© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717806

# A Lightweight Convolutional Neural Network for Automated Ependymoma Detection from Brain MRI

SNEHA M[1], NAVINKUMAR D[2], SURYA PRAKASH M[3], SIVA K[4]
[1]*Assistant professor, Department of computer science and information technology, Kalasalingam Academy of Research and Education*
[2,3,4]*Department of computer science and information technology, Kalasalingam Academy of Research and Education, Krishnankoil-626126, Tamilnadu, India.*

***Abstract-* *Ependymoma is a rare tumour of the central nervous system, which can be successfully diagnosed by Magnetic Resonance Imaging (MRI) in order to provide treatment in initial stages. Automated diagnostic assistance is important since manual examination of MRI scans is labor-intensive and plagued with observer variability. This research offers the suggestion of a lightweight Convolutional Neural Network (CNN) which aims to classify brain MRI scans as ependymoma or non-tumor. The framework is trained and tested on a publicly available dataset of MRIs (3,264 images) split (with the help of stratified 70/15/15 train-validation-test dataset) and 5-fold cross validation. There are approximately 0.9 million parameters in the proposed network architecture, which ensures computational efficiency. Experimental evaluation has shown that the performance is good, which reaches 94.1% accuracy, 94.9% accuracy, 94.3% recall and F1-score, and 0.968 AUC value when given the test data. Comparative analysis against transfer learning models such as Resnet-50, VGG16 and MobileNetV2 proves that the proposed method achieves competitive results with a significantly lower formality of parameters. A lightweight Web-based deployment system coupled with PostgreSQL is implemented for structured prediction logging.***

***Index Terms* - *Ependymoma, brain magnetic resonance imaging, convolutional neural network, deep learning, medical image classification, computer aided diagnosis.***

## I. INTRODUCTION

The brain tumors are one of the greatest health problems facing the world and, in most cases, it is important to note that early and proper diagnosis of the tumors is necessary to provide successful planning of treatment. Ependymoma is a little-known CNS tumor which is a product of ependymal cells which line the ventricles and the spinal canal. Even though it is not as prevalent as other types of brain tumours, untimely or inaccurate diagnosis can have a great deal of implications on the patients. Nevertheless, the laborious aspect of manual MRI scan interpretation is lengthy, it relies on the experience of the radiology, and inter-observer error especially in large clinical volumes.

Recent innovations in deep learning have greatly influenced the automatic medical image analysis. Convolutional Neural Networks (CNNs) are especially powerful in automatically extracting hierarchical features in the spatial domain, which makes them useful for effective tumor classification of image inputs. Numerous studies have used transfer learning in the context of pretrained deep architectures such as ResNet-50 and VGG16 architectures for brain tumor detection. Despite their high levels of classification accuracy, these architectures often come with massive parameter sizes and computing demands that may restrict the deployment of these architectures in the environment with limited computing capabilities.

A customized CNN architecture (lightweight) is developed in this work to detect automated ependymoma based on the brain MRI images. It is aimed to perform competitively in diagnostics with a lower level of computational complexity. Besides the model evaluation, a deployment-based framework is carried out to facilitate inference available one in real-time and a historical collection of the prediction. The suggested solution is an attempt to make the accuracy, efficiency, and usability of AI-based medical screening systems balanced.

## II. RELATED WORK

One of the recent trends is the usage of deep learning as a dominant trend in automated brain tumor analysis based on MRI data. Convolutional neural networks (also known as CNNs) are especially good because they are able to learn hierarchical spatial representations. Pereira et al. [1] proved the usefulness of deep CNNs in segmentation of brain tumors in.

MRI showing substantial improvement compared to other traditional feature-engineering techniques. On the same note, Akkus et al. [2] uniformly reviewed the application of deep learning in the analysis of brain by MRI, suggesting CNN-based classification as an emerging field of rapid development.

Transfer learning has also enhanced learning in tumor detection challenges. Swati et al. [3] used fine-tuning strategies of pretrained CNN models to classify brain tumor and showed better generalization. Hossain et al. [4] studied the deep residual frameworks in medical imaging practice and proved that it has an improved capacity to extract features. Deep network models like ResNet-50 and VGG16 have become popular in classification tasks related to classification of MRI because they exhibit a deep ability to represent. Lightweighting MobileNetV2 has also been considered to lighten the computational load.

The new literature and architectural works have taken an interest in architectural efficiency. To enhance spatial features depiction, Afshar et al. [5] came up with capsule networks in brain tumor classification. Paul et al. [6] suggested a hybrid CNN-based model that involves handcrafted and learned features to obtain a better diagnostic. Talo et al. [7] examined the models of deep transfer learning and have exhibited high classification rates among publicly accessible MRI data sets.

Optimization and validation strategies were also the subject of a number of studies. Khan et al. [8] tested the CNN models based on large data augmentation and cross-validation to promote strength. Several pretrained architectures to detect tumors were compared by Noreen et al. [9]. Saba et al. [10] have a more recent review of AI-based diagnostic systems in medical imaging and highlighted that validation, reproducibility, and deployment readiness are important.

In spite of reported high classification accuracy of existing strategies, most of them use computationally intensive or dense architectures or only concentrate on performance data but do not target model efficiency and deployment combination. Conversely, the current study analyses a parameter-efficient custom CNN model to detect ependymoma automatically, compares it to the existing deep learning models under the same test conditions, and introduces cross-validation and deployment-oriented validation.

## III. DATASET DESCRIPTION

These experiments have been carried out based on publicly available brain MRI dataset that had been obtained on Kaggle. This dataset will include 3, 264 axial MRI images that will be divided into two categories (tumor (ependymoma)) and non-tumor. The tumor class is made up of 1,683 images and the non-tumor class consists of 1,581 images. The data is in the form of two-dimensional MRI cross-section of varied subjects saved in a form of an image.

The original images had varying spatial resolutions and to have a uniform input size to the CNN model, all the samples had been resized to 224 x 224. To stabilize training convergence, pixel intensity values were normalized to [0,1].To retain the same amount of training and evaluation classes, a stratified splitting strategy was used. The data was characterized by a 70 per cent training data, 15 per cent validation data, and 15 per cent testing data. This will give around 2,284 of training images, 490 of validation images and 490 of testing images.

The training set was data-augmented to enhance the generalization performance and minimize the overfitting. There were augmentation methods of small angle rotations (+-15deg), horizontal flipping, zoom distortions, and slight intensity differences. Only training data was augmented to avoid data leakageFive-fold cross-validation was also done to test the model robustness and minimize variance bias in estimating performance. Average and variance of assessment measures were provided by folds.

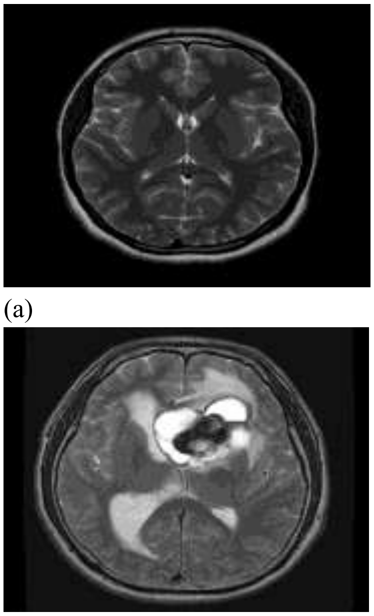

Fig. 1. Representative MRI samples from the dataset: (a) non-tumor image, (b) Tumor (ependymoma) image

## IV. METHODOLOGY

The proposed system uses a lightweight CNN architecture that can be used in the binary classification of brain MRI images as ependymoma or non-tumor. The architecture design aim is to attain the performance of competitive classification with lower computational complexity than the large transfer learning models.

### A. Data Collection

The datasets that was used in this paper was accessed via a publicly available brain MRI repository available at Kaggle. It is made up of 3 264 axial MRI images that have labels and are divided into two classes, which include tumor (ependymoma) and non-tumor.It contains 1,683 tumor images and 1,581 non-tumor images. The images are supplied in the form of two-dimensional MRI slices of various subjects that were resized to 224 × 224 before training.To provide consistency in the model input all the images have been resized before training.

### B. Data Preprocessing

To standardize the input data, preprocessing was done to enhance convergence of the models. The values of pixel intensity were scaled to be within the range [0,1] to minimize the variation in scale and stabilize gradient changes. Data augmentation was also used on the training set only to enhance generalization and lessen overfitting in data analysis.

The augmentation operations were:

a) Random rotations within ±15°
b) Horizontal flipping
c) Zoom transformations
d) Minor intensity variations

To balance the distributions of classes across splits, there was a stratified division of the dataset into 70 percent training, 15 percent validations and 15 percent of testing subsets.

### C. CNN Architecture

A binary Convolutional Neural Network (CNN) was developed that was based on a lightweight network. The architecture has four convolutional blocks, and each of their blocks contains:

a) A 3 × 3 convolutional layer
b) Batch normalizations
c) ReLU activations
d) 2 × 2 max-pooling

The number of filters become increasingly larger (32, 64, 128, 256 filters) which allows the hierarchical formation of feature extraction starting with low level edges to high level structural forms.

After the last convolutional layer, a dropout layer with the rate of 0.5 is applied to reduce overfitting of the model. The resulting feature maps are flattened and fed in fully connected dense layers.

The overall size of the trainable parameters in the proposed architecture is about 0.9 million which is significantly lower compared to the deep transfer learning architecture like the ResNet-50 and VGG16. The small number of parameters makes the inference process faster and consumes less memory.

### D. Model Optimization

Binary cross entropy was used as the loss function to train the model. The Adam optimizer was used with a scary rate 1 * 10-4 and a batch size of 32 as the model optimizer. Training for the maximum number of epochs, which was 40, was run and early stopping strategy by checking the validation loss was used to prevent overfitting.

There was also five-fold cross-validation with the view of assessing the robustness of the model and to ascertain stability between the various data partitions.

### E. System Architecture

The system in question is in the form of a modular architecture aimed at automated detection of ependymoma and proper management of results. The MRI images are sent using a web-based application and sent through the preprocessing module where they are resized to 224 224 pixels and then normalized in terms of intensity to make them compatible with the trained CNN model. In training, data augmentation is used to enhance the ability of generalization.

Depending on the probability as predicted, the decision module would classify the image to Tumor (Ependymoma) or Non-Tumor. The prediction output and its relevant metadata including image identifier, confidence score and data creation time are logged into a PostgreSQL database in a well structured way and traceable manner..The out separation in preprocessing, inference, storage and presentation functions can guarantee scalability and deployment flexibility as demonstrated in Fig 2.

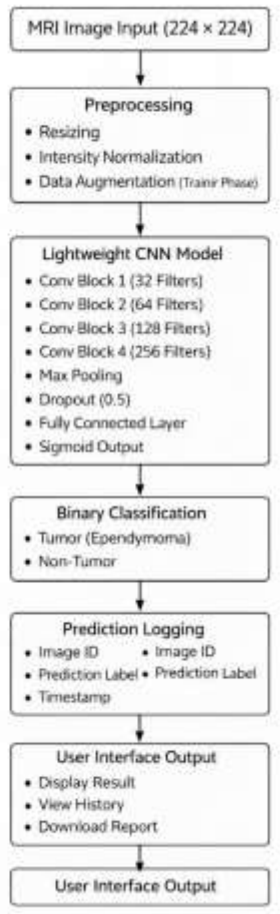

Fig. 2. Block diagram of the proposed automated system of detecting ependymoma.

## V. EXPERIMENTAL SETUP

All experiments were performed using the data described above to have stratified 70/15/15 split for training, validation and testing. The CNN model was trained with the Adam optimizer with learning rate 1*10-4, and batch size of 32. For the loss function, binary cross-entropy was used. Training was done for a maximum of 40 epochs with early stopping criteria based on validation loss to ensure that the model is not over fitted. The models employed TensorFlow/Keras implementation and a workstation with an Nvidia GPU for acceleration of the computational performance. The system used for the training consisted of an Nvidia GPU with a memory of 8GB.

Model performance was assessed with several measures of accuracy, precision, recall (sensitivity), specificity, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). A confusion matrix was created in order to analyze classification performance in detail. To evaluate the robustness, 5-fold cross-validation was carried out and the average and standard deviation of the performance measures were reported.

For purposes of comparison, the proposed lightweight CNN was tested with the existing and established architecture of transfer learning approaches like ResNet-50, VGG16 and MobileNetV2 under the same data partitioning conditions.

## VI. RESULTS AND COMPARATIVE ANALYSIS

The proposed lightweight CNN model was tested on a held out test dataset of roughly 490 MRI images. Performance assessment involved measures such as accuracy, precision, recall, and specificity as well as F1- score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The model showed excellent discriminative ability in separating ependymoma from non-tumor MRI scans with a test accuracy score of 94.1%.

The full performance metrics of the model are presented in Table I.

TABLE I
PERFORMANCE OF PROPOSED CNN MODEL

| Metric | Value |
|---|---|
| Accuracy | 94.1% |
| Precision | 94.9% |
| Recall | 93.7% |
| F1-Score | 94.3% |
| AUC | 0.968 |

The confusion matrix is a detailed summary of the results of classification. The model correctly identified 225 images of tumors and 238 images of healthy tissues, as well as had very low false positive and false negative rates.

The model correctly classified 225 tumor and 238 non-tumor images, with minimal false positives and false negatives.

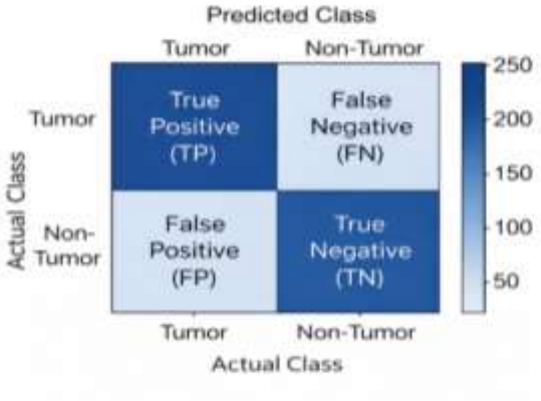

Fig. 3. Confusion matrix for CNN Model proposed on the test dataset.

The Receiver Operating Characteristics (ROC) curve, which is observed in Fig. 4, indicates a good class separability with an AUC value of 0.968.

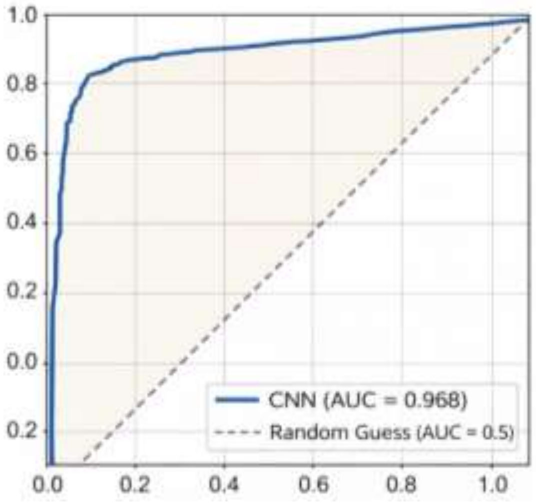

Fig. 4. ROC curve of the proposed CNN model with AUC = 0.968.

To determine robustness of the models, five-fold cross validation was used. The mean and standard deviation of the evaluation metrics of the folds is presented in Table II.

TABLE II
Results of CROSS-VALIDATING(5-FOLD)

| Metric | Mean | Std Dev |
|---|---|---|
| Accuracy | 93.8% | ±1.1 |
| Precision | 94.2% | ±1.3 |
| Recall | 93.4% | ±1.4 |
| F1-Score | 93.7% | ±1.2 |
| AUC | 0.964 | ±0.006 |

The proposed architecture was further benchmarked against the established transfer learning models with the same data partitions. The result of the comparison are shown in table III.

TABLE III
COMPARISON WITH TRANSFER LEARNING MODELS

| Model | Accuracy | F1-score | AUC | Parameters |
|---|---|---|---|---|
| Proposed CNN | 94.1% | 94.3% | 0.968 | 0.9M |
| ResNet-50 | 95.2% | 95.1% | 0.974 | 25M |
| VGG16 | 95.6% | 95.4% | 0.976 | 138M |
| MobileNetV2 | 94.8% | 94.4% | 0.971 | 3.4M |

## VII. DEPLOYMENT FRAMEWORK

The trained lightweight CNN model was implemented in a web-based inference framework to carry out real-time analysis of MRI. The deployment architecture is made of frontend interface for uploading MRI image, backend serves for preprocessing and CNN inference, and a backend PostgreSQL database for a structured prediction log. Upon submission of images, preprocessing (resizing and normalization) of images is carried out prior to model inference and then, a binary classification output with confidence score is produced.

Prediction records, with image ID, predicted label, confidence score and timestamp are added to database in order to trace the trails and keep the history. The deployed interface is shown in Fig. 5, which has shown automated upload of MRI and prediction visualization.

Fig. 5. Web-based interface for uploading of the MRI and automated output of the prediction.

## VIII. LIMITATIONS

There are several limitations existing despite the promising performance of the proposed lightweight CNN model. First, the information used in this survey is publicly-available, and is composed of slices of 2D MRI images rather than actual volumetric scans. The lack of metadata at the patient level and the lack of multi-sequence MRI modalities may have some limitations in how it can be generalized to clinical environments. In addition, the model is only used for binary classification and does not get tumor localization and segmentation.

Second, the evaluation was done on one dataset without outside validating ccs in independent clinical cohorts. Although five-fold cross validations was used to the determine robustness, additional validation on multi-center data sets is needed to determine clinical reliability on a larger scale. Future work may include the inclusion of multi-modal MRI data, external validation, and tumour localization mechanisms, to facilitate the diagnostic applicability.

## IX. CONCLUSION

This paper introduced a small size of Convolutional Neural Network (CNN) for automatic diagnosis of ependymoma on the brain MRI image. The training of the proposed framework and evaluation using a publicly available dataset using stratified data splitting and the five-fold cross-validation method were used. Experimental results showed good classification performance with an accuracy of 94.1% and the AUC of 0.968, ensuring the lightweight model with about 0.9 M trainable parameters. Evaluative comparison of the proposed approach with existing transfer learning models showed that the proposed approach provides diagnostic performance on par with much lower computational complexity.

In addition to algorithmic validation, an online deployment framework that runs in the web and is connected to the PostgreSQL implementation was achieved allowing inference to be performed in real-time with structured prediction logging. The proposed system is shown to have found an effective trade off among accuracy, efficiency, and deployment feasibility. Future efforts will be directed towards multi-class tumor classification, tumor localization, and testing on outside clinical data to improve applicability in the clinic.

## REFERENCES

[1] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain tumor segmentation with convolutional neural networks in MRI images," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1240-1251, 2016.

[2] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," J. Digit. Imaging, vol. 30, no. 4, pp. 449-459, 2017.

[3] Z. N. K. Swati et al., "Brain tumor classification for MR images utilizing transfer learning and fine tuning," Comput. Med. Imaging Graph., vol. 75, pp. 34-46, 2019.

[4] M. Talo, U. B. Baloglu, O. Yildirim and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification," Comput. Methods Programs Biomed., 2019, vol. 176, pp. 227-235.

[5] H. A. Khan et al., "Brain tumor classification using deep CNN and data augmentation," in the Proceedings of the Institute of Electrical and Electronics Engineers (IEE) Access, vol. 8, pp. 179599-179608, 2020.

[6] N. Noreen et al., "A deep learning model based on concatenation approach for brain tumor classification," in Atharva Alshayk, E. Atharva Alshayk, A. Hamza Azfar, and Obaid Alafifi, "IEEE Access," vol. 8, pp. 55135-55144, 2020.

[7] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification using the capsule networks," in Proc. IEEE ICIP, 2018, pp. 3129-3133.

[8] S. Rehman, M. Khan, T. Saba and A. Mehmood, "Microscopic brain tumor detection using deep convolutional neural networks," Microscopy Research and Technique, vol. 83, no. 11, pp. 1361-1370, 2020.

[9] A. Abiwinanda et al., "Brain tumor classification using convolution neural network," IFMBE Proc., vol. 68, pp. 183-189, 2019.

[10] M Sajjad et al, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation", J. Comput. Sci., vol. 30, pp. 174-182, 2019.

[11] T. Saba, A. Rehman, and M. Sharif, “Brain tumor detection and classification using artificial intelligence techniques,” J. Med. Syst., vol. 45, no. 4, 2021.

[12] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016.

[14] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. ICLR, 2015.

[15] M. Sandler et al., “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE CVPR, 2018.

[16] A. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, 2017.

[17] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep learning applications in medical image

analysis," IEEE Access, vol. 6, pp. 9375–9389, 2018.

[18] M. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019.