
Predicting Early Functional Outcomes in Acute Ischemic Stroke Using Interpretable Deep Learning

Arun Kumar R¹, Dayananda P¹, Niranjnamurthy M²

1School of Computer Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education (MAHE), Bengaluru, India

2 Department of AI and ML, BMS Institute of Technology and Management, Bengaluru, India

Abstract.

Acute ischemic stroke is one of the major causes for the death and lifelong disability, where early treatment planning is very essential to improve the patient's functional outcome. However, clinicians restrict the adoption of current deep learning models because of its limited interpretability. This study puts forward an interpretable framework for deep learning that connects Convolutional Neural Network (CNN) with SHapley Additive exPlanations (SHAP) to predict the patient's outcome using neuroimaging data, which also explains the imaging features contributing to model predictions. Compared with conventional approaches that mainly focus on accuracy of predication, the proposed framework brings together prediction and explainability to improve clinical decision making and transparency. The framework has been evaluated using the publicly available ischemic stroke datasets with Accuracy, Balanced F-score, Recall and Area Under the Curve (AUC). The Benchmark evaluation shows that the framework predicts the functional outcomes more effectively and also provide the clinically meaningful explanations through SHAP. Future work will focus on multicentre validation and integration of additional clinical variables.

Keywords. Acute ischemic stroke, functional outcome prediction, deep learning, explainable artificial intelligence, SHAP, neuroimaging, clinical decision support.

1. INTRODUCTION

Stroke being one of the main causes for the death and lifetime neurological disability, creating a significant burden on healthcare system, patients and the society. Among various categories of stroke, acute ischemic stroke reports nearly 85% of the cases and requires quick assessment by the clinicians to minimize the brain injury and enhancing the patient outcomes [1]. Early diagnosis, prompt treatment, and accurate prognostic assessment are very essential for selecting right therapeutic interventions and planning rehabilitation strategies [2,3].

Though there has been a significant advancement in stroke management, predication of early functional outcome always remains challenging because of the differences in the lesion

characteristics, treatment response and the neurological status, that contributes to the recovery of the patients. Even though clinicians depend on the neurological assessment scores and neuroimaging findings for the treatment plan, patients with almost similar clinical presentations mostly exhibit different trajectories of recovery, limiting the dependency of the conventional prognostication assessment [3,7].

Artificial intelligence has been steadily explored for the prognostication of the stroke due to its ability to learn the complex patterns from neuroimaging and the clinical data [4,7]. Many deep learning models often function as a black-box system, with limited knowledge to the factors that influence the outcome predication and reducing the trust of clinicians in artificial intelligence-assisted systems [8].

To overcome these limitations, this research presents an interpretable framework, involving deep learning to predict the functional outcomes in acute ischemic stroke. The proposed framework combines the Convolutional Neural Network (CNN) with SHapley Additive exPlanations (SHAP), which identifies the imaging features that contribute to the prediction of the patient outcome and model transparency. However, the conventional approaches mainly focused on the prediction of accuracy.

The primary innovation of this research is as follows

- An interpretable approach based on CNN, for early functional outcome prediction using neuroimaging data.
- SHAP integration for the improvement of transparency and model predictions.
- Publicly Available Dataset – For model evaluation and Prognostic assessment.

2. RELATED WORK

2.1. *Deep Learning in Stroke Outcome Prediction*

Deep learning and Artificial intelligence emerged as a very important factor for the stroke prognosis due to their ability to analyse the complex neuro-imaging and clinical data. Previously, the prediction methods mainly depended on clinical score and statistical models which were capturing only the non-linear relationships linked with stroke progression [7]. However, of late Convolutional Neural Network (CNNs) did show very competent results in learning imaging features for characterization of lesions and prediction of functional outcomes [5-7]. Yet the current models provide very limited explanation of the factors influencing outcome prediction and limiting the acceptance by the clinicians.

2.2. *Neuroimaging-Based Prognostic Analysis*

Neuroimaging is crucial in assessment of the acute ischemic stroke for improving the patient recovery, through the detailed information on lesion extent, infarct location and abnormalities in the tissues [6]. MRI-based imaging biomarkers are mainly studied for prediction of functional outcome and prognosis estimation [7]. The prediction reliability always remains affected by variations in patient specific factors and lesion characteristics, which results in need for the more reliable prognostic models [2].

2.3. *Explainable Artificial Intelligence in Healthcare*

The growing use of artificial intelligence in healthcare has demonstrated the need for model clarity, mainly in clinical applications where the decision of the treatment needs a clear justification. The transparency of the model is achieved by the Explainable Artificial Intelligence (XAI) which highlights the factors leading to outcome predication [4]. Clubbing the SHAP with Deep Learning will improve the clinicians understanding of the model behaviour and helps them in decision-making.

3. RESEARCH GAP AND MOTIVATION

Latest developments in the Deep Learning and Artificial Intelligence have increased the outcome prediction of the patients using the neuroimaging data. On the other hand, the current approach focuses on the accuracy of the prediction rather than clinical explainability.

Along with this, because of the diverse nature of the ischemic stroke, including response to the treatment, neurological status and the variations in the characteristics of the lesions, the prediction of the functional outcome becomes more difficult and challenging. Even though the current models can capture the complex relationships, they lack transparency in explaining the prediction outcomes.

To overcome these limitations, the proposed study uses the neuroimaging data for the prediction of the patient's outcome with an explainable framework using Deep Learning. The framework combines the SHapley Additive exPlanation (SHAP) with CNN for making the model more transparent and explainable, and that encourages the clinicians to accept the prediction and improve the stroke management.

4. PROPOSED METHODOLOGY

This section presents the proposed interpretable deep learning framework for predicting early functional outcomes in acute ischemic stroke using neuroimaging data. The framework integrates a Convolutional Neural Network (CNN) for outcome prediction with SHapley Additive exPlanations (SHAP) to improve model interpretability. The overall workflow comprises dataset preparation, image preprocessing, CNN-based prediction, and SHAP-based interpretability analysis, as illustrated in Figure 4.1.

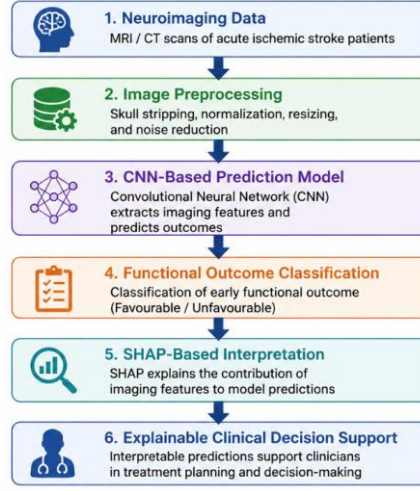


Figure 4.1. Overall workflow of the proposed interpretable deep learning approach for early functional outcome prediction in acute ischemic stroke.

4.1. Dataset Description

Publicly This framework includes the publicly available ischemic stroke dataset. The neuroimaging data provides the relevant information on infract location, characteristics of the lesion and the abnormalities within the brain tissues that affects the patient's recovery. The dataset has been pre-processed, normalized and resized to reduce the level of variations and to achieve the standard input data.

4.2. Proposed Deep Learning Model

To automatically learn the imaging features without any handcrafted descriptors, the CNN architecture was selected. The proposed framework generates the prediction of the functional outcomes as a binary classification, which includes the favourable and unfavourable recovery. The CNN also consists of the classification stages and feature extraction, that shows the imaging feature linked with severity of the stroke.

4.3. Interpretability Through SHAP

To increase the prediction transparency, SHAP was included into the framework to measure the value of imaging features to predict the outcomes individually. SHAP basically provides both local and global explanations, and helps visualize the elements that influenced the model prediction.

5. EXPERIMENTAL SETUP

The proposed interpretable Deep Learning framework was evaluated using the publicly available dataset for ischemic stroke. The input data was pre-processed with normalization and resizing to ensure the consistency.

The supervised learning technique was used to train the model to perform the binary classification such as favourable and unfavourable outcomes. The classification objective was Cross-Entropy loss. The evaluation of the performance was done using Recall, F1-Score, Precision and Area Under the Receiver Curve. The complete experimental configuration is shown in **Table 1**.

Table 1. Experimental configuration of the proposed interpretable deep learning model

Parameter	Configuration
Prediction Task	Functional Outcome Prediction
Learning Type	Binary Classification
Model Architecture	CNN-based Deep Learning
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, AUC
Explainability Method	SHAP

6. RESULTS AND DISCUSSION

The interpretable Deep Learning model was evaluated as shown in the **Table 2**.

Table 2. Predictive performance of the proposed interpretable deep learning model

Metric	Score
Accuracy	0.8921
Precision	0.8817
Recall	0.8679
F1-score	0.8747
AUC	0.9125

Beyond the predictive performance, the individual prediction outcomes were achieved from the enhanced model transparency using SHAP. The combination of SHAP and CNN provides the provide the meaningful insight to the clinicians for the better decision making. Nevertheless, further validation using larger multicentre datasets is required to improve model generalizability.

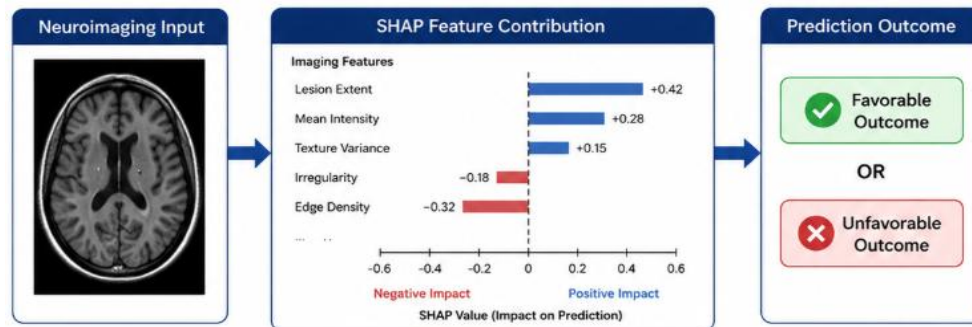


Figure 6.1. Conceptual illustration of SHAP-based explainability showing the contribution of neuroimaging features to favourable and unfavourable functional outcome prediction.

7. CONCLUSION

This study showcases the interpretable framework using Deep Learning for predicting the functional outcomes of the patient's, which combines the CNN and SHAP predictive performance with model interpretability, supporting more transparent clinical decision-making system.

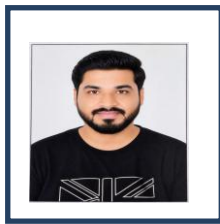
Experimental results demonstrated encouraging predictive performance in distinguishing favourable and unfavourable functional recovery outcomes. In addition, SHAP-based explainability provided insights into the imaging features influencing model predictions, improving the transparency of the proposed framework. Overall, the study shows that integrating prediction with explainability can support more transparent early stroke outcome assessment.

8. REFERENCES

- [1] Feigin V L, Stark B A, Johnson C O, et al.: Global, regional, and national burden of stroke and its risk factors. *The Lancet Neurology*, 20(10), 795–820, 2021.
- [2] Campbell B C V, Khatri P.: Stroke. *The Lancet*, 396(10244), 129–142, 2020.
- [3] Powers W J, Rabinstein A A, Ackerson T, et al.: Guidelines for the early management of patients with acute ischemic stroke. *Stroke*, 50(12), e344–e418, 2019.
- [4] Topol E J.: High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56, 2019.
- [5] Esteva A, Robicquet A, Ramsundar B, et al.: A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29, 2019.
- [6] Lundervold A S, Lundervold A.: An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, 29(2), 102–127, 2019.
- [7] Heo J, Yoon J G, Park H, et al.: Machine learning-based model for prediction of outcomes in acute stroke. *Stroke*, 50(5), 1263–1265, 2019.

- [8] Tjoa E, Guan C.: A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813, 2021.
- [9] Lundberg S M, Lee S I.: A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 4765–4774, 2017.
- [10] Samek W, Wiegand T, Müller K R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries*, 1(1), 39–48, 2017.

Biographies



Arun Kumar R is a Research Scholar at the School of Computer Engineering, Manipal Institute of Technology Bengaluru, MAHE. He has over 13 years of experience in Healthcare IT. His research interests include explainable AI, medical image analysis, and stroke prognosis.