Dean&Francis

ISSN 2959-6157

Deep Learning Architectures for Brain Tumor Classification: A Comprehensive Investigation of Advances, Challenges and Clinical Translation

Canxu Wang

Department of Leeds, Southwest Jiaotong University, Chengdu, China
wangcanxu@my.swjtu.edu.cn

Abstract:

Brain tumor categorization from medical pictures is still a difficult and skill-dependent endeavor for radiologists, despite the fact that it is vital for efficient planning of treatment and better patient outcomes. The pressing demand for more dependable and effective diagnostic tools is addressed by this review, which thoroughly examines cutting-edge Deep Learning (DL) techniques used to automate brain tumor classification. This paper systematically examined key DL architectures employed in this domain, including Convolutional Neural Networks (CNNs) – highlighting hierarchical frameworks and Bayesian optimization approaches, Capsule Networks (CapsNets) – focusing on boundary-guided and Bayesian variants like BayesCap for uncertainty quantification, and Vision Transformers (ViTs) – particularly ensembled models leveraging multi-head self-attention. The analysis highlights the remarkable accuracies (e.g., over 98% in multiple experiments) attained by these sophisticated methodologies, synthesizing evidence on model performance. The review also critically examines the main obstacles to clinical adoption, including the restrictive interpretability of "black-box" models, the lack of expert-annotated data and the challenge of obtaining multidisciplinary cooperation for widespread implementation, problems with model applicability across various datasets and imaging protocols, and strict patient data privacy concerns. It can be concluded that while DL offers immense potential for revolutionizing brain tumor diagnosis, future research must prioritize developing explainable Artificial Intelligence (AI) techniques, robust domain adaptation methods, efficient lightweight models, and privacy-preserving frameworks like federated learning to enable trustworthy and widespread clinical implementation.

ISSN 2959-6157

Keywords: Deep learning; brain tumor classification; machine learning.

1. Introduction

Abnormal cell growth in the brain is known as brain tumors [1]. Their ability to interfere with normal brain activities, resulting in a range of symptoms like immobility, blindness or deafness, and potentially threatening life, makes them a serious hazard to human health [2].

Data from NBTS reveal that over 4,200 individuals in the UK are living with primary brain tumors. In the USA, annual statistics show 13,000 deaths alongside 29,000 new diagnoses of primary brain tumors [3]. Therefore, in clinical practice, it is crucial to accurately and promptly classify brain tumors. The tumor‘s size, location, grade, and type all influence the treatment option; a misclassification can result in inefficient therapy, needless side effects, and unfavorable patient outcomes.

However, for human professionals, interpreting these medical images for tumor categorization is a difficult and complex undertaking. Tumors can seem very different in pictures, and it frequently takes a great deal of experience and skill to tell one form from another. To comprehend and address this actual circumstance, it is necessary to require more precise computer-based automatic tumor detection/diagnosis techniques.

Deep learning, a subfield of machine learning, has emerged as a powerful technology with the ability to automatically learn hierarchical feature representations from data. It has shown remarkable success in various fields, and its application in medical image analysis, especially for brain tumor classification, has been growing rapidly in recent years [4]. Deep learning algorithms can process large amounts of medical image data and extract intricate patterns and features that may be difficult for human observers to detect. This has the potential to improve the accuracy and efficiency of brain tumor classification, providing valuable assistance to radiologists and oncologists in the diagnosis and treatment planning process. An end-to-end, fully autonomous deep learning system for brain tumor discrimination was created by the author in [5]. Transfer Learning technology was used to refine the pre-trained DenseNet201 model, which resulted in an extremely high accuracy rate. For multi-classification tasks involving brain tumors, Afshar et al. created the Bayesian Capsule network (BayesCap) framework, an enhanced convolutional neural network architecture [6]. A differential deep convolutional neural network model for the automatic classification of MRI images of brain tumors was proposed by the author in [7]. The model's 99.25% classification accuracy rate was confirmed using the TUCMD dataset, which included 25,000 normal and malignant (6 tumor kinds) brain MRIs. In order to automatically classify brain cancers, Tummala et al. investigated both single and integrated multiple pre-trained fine-tuned visual transformer (ViT) models based on ImageNet. The findings demonstrate that the integrated model outperforms all separate models and alternative configurations, achieving a maximum test accuracy rate of 98.7% at a resolution of 384×384 [8]. ViT integration is an effective computer-aided brain tumor diagnosis technique that can reduce radiologists‘ workload, according to this study.

This article‘s primary goal is to perform a thorough examination of the current deep learning-based brain tumor classification techniques. It seeks to give a thorough rundown of the many deep learning architectures that are employed, including convolutional neural networks (CNNS) and their variations. Furthermore, the assessment criteria employed to gauge the effectiveness of these classification models were thoroughly examined, along with the difficulties and potential paths forward in this area of study.

2. Method

2.1 Convolutional Neural Networks

Convolutional neural network is a deep learning model primarily used for visual data tasks like image recognition. It processes input data (e.g., images) through layers: convolutional layers apply learnable filters to extract spatial features via convolution operations, pooling layers reduce dimensionality to maintain efficiency, and activation functions introduce non-linearity. This architecture allows CNNs to hierarchically learn complex patterns, making them robust for image classification and other applications like speech recognition or natural language processing. Training involves optimization algorithms (e.g., stochastic gradient descent) and backpropagation to minimize prediction errors. In essence, CNNs excel at automating feature extraction from structured data [9].

2.1.1 Hierarchical CNN framework

To automate the identification and classification of brain tumors, Khan et al. suggested a CNN-based Hierarchical

Deep Learning-Based Brain Tumor Classifier (HDL2BT). A hierarchical deep learning framework that was created to categorize brain MRI pictures into four groups—glioma, meningioma, pituitary, and no-tumor—was their main breakthrough. Their method began with data acquisition, sourcing a total of 3,264 raw MRI images from Kaggle via an IoMT layer. Subsequently, these input images underwent preprocessing, involving normalization and resizing to ensure compatibility with the CNN architecture. Finally, the core of the method involved CNN application and classification: a CNN architecture was implemented for feature extraction, utilizing convolutional layers, ReLU activation functions, and pooling layers. Following their extraction, the features were sent into a classifier, where they were transformed into probabilities for each of the four output classes using a Softmax transformation function. With an overall accuracy of 92.13% on the validation set (13% of the data), they used backpropagation to optimize the model during training (with 87% of the data), which they claimed was better than a number of current approaches [10].

2.1.2 The CNN model optimized through bayesian optimization

Ait Amou et al. proposed a novel CNN model optimized using Bayesian Optimization for classifying brain tumors (Glioma, Meningioma, Pituitary) from T1-weighted CE-MRI images. Their key methodological innovation was the automated selection of critical CNN hyperparameters (activation function, batch size, dropout rate, dense nodes, optimizer) using Bayesian Optimization (implemented with Scikit-optimize). This replaced manual tuning, significantly improving efficiency and performance. They first designed a custom base CNN architecture (5 convolutional layers, 5 max-pooling layers, 2 dense layers, 1 dropout layer), then used Bayesian Optimization over 40 iterations to find the optimal hyperparameter set. The optimized model, trained from scratch on the Figshare dataset (3,064 images), achieved 98.70% validation accuracy, outperforming five fine-tuned pre-trained models (VGG16, VGG19, ResNet50, InceptionV3, DenseNet201) and other state-of-the-art methods, demonstrating the effectiveness of automating hyperparameter tuning for this task [11].

2.2 Capsule Network

Capsule Networks (CapsNets) are an advanced deep learning architecture designed to overcome limitations of CNNs, such as inefficiency in encoding spatial hierarchies (e.g., pose, orientation) and reliance on large datasets. Unlike CNNs, which output scalar activations, capsules output vectors whose magnitude represents the probability of an entity's presence and orientation captures its instantiation parameters (e.g., pose, deformation). CapsNets employ dynamic routing-by-agreement mechanisms (e.g., iterative cosine similarity or expectation-maximization) between capsule layers to establish part-whole relationships, ensuring equivariance to transformations. This allows CapsNets to recognize objects robustly under affine variations (e.g., rotated faces) without extensive data augmentation. They demonstrate superior performance on tasks requiring spatial awareness (e.g., image segmentation, medical imaging) and adversarial robustness but face challenges with complex datasets (e.g., CIFAR-10). CapsNets offer enhanced interpretability and parameter efficiency but remain computationally intensive for real-time applications [12].

2.2.1 Boundary-guided CapsNet

The sensitivity of CapsNets to irrelevant background in the classification of magnetic resonance imaging of brain tumors was examined by Afshar et al. They suggested a better capsule network architecture that makes use of approximate tumor boundaries, or simple bounding boxes, so that the model can concentrate on the tumor region and the surrounding tissues that are important for diagnosis without requiring exact segmentation. Convolutional and capsule layers are used in their model to process the full brain magnetic resonance imaging pictures. Before going through the entire connection layer and a softmax classifier, the output vector of the last capsule layer is concatenated with the tumor border coordinates. This is crucial. This integration directs attention while preserving the context. The technique obtained a classification accuracy of 90.89% when tested on a test dataset with 3064 photos [13]. This is significantly superior to the performance of its competitors. This method indicates that combining rough spatial guidance can effectively improve the performance of capsule networks while reducing the reliance on expert annotations.

2.2.2 BayesCap model

Afshar et al. propose BayesCap, a Bayesian CapsNet framework for brain tumor classification from MRI images. Their key innovation is integrating Bayesian inference into the CapsNet architecture to quantify prediction uncertainty, a critical aspect often missing in standard deep learning models for medical diagnosis. They model the CapsNet's prediction weight matrices as probability distributions (using variational inference approximated by Gaussians) rather than point estimates. During testing, they perform Monte Carlo sampling from these learned distributions over multiple forward passes. The mean prediction provides the tumor class, while the entropy (H) across these samples serves as a measure of prediction

ISSN 2959-6157

uncertainty. This allows filtering uncertain predictions for expert review, improving overall accuracy and enabling a human-in-the-loop system [6].

2.3 Vision Transformer

Deep learning models called Vision Transformers (ViTs) modify the Transformer architecture, which was first created for natural language processing, to fit computer vision applications. They divide images into fixed-size patches, then add positional information and transform the patches into linear embeddings (tokens). Unlike CNNs, which concentrate on local features via convolutional filters, ViTs can simulate long-range, global dependencies throughout the entire image because to its fundamental mechanism, multi-head self-attention (MSA). ViTs attain cutting-edge outcomes in tasks such as segmentation, object identification, and picture classification [14].
Tummala et al. pioneered the use of ensembled ViTs for brain tumor classification from T1-weighted contrast-enhanced MRI. The authors fine-tuned four pretrained ViT models (B/16, B/32, L/16, L/32) originally trained on ImageNet, adapting them to medical imaging by replicating single-channel MRI data into three-channel inputs. They processed 3,064 MRI slices across three tumor types (meningiomas, gliomas, pituitary tumors) at dual resolutions (224×224 and 384×384), optimizing hyperparameters (e.g., Adadelta optimizer for high-resolution training). The core innovation involved averaging softmax outputs from all individual ViTs to form an ensemble classifier, which significantly enhanced robustness. This ensemble achieved 98.7% test accuracy at 384×384 resolution—outperforming both single ViT models and prior CNN-based methods—while attaining 100% accuracy for glioma detection. The work demonstrated ViTs‘ superiority in capturing long-range dependencies in MRI and established resolution-dependent performance gains [8].

3. Discussion

3.1 Interpretability

One of the most crucial aspects of using deep learning models in clinical practice is their interpretability. Deep neural networks have produced impressive results, but their use has been constrained by their opaque decision-making [15]. Clinicians must comprehend how deep neural network models produce predictions if they are to have faith in their judgment. In order to increase trust and adoption in the clinical setting, future research should concentrate on creating strategies to make the models used in the classification of brain tumors more interpretable. This will allow clinicians to comprehend the logic underlying the predictions.
To address this critical need for interpretability, visual explanation methods have emerged as powerful tools. These techniques generate intuitive visual representations, such as heatmaps (e.g., Grad-CAM, Layer-wise Relevance Propagation - LRP) or class activation maps (CAM), that highlight the specific regions within medical images (like MRI or CT scans) most influential to the model‘s prediction. By revealing where the model is „looking" and attributing importance to anatomical structures or potential lesions, these visualizations provide clinicians with transparent insights into the model‘s decision-making process [16]. This direct visual correlation helps clinicians understand the reasoning behind a classification (e.g., benign vs. malignant tumor), verify if the model focuses on clinically relevant features, and identify potential errors or biases.
Consequently, enhancing trust and facilitating the adoption of deep learning models in critical clinical settings like brain tumor classification becomes significantly more achievable when the model‘s predictions are accompanied by such human-interpretable visual evidence.

3.2 Limited access to experts

The application of deep learning in the diagnosis and classification of brain tumors faces a core challenge of ensuring the clinical effectiveness of the model results and the feasibility of its large-scale deployment. This process is highly complex and requires the collaborative efforts of deep learning experts and clinical medical experts. The deep learning experts are responsible for data preparation, technology selection, model development and interpretation. However, the output of the model must undergo strict clinical verification: medical professionals need to evaluate the diagnostic accuracy, reliability and clinical value of the predictions. This verification is usually accomplished by comparing the model results with the gold standard clinical diagnosis. Due to the strict requirements for the deep integration of these two professional fields (technology development and clinical evaluation), building a system that serves the general public on a large scale becomes extremely difficult. The continuous coordination and mobilization of the required multidisciplinary expert resources to meet the large-scale public needs are almost impossible to achieve in reality. Therefore, the current opportunities to fully utilize ML technology to improve the efficiency of brain tumor diagnosis are mainly limited to specific scenarios with relatively concentrated resources.

3.3 Applicability

The datasets used for training the model are generally obtained through medical diagnostic methods. It should be noted that many methods for diagnosing brain tumors and cancers have been established, resulting in numerous different datasets. Furthermore, there are frequently variations in the brain tumor images acquired from different hospitals and imaging equipment. As a result, a model that has been trained on one dataset could not work on another. Therefore, a portion of the research should be devoted to the model‘s domain adaptation strategies in order to guarantee the model‘s universality and applicability. Using adversarial learning or self-supervised learning as domain adaptation techniques, for example, might help the model adjust to new target domains by decreasing the distribution discrepancies between the source and target data [15].

3.4 Privacy

Privacy and security are of utmost importance, especially in the medical field. Sensitive medical data, such as pictures of brain tumors, necessitate stringent privacy and security protocols. As a result, scientists that use deep learning to categorize brain tumors always encounter extremely challenging problems. Obtaining research data is extremely challenging since patient and medical institution privacy must be maintained. It is challenging to train an effective and highly accurate model without a lot of data assistance. Future studies should focus on the privacy problem and provide solutions that guarantee the transfer learning models‘ privacy protection features. Technologies such as federated learning may be used. This technology creates a decentralized machine learning environment, consisting of multiple clients and one or more central servers working collaboratively. It retains the original training data locally on the clients and only shares the encrypted model update content with the central server, without revealing the original data [17].

4. Conclusion

The most recent deep learning techniques for classifying brain tumors were thoroughly examined in this paper. CNNs like Hierarchical CNN and Bayesian-optimized CNN, capsule networks like Boundary-Guided CapsNet and BayesCap, and ViTs, especially ensembled ViTs, are among the important designs examined. Critical challenges identified encompass limited model interpretability hinders clinical trust; the need for professionals and professional skills has restricted the deployment of the system; and patient privacy concerns restrict data access. Future research must prioritize developing explainable AI techniques, robust domain adaptation methods, efficient lightweight models, and privacy-preserving frameworks like federated learning to enable reliable clinical adoption of DL-based brain tumor classification systems.

References

[1] Sultan HH, Salem NM, Al-Atabany W. Multi-Classification of Brain Tumor Images Using Deep Neural Network. IEEE Access. 2019;7:69215–25.
[2] Saleh A, Sukaik R, Abu-Naser SS. Brain Tumor Classification Using Deep Learning. In: 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech) IEEE; 2020 p. 131–6.
[3] Ben naceur M, Akil M, Saouli R, Kachouri R. Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. Medical Image Analysis. 2020 Jul 1;63:101692.
[4] Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. Multimedia Systems. 2021 Aug 1;27(4):599–613.
[5] Sharif MI, Khan MA, Alhussein M, Aurangzeb K, Raza M. A decision support system for multimodal brain tumor classification using deep learning. Complex Intell Syst. 2022 Aug 1;8(4):3007–20.
[6] Afshar P, Mohammadi A, Plataniotis KN. BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks. IEEE Signal Processing Letters. 2020;27:2024–8.
[7] Abd El Kader I, Xu G, Shuai Z, Saminu S, Javaid I, Salim Ahmad I. Differential Deep Convolutional Neural Network Model for Brain Tumor Classification. Brain Sciences. 2021 Mar;11(3):352.
[8] Tummala S, Kadry S, Bukhari SAC, Rauf HT. Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling. Current Oncology. 2022 Oct;29(10):7498–511.
[9] Krichen M. Convolutional Neural Networks: A Survey. Computers. 2023 Aug;12(8):151.
[10] Khan AH, Abbas S, Khan MA, Farooq U, Khan WA, Siddiqui SY, et al. Intelligent Model for Brain Tumor Identification Using Deep Learning. Ejbali R, editor. Applied Computational Intelligence and Soft Computing. 2022 Jan 21;2022:1–10.
[11] Ait Amou M, Xia K, Kamhi S, Mouhafid M. A Novel MRI Diagnosis Method for Brain Tumor Classification Based on CNN and Bayesian Optimization. Healthcare. 2022 Mar 8;10(3):494.
[12] Kwabena Patrick M, Felix Adekoya A, Abra Mighty A, Edward BY. Capsule Networks – A survey. Journal of King Saud University - Computer and Information Sciences. 2022 Jan 1;34(1):1295–310.
[13] Afshar P, Plataniotis KN, Mohammadi A. Capsule Networks for Brain Tumor Classification Based on MRI Images

ISSN 2959-6157

and Coarse Tumor Boundaries. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, United Kingdom: IEEE; 2019 p. 1368–72.

[14] Khan A, Rauf Z, Sohail A, Rehman A, Asif H, Asif A, et al. A survey of the Vision Transformers and their CNN-Transformer based Variants. Artif Intell Rev. 2023 Dec;56(S3):2917–70.

[15] Anwar RW, Abrar M, Ullah F. Transfer Learning in Brain Tumor Classification: Challenges, Opportunities, and Future Prospects. In: 2023 14th International Conference on Information and Communication Technology Convergence (ICTC). Jeju Island, Korea, Republic of: IEEE; 2023 p. 24–9.

[16] Teng Q, Liu Z, Song Y, Han K, Lu Y. A survey on the interpretability of deep learning in medical diagnosis. Multimedia Systems. 2022 Dec 1;28(6):2335–55.

[17] Li L, Fan Y, Tse M, Lin KY. A review of applications in federated learning. Computers & Industrial Engineering. 2020 Nov 1;149:106854.