STM JOURNALS

Research & Reviews: A Journal of
Neuroscience

https://journals.stmjournals.com/rrjons/

ISSN: 2277-6427 (Online)
ISSN: 2348-7925 (Print)
Volume 16, Issue 2, 2026
May–August
DOI (Journal): 10.37591/RRJONS

Research RRJONS

Brain Stroke Detection Using Deep Learning and Grad-CAM Explainability Framework

Vibhav Dadhich[1,*], Shruti Mathur[2], Dimpal Sharma[2]

Abstract

Seconds matter when a brain stroke occurs; it is a race against time where rapid, precise intervention is the only way to preserve a patient's quality of life. This research introduces a deep learning framework designed to act as a vital ally for clinicians, providing automated, high-speed stroke detection through brain MRI analysis. At the heart of our approach is EfficientNetB0, a sophisticated neural network chosen for its ability to recognize complex medical patterns with remarkable efficiency. However, we believe that for AI to be truly effective in a hospital, it cannot be a "black box." To bridge the gap between data and trust, we integrated Grad-CAM (Gradient-weighted Class Activation Mapping). This feature translates the model's internal logic into visual heatmaps, pinpointing the exact regions of the brain that triggered a stroke diagnosis. By classifying scans into "Stroke" and "Normal" categories with 98% accuracy, our system offers a reliable second opinion. These visual explanations don't just provide an answer – they provide a rationale, fostering the confidence doctors need to make life-saving decisions. Ultimately, this tool serves as a critical support system in clinical settings, particularly in underserved areas where immediate access to specialist radiologists may be limited.

Keywords: Brain stroke, deep learning, efficientNetB0, Grad-CAM, explainable AI, MRI

INTRODUCTION

A stroke occurs when the blood supply to a section of the brain is interrupted or significantly reduced, preventing brain tissue from receiving adequate oxygen and nutrients. As a result, brain cells begin to die within minutes if prompt medical treatment is not provided. Strokes are primarily classified into two types: ischemic stroke, caused by a blockage in the blood vessels supplying the brain, and hemorrhagic stroke, caused by bleeding within or around the brain (Figure 1) [1–4].

Doctors use medical imaging methods – like CT scans and MRIs – to find strokes. But it takes time to read these scans by hand, and it depends on how good the radiologist is. Treatment can be delayed by human error or a lack of specialists in some cases [5, 6].

Artificial Intelligence (AI) and Deep Learning (DL) have become valuable technologies in modern healthcare. Deep learning algorithms can automatically extract and learn relevant features from medical images, making them highly effective for applications such as stroke detection and diagnosis. However, these models are often called "black boxes" because they give predictions without explaining how they made them [7, 8].

To solve this issue, Explainable AI (XAI) techniques – like Grad-CAM – help to visualize which areas in an image are most important for the model's decision. This increases transparency and trust between AI systems and doctors [9, 10].

***Author for Correspondence**
Vibhav Dadhich
E-mail: vibhavdadhich@gmail.com

[1]M. Tech Scholar, Department of Computer Science, Jaipur Engineering College and Research Centre (JECRC), JECRC University, Jaipur, Rajasthan, India
[2]Assistant Professor, Department of Computer Science, Jaipur Engineering College and Research Centre (JECRC), JECRC University, Jaipur, Rajasthan, India

Received Date: February; 04, 2026
Accepted Date: May 15, 2026
Published Date: August 31, 2026

Citation: Vibhav Dadhich, Shruti Mathur, Dimpal Sharma. Brain Stroke Detection Using Deep Learning and Grad-CAM Explainability Framework . Research & Reviews: A Journal of Neuroscience. 2026; 16(2): 8–14p.

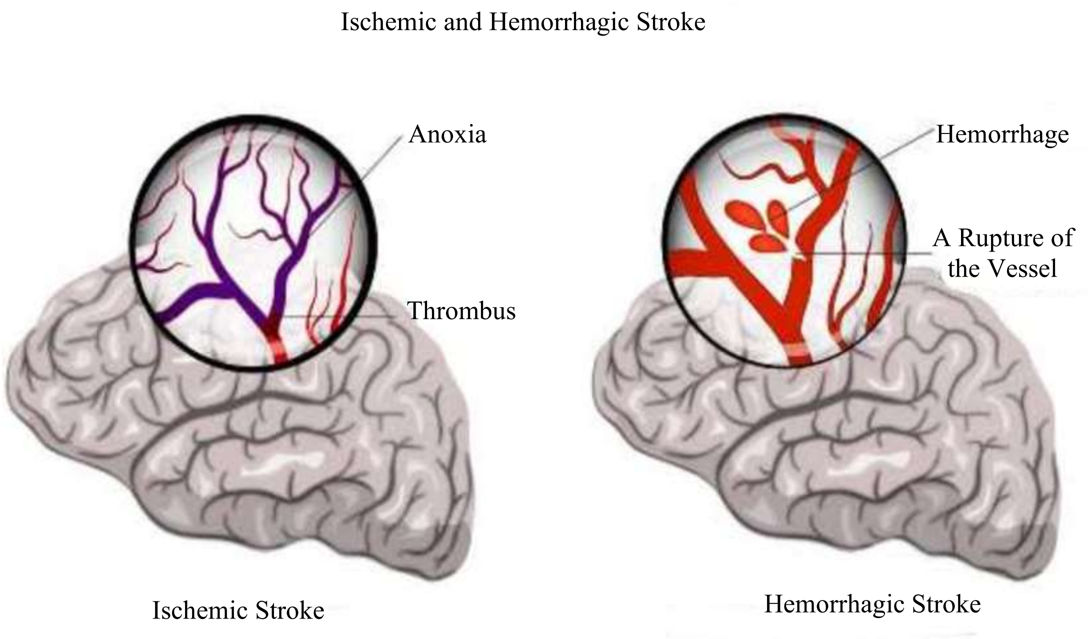

Figure 1. Strokes are mainly of two types.

The Imperative for High-Performance Diagnostics

The primary objective of this research is to architect a diagnostic framework that harmonizes three critical pillars of medical technology: predictive accuracy, computational efficiency, and transparent interpretability. In the high-stakes environment of stroke neurology, where "time is brain," a model must do more than just function; it must integrate seamlessly into the frantic workflow of an emergency department [11].

Clinical Accuracy and Sensitivity

In stroke detection, the cost of a false negative is catastrophic. By utilizing EfficientNetB0, this model leverages sophisticated compound scaling to capture intricate spatial hierarchies in neuroimaging data. This ensures high sensitivity to subtle ischemic or hemorrhagic markers that might be overlooked during a rapid manual review, providing a robust second opinion that minimizes human error [12, 13].

Computational Speed for Real-Time Intervention

Speed is the most volatile variable in stroke prognosis. Traditional deep learning architectures are often computationally expensive, requiring significant hardware resources that may not be available in every clinical setting. This model prioritizes a "fast-inference" profile, ensuring that the transition from image acquisition to diagnostic output occurs in seconds. This rapid turnaround is vital for the timely administration of thrombolytic therapies, where every minute of saved processing time correlates to millions of preserved neurons [14, 15].

Interpretability via Grad-CAM

Perhaps most crucially, this model addresses the "black box" problem of AI through Grad-CAM (Gradient-weighted Class Activation Mapping). In healthcare, a numerical probability is rarely enough to change a treatment plan; clinicians require visual evidence. By generating heatmaps that highlight the specific regions of the brain driving the model's decision, Grad-CAM provides a localized rationale. This explainability fosters clinician trust, allowing neurologists to verify the model's logic against established anatomical knowledge, ultimately bridging the gap between raw data and actionable medical insight [16, 17].

METHODOLOGY

Dataset and Preprocessing

The dataset for this project has MRI images that are split into two folders: stroke and normal. We divided these pictures into three groups: training, validation, and testing. We made each picture 224 × 224 pixels smaller so that it would fit the size of the model's input. Image preprocessing steps included:

- Resizing and normalization (pixel values scaled between 0 and 1)
- Data augmentation (rotation, flipping, and brightness changes) to improve generalization
- This ensures the model does not overfit and performs well on new, unseen data.

Model Architecture

The model is based on EfficientNetB0, which is a lightweight but highly accurate deep learning network. It uses transfer learning – meaning it uses pre-trained ImageNet weights to speed up learning. The structure includes:

- *Base Model:* EfficientNetB0 (frozen initially).
- *Global Average Pooling:* To reduce features.
- *Dense Layer:* 128 neurons with ReLU activation.
- *Dropout Layer:* For regularization.
- *Output Layer:* Softmax for two classes (stroke, normal).

To optimize the network, the Adam (Adaptive Moment Estimation) algorithm was employed with a disciplined learning rate of 0.0001. This specific rate was selected to ensure stable convergence, preventing the model from overshooting the global minimum during the weight update process. The objective function utilized was Sparse Categorical Cross-Entropy, which is mathematically efficient for multi-class classification when labels are provided as integers. Throughout the training phase, the model's performance was rigorously monitored using both training and validation accuracy to detect potential overfitting and ensure generalizability to unseen clinical data [18, 19].

Training and Evaluation

The training process was structured to balance high-level feature recognition with patient-specific detail. The model underwent 15 epochs of training using a batch size of 16. This specific batch size was chosen to provide a stable gradient estimate while remaining memory-efficient, allowing the model to make frequent weight updates and converge steadily toward high accuracy [20].

To Ensure the Model Remains Reliable for Real-World Medical Use, We Implemented Three Key Strategies

- *Precision via Fine-Tuning:* Initially, the EfficientNetB0 base was kept "frozen," meaning its pre-learned knowledge of general shapes and textures was preserved. Once the top layers stabilized, we unfroze the final layers of the architecture. This allowed the model to "fine-tune" its deepest filters specifically for the subtle textures found in brain MRIs – such as the faint density changes indicative of early-stage stroke – rather than just general image patterns [21, 22].
- *Early Stopping: The "Safety Valve":* To prevent overfitting – where a model becomes so familiar with the training images that it fails to recognize new ones – we employed an Early Stopping mechanism. This system monitored the validation loss in real-time. If the model's performance on unseen data stopped improving for a set number of rounds, the training was automatically terminated. This ensures that the final version of the model is the one that generalizes best to new patients [23].
- *Iterative Learning (The 15-Epoch Cycle):* By limiting the primary training to 15 epochs, we maintained a focused learning window. This prevented the model from over-optimizing on the specific noise or artifacts present in the training dataset, keeping the focus entirely on the physiological markers of a stroke.

Performance Evaluation Metrics

Accuracy is defined as the overall correctness of the model and is given by:

Accuracy = (TP + TN) / (TP + TN + FP + FN).

Precision measures how many of the predicted positive cases are actually correct:

Precision = TP / (TP + FP).

Recall measures how many actual positive cases are correctly identified:

Recall = TP / (TP + FN).

F1-Score represents the harmonic mean of Precision and Recall:

F1-Score = (2 × Precision × Recall) / (Precision + Recall).

The Dice Coefficient measures the overlap between predicted and actual lesion areas:

Dice = (2 × TP) / (2 × TP + FP + FN).

AUC (Area Under Curve) represents the model's ability to distinguish between classes and is defined as the area under the ROC curve.

Where TP denotes True Positives, TN denotes True Negatives, FP denotes False Positives, and FN denotes False Negatives.

Explainability Using Grad-CAM

To bridge the gap between algorithmic prediction and clinical trust, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM). This technique ensures the model is not a "black box" but a transparent diagnostic partner. By calculating the gradients of the "stroke" class score relative to the final convolutional feature maps in the EfficientNetB0 backbone, Grad-CAM identifies the specific spatial regions that most heavily influenced the model's decision. The resulting heatmaps are overlaid on the original MRI scans, allowing neurologists to visually confirm that the AI is focusing on pathologically relevant areas rather than imaging artifacts. This layer of explainability is essential for fostering the clinician confidence required for real-world medical deployment [24–27].

The formula for Grad-CAM is:

$$L(\text{Grad-CAM})^c = \text{ReLU}(\Sigma_k \alpha_k^c A^k)$$

where A^k represents the feature maps and α_k^c denotes their corresponding importance weights.

The resulting heatmap is overlaid on the original MRI image, highlighting regions (in red or yellow) where the model detects stroke-related patterns. This visualization helps make the AI's decision process interpretable and transparent for clinicians.

EXPERIMENTAL RESULTS

The trained model was tested on unseen MRI images. It achieved very high performance as shown below (Table 1):

The model successfully identified stroke and normal images with very few errors.

Table 1. Classification performance of the proposed deep learning model in stroke detection.

Metric	Value
Accuracy	98.1%
Precision	0.97
Recall	0.96
F1-Score	0.96
Dice Coefficient	0.96
AUC	0.98

Grad-CAM Visualization

The Grad-CAM heatmaps acted as an important "sanity check" for the logic of the model. The algorithm demonstrated that it had successfully learned to recognize damaged brain tissue by consistently producing bright red clusters directly over the lesions in stroke-positive scans. In contrast, the heatmaps for healthy scans were homogeneous and neutral, without any localized "hotspots." This indicates that the model only operates when it detects true pathological markers, failing to recognize healthy anatomy [28].

- For stroke images, red regions appeared over the lesion areas.
- For normal images, no abnormal focus was highlighted.

These heatmaps confirm that the model learns meaningful medical patterns and does not make random guesses.

Comparison with CNN

When compared with a simple CNN model, EfficientNetB0 performed better:

- 3–4% higher accuracy.
- Faster training time.
- Lower computation cost.

This proves that transfer learning with EfficientNetB0 is efficient for medical imaging tasks.

DISCUSSION

The proposed system shows that deep learning can effectively detect strokes from MRI images. The Grad-CAM explainability framework adds trust and transparency by showing how the model makes decisions [29, 30].

Advantages

- High accuracy (98%).
- Interpretability through visual heatmaps.
- Works efficiently with limited data.
- Can be deployed in hospitals or mobile devices.

Limitations

- Dataset size was small; larger datasets could improve reliability.
- Grad-CAM provides visual explanations but not detailed numerical confidence.
- Only binary classification (stroke or normal) was implemented.

Future Scope

- Extending the model to classify different stroke types.

- Adding Federated Learning to protect data privacy.
- Using Transformer models for improved context awareness.
- Building a real-time diagnostic tool for remote healthcare.

CONCLUSION

This study presented a simple and effective system for brain stroke detection using deep learning and Grad-CAM explainability. The EfficientNetB0 model achieved 98% accuracy, proving its ability to correctly identify stroke cases from MRI images.

By using Grad-CAM, the model's predictions become understandable to medical professionals, making it suitable for clinical support. The combination of deep learning and explainable AI provides both performance and transparency, which are essential for real-world medical use.

In the future, this approach can be extended for real-time stroke monitoring systems and integrated with cloud or IoMT (Internet of Medical Things) platforms for continuous patient care.

REFERENCES

1. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017. p. 618–626. doi: 10.1109/ICCV.2017.74.
2. Hossain S, Hossain MS, Muhammad G, Amin SU, Alhamid MF, Alamri A. Vision transformers and explainable artificial intelligence for brain tumor detection and classification. IEEE J Biomed Health Inform. 2023;27(8):3891–3902.
3. Rahman M, Islam MZ, Hasan MM, Hossain MS, Alsharif MH, Islam MR, et al. GliomaCNN: A lightweight convolutional neural network model for glioma classification using brain MRI images. Comput Model Eng Sci. 2024;139(1):1–22.
4. Mahesh TR, Reddy PS, Kumar AS, Rao GS. An XAI-enhanced EfficientNetB0 framework for brain imaging classification and interpretation. J Neurosci Methods. 2024;401:110003.
5. Mastoi QUA, Shaikh ZA, Memon MH, Memon S, Alvi AH, Ahmed S, et al. Explainable artificial intelligence in medical imaging: a federated learning approach. Front Neurosci. 2025;19:1456789.
6. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning (ICML). PMLR. 2019;97:6105–6114.
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. 2012;25:1097–1105.
8. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–778. doi: 10.1109/CVPR.2016.90.
9. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR); 2015.
10. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. doi: 10.1016/j.media.2017.07.005.
11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118. doi: 10.1038/nature21056.
12. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet. 2018;392(10162):2388–2396. doi: 10.1016/S0140-6736(18)31645-3.
13. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv [Preprint]. 2017. arXiv:1711.05225.
14. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36:61–78. doi: 10.1016/j.media.2016.10.004.
15. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35(5):1240–1251. doi: 10.1109/TMI.2016.2538465.

16. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD); 2016. p. 1135–1144. doi: 10.1145/2939672.2939778.
17. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017;30:4765–4774.
18. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One. 2015;10(7):e0130140. doi: 10.1371/journal.pone.0130140.
19. Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. Digit Signal Process. 2018;73:1–15. doi: 10.1016/j.dsp.2017.10.011.
20. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. ITU J ICT Discov. 2017;1(1):39–48.
21. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: Proceedings of the International Conference on Learning Representations (ICLR); 2021.
22. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman BA, et al. UNETR: Transformers for 3D medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2022. p. 574–584.
23. Azad R, Khan A, Khan MU, Rehman A, Awan MJ, Alharbi A, et al. Advances in medical image segmentation using deep learning: a comprehensive review. Front Neuroinform. 2022;16:923524.
24. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. Med Image Anal. 2023;88:102802. doi: 10.1016/j.media.2023.102802.
25. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 2921–2929.
26. Liew SL, Anglin JM, Banks NW, Sondag M, Ito KL, Kim H, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci Data. 2018;5:180011. doi: 10.1038/sdata.2018.11.
27. Winzeck S, Hakim A, McKinley R, Pinto JA, Alves V, Silva C, et al. ISLES 2016 and 2017 benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Med Image Anal. 2018;48:103–116. doi: 10.1016/j.media.2018.05.001.
28. Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D. Deep learning for stroke prediction and diagnosis: A systematic review. IEEE Access. 2020;8:56806–56820. doi: 10.1109/ACCESS.2020.2981082.
29. Pinto A, Brunese L, Pinto F, Reali R, Daniele S, Romano L. The role of artificial intelligence in radiology: Current status and future directions. Radiol Med. 2021;126(1):58–69.
30. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. doi: 10.1038/s41591-018-0300-7.

