinically relevant findings	41 (29.7%)

*Data are mean ± 1SD, with ranges in parentheses.

Interrater reliability

The inter-rater agreement showed moderate to substantial results between the AI algorithm and the human readers with Fleiss' Kappa (χ) = 0.59 (CI: 0.50–0.68) for R1 and (χ) = 0.65 (CI: 0.56–0.73) for R2. An almost perfect inter-rater agreement was demonstrated between the human readers $\chi = 0.90$ (CI: 0.81–0.99).

Discussion

Our study aimed to evaluate the performance of a DL-based algorithm for detecting and classifying intracranial tumors on MRI. Our analysis revealed that the tested CNN model, although currently inferior to experienced radiologists, showed good overall accuracy. These results highlight the potential benefits of CNN, resulting in improved workflow. Additionally, it can be a valuable aid, especially for young, inexperienced radiologists.

Due to the increasing number of medical examinations, an improved radiology workflow is needed [19, 20]. Implementing such a model can help create prioritized worklists, enabling radiologists and neuroradiologists to focus on urgent and complex studies. However, such a model's high sensitivity and specificity is essential to prevent overloading radiologists with false positive cases, which could lead to alarm fatigue. This can also expedite the process from diagnosis to treatment [21, 22]. Additionally, an overall improvement in workflow efficiency allows for valuable learning opportunities for trainees and junior physicians.

Another benefit of the deep learning-based model in our study is the automated pathology segmentation and volumetry. Manual segmentation and volumetry are very time-consuming and often restricted to qualitative and visual assessment in the clinical routine. By implementing such an AL algorithm, physicians' workflow remains unaltered, as images are automatically transferred from PACs to the working station, and a suggested report is generated and supplied to physicians automatically.

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

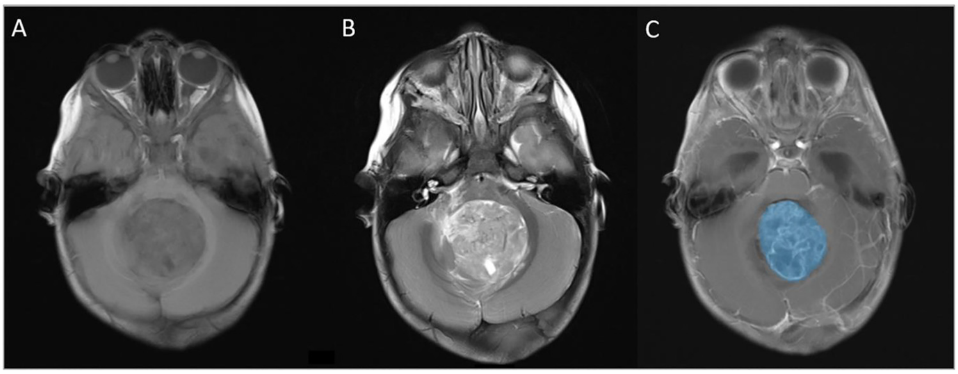

Fig. 3. Medulloblastoma in T1-weighted (A), T2-weighted (B) and with region of interest (ROI) in the T1-weighted post-contrast sequence (C)

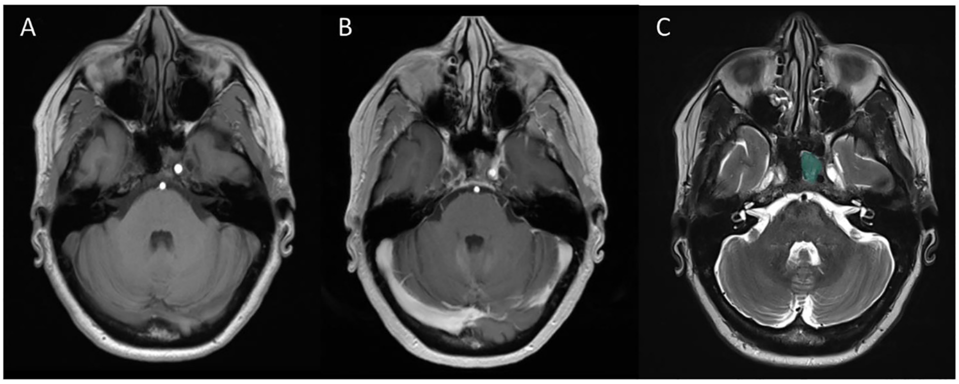

Fig. 4. Normal findings in (A) T1-weighted pre-contrast sequence, (B) T1-weighted post-contrast sequence and (C) T2-weighted image with falsely segmented and incorrectly classified region of interest (ROI) as pituitary tumor

While 3D imaging is increasingly applied in clinical routines, the software in our study processes only two-dimensional (2D) images. Over the past years, the use of three-dimensional (3D) image datasets has increased significantly, making optimization of 3D CNN architectures necessary. However, a significant limitation to implementing 3D models is the excessive computational load they require. Alternative methods to 3D CNNs are emerging, including integrating 2D CNNs with a neural network designed for sequence data to analyze sequential 2D images within a 3D volume [23].

The limitation of our study lies in its retrospective single-centered design. It has only been tested in an experimental setting but has yet to be in a broad clinical dataset; thus, ensuring the model's robustness requires further validation in a prospective setting. In Addition, it was trained on a Chinese population. Some studies observed reduced model performance when test sets were collected from institutions that differed from those where the training data was collected [19]. Another critical aspect to consider is the quality of the training data used, which directly affects the accuracy of an AI model.

Furthermore, the study dataset was limited mainly to the most common intracranial tumor types, and specific tumor types were underrepresented due to their low incidence rate. Therefore, we were unable to assess the model's generalizability effectively. Moreover, the AI algorithm was trained on patients older than 10 years, and our dataset included both adult and pediatric patients <10 years old. Future studies could investigate whether individualized pediatric and adult models could enhance performance.

The AI model in our study uses axial images as input. This could complicate the detection of specific tumor types depending on the region. For example, tumors in the pituitary regions are more easily detected on a sagittal or coronal plane. Additionally, a FLAIR-sequence seems the most suitable for detecting tumors [24]; however, the evaluated CNN algorithm detects tumors using a T2-sequence. Moreover, it does not consider additional sequences, such as diffusion-weighted imaging or SWI, which could aid the diagnosis. Nevertheless, it is expected that AI algorithms will continue to develop in the near future, enabling the simultaneous analysis of more sequences and the processing of 3D images.

Furthermore, there was considerable variation among the analyzed data sets. Some of the MRI images were acquired over a decade ago, and not all images had the exact spatial resolution due to the developing MRI scanner technology. The variation among patients, scanners (different manufacturers), scanner operators, higher spatial resolution,

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

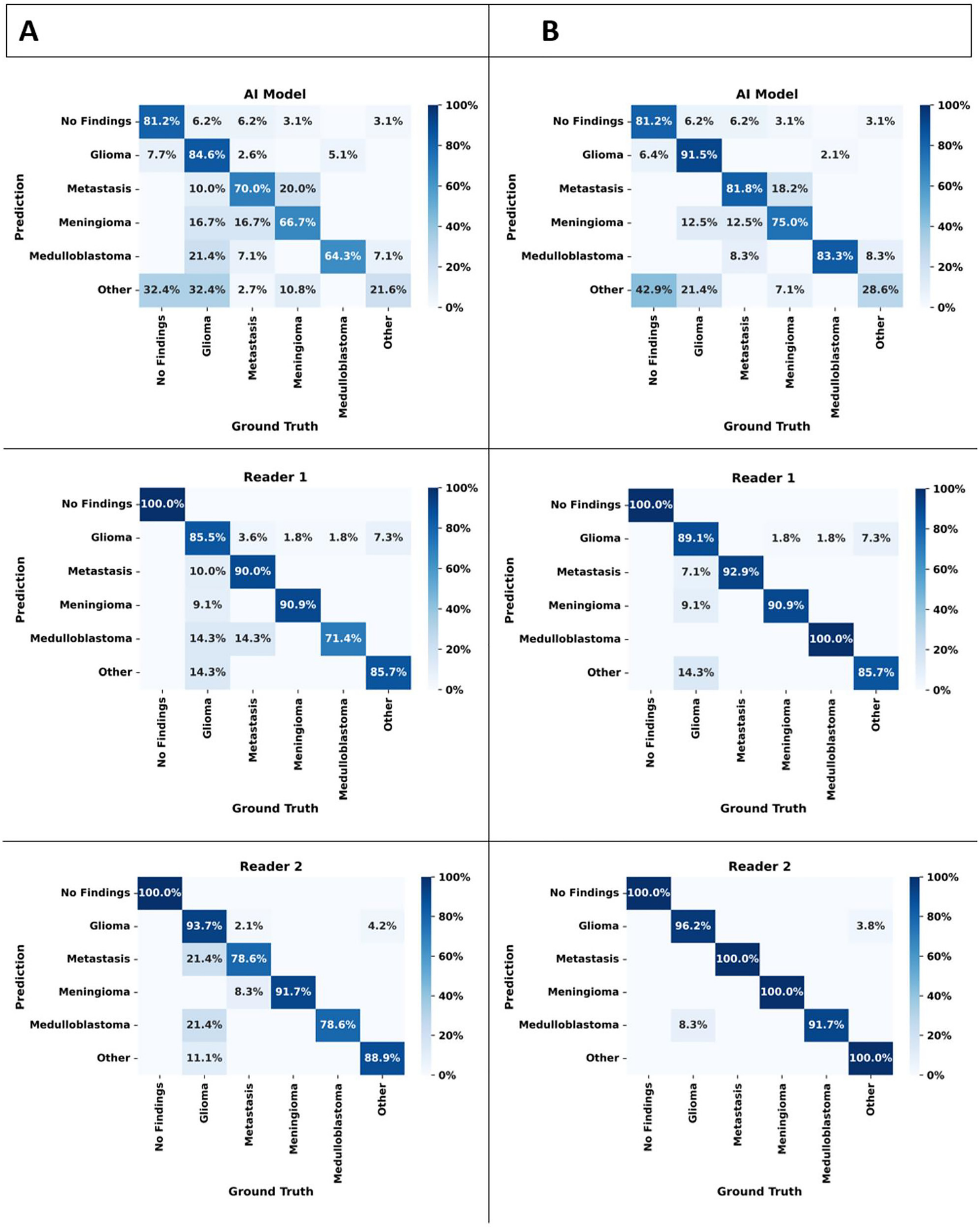

Fig. 5. The image shows the testing performance demonstrated by the AI model and the human readers in our study. Column A demonstrates the accuracy when only the proposed principal diagnosis was considered, and column B when the two most likely differential diagnoses were considered

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

and disease incidence in diverse populations is a crucial aspect to be considered [19, 25].

Conclusion

This AI algorithm, used in the underlying study, can independently identify and segment intracranial tumors and provides satisfactory results regarding the establishment of the correct diagnosis. Despite its current inferiority compared to experienced neuroradiologists, it still experiences ongoing development and it is a meaningful step towards developing an artificial intelligence-augmented radiology workflow.

Author contributions: Conceptualization, M.K. and M.A.B.; methodology, M.K.; software, W.Ch.; validation, M.K. and A.S.; formal analysis, M.K. and A.S.; investigation, M.K. and M.M.-E.; resources, M.A.B., M.K. and O.K.; data curation, M.K., S.A. and O.K.; writing—original draft preparation, M.K., A.S., W.Ch., A.E.O., M.A.B. and S.A.; writing—review and editing, M.K., A.S., W.Ch., A.E.O., M.A.B., S.A., M.M.-E. and O.K.; visualization, M.K., S.A., A.S. and W.Ch.; supervision, M.A.B., A.E.O und S.A.; project administration, M.A.B., M.K. and M.M.-E.; funding acquisition, M.M.-E. All authors reviewed the final version of the manuscript and agreed to submit it to IMAGING for publication.

Funding sources: This research was funded by BioMind, HANALYTICS PTE LTD 151 Lorong Chuan, New Tech Park, Singapore 556741.

Conflicts of interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; or in the decision to publish the results. The rest of the authors have no conflict of interest to disclose.

Ethical statement: The study was in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

[1] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al.: Brain tumor segmentation with deep neural networks. Med Image Anal 2017; 35: 18–31.

[2] Magadza T, Viriri S: Deep learning for brain tumor segmentation: A survey of state-of-the-art. J Imaging 2021; 7(2): 19.

[3] Scheichel F, Marhold F, Pinggera D, Kiesel B, Rossmann T, Popadic B, et al.: Influence of preoperative corticosteroid treatment on rate of diagnostic surgeries in primary central nervous system lymphoma: A multicenter retrospective study. BMC Cancer 2021; 21: 1–10.

[4] Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S: Emerging applications of artificial intelligence in neuro-oncology. Radiology 2019; 290(3): 607–18.

[5] Srinivas C, K SN, Zakariah M, Alothaibi YA, Shaukat K, Partibane B, Awal H: Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images. J Healthc Eng 2022; 2022: 3264367.

[6] Kamepalli H, Kalaparti V, Kesavadas C: Imaging recommendations for the diagnosis, staging, and management of adult brain tumors. Indian J Med Paediatr Oncol 2023; 44(01): 026–38.

[7] Yates E, Yates L, Harvey H: Machine learning "red dot": Open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. Clin Radiol 2018; 73(9): 827–31.

[8] Syed AB, Zoga AC, editors. Artificial intelligence in radiology: Current technology and future directions. Seminars in musculoskeletal radiology; 2018: Thieme medical publishers.

[9] Prevedello LM, Erdal BS, Ryu JL, Little KJ, Demirer M, Qian S, White RD: Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 2017; 285(3): 923–31.

[10] Lee E-J, Kim Y-H, Kim N, Kang D-W: Deep into the brain: Artificial intelligence in stroke imaging. J Stroke 2017; 19(3): 277.

[11] Shiee N, Bazin P-L, Cuzzocreo JL, Blitz A, Pham DL, editors. Segmentation of brain images using adaptive atlases with application to ventriculomegaly. Information processing in medical imaging: 22nd international conference, IPMI 2011, Kloster Irsee, Germany, July 3–8, 2011 Proceedings 22; 2011: Springer.

[12] Siddique N, Paheding S, Elkin CP, Devabhaktuni V: U-Net and its variants for medical image segmentation: a review of theory and applications. IEEE Access 2021; 9: 82031–57.

[13] Hu J, Shen L, Sun G, editors. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.

[14] Rundo L, Han C, Nagano Y, Zhang J, Hataya R, Militello C, et al.: USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. Neurocomputing 2019; 365: 31–43.

[15] He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

[16] Diakogiannis FI, Waldner F, Caccetta P, Wu C. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 2020; 162: 94–114.

[17] Hansen LK, Salamon P: Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 1990; 12(10): 993–1001.

[18] Luong M-T, Pham H, Manning CD: Effective approaches to attention-based neural machine translation. arXiv Preprint arXiv: 150804025 2015.

[19] Gauriau R, Bizzo BC, Kitamura FC, Landi Junior O, Ferraciolli SF, Macruz FB, et al.: A deep learning–based model for detecting abnormalities on brain MR images for triaging: Preliminary results from a multisite experience. Radiol Artificial Intell 2021; 3(4): e200184.

[20] Andriole KP: Addressing the coming radiology crisis: the society for computer applications in radiology transforming the radiological interpretation process (TRIP™) initiative. 2003.

[21] Soffer S, Klang E, Shimon O, Barash Y, Cahan N, Greenspana H, Konen E: Deep learning for pulmonary embolism detection

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

on computed tomography pulmonary angiogram: A systematic review and meta-analysis. Sci Rep 2021; 11(1): 15814.
[22] Mitka M: Joint commission warns of alarm fatigue: Multitude of alarms from monitoring devices problematic. Jama 2013; 309(22): 2315–6.
[23] Cheng PM, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón Romero F, et al.: Deep learning: An update for radiologists. Radiographics 2021; 41(5): 1427–45.
[24] Anaya-Isaza A, Mera-Jiménez L, Verdugo-Alejo L, Sarasti L: Optimizing MRI-based brain tumor classification and detection using AI: a comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network. Eur J Radiol Open 2023; 10: 100484.
[25] Hyde RJ, Ellis JH, Gardner EA, Zhang Y, Carson PL: MRI scanner variability studies using a semi-automated analysis system. Magnetic Resonance Imaging 1994; 12(7): 1089–97.

Open Access statement. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

Unauthenticated | Downloaded 12/18/24 01:00 PM UTC

3. Literaturverzeichnis

1. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Medical image analysis. 2017;35:18-31.

2. Magadza T, Viriri S. Deep learning for brain tumor segmentation: a survey of state-of-the-art. Journal of Imaging. 2021;7(2):19.

3. Scheichel F, Marhold F, Pinggera D, Kiesel B, Rossmann T, Popadic B, et al. Influence of preoperative corticosteroid treatment on rate of diagnostic surgeries in primary central nervous system lymphoma: a multicenter retrospective study. BMC cancer. 2021;21:1-10.

4. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging applications of artificial intelligence in neuro-oncology. Radiology. 2019;290(3):607-18.

5. Srinivas C, K SN, Zakariah M, Alothaibi YA, Shaukat K, Partibane B, Awal H. Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images. J Healthc Eng. 2022;2022:3264367.

6. Gauriau R, Bizzo BC, Kitamura FC, Landi Junior O, Ferraciolli SF, Macruz FB, et al. A deep learning–based model for detecting abnormalities on brain MR images for triaging: preliminary results from a multisite experience. Radiology: Artificial Intelligence. 2021;3(4):e200184.

7. Andriole KP. Addressing the Coming Radiology Crisis: The Society for Computer Applications in Radiology Transforming the Radiological Interpretation Process (TRIP□) Initiative. 2003.

8. Yates E, Yates L, Harvey H. Machine learning “red dot”: open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. Clinical radiology. 2018;73(9):827-31.

9. Syed AB, Zoga AC, editors. Artificial intelligence in radiology: current technology and future directions. Seminars in musculoskeletal radiology; 2018: Thieme medical publishers.

10. Prevedello LM, Erdal BS, Ryu JL, Little KJ, Demirer M, Qian S, White RD. Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology. 2017;285(3):923-31.

11. Lee E-J, Kim Y-H, Kim N, Kang D-W. Deep into the brain: artificial intelligence in stroke imaging. Journal of stroke. 2017;19(3):277.

12. Sad-classification using advanced deep learning techniques. Microsc Res Tech. 2021;84(6):1296-308.ad T, Rehman A, Munir A, Saba T, Tariq U, Ayesha N, et al. Brain tumor detection and multi

4. Danksagung

5. Tabellarischer Lebenslauf