World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

AI-enhanced predictive modeling for acute ischemic stroke: Advancing diagnosis accuracy and patient outcomes

Suhani Gupta [1,*], Harveer Saini [2] and Mayur Dalvi [3]

[1] *Department of Youth Wellness and Research, Neuro Health Alliance, Dublin, California, USA.*

[2] *Department of Innovation & Research, National AI Youth Council, Dublin, California, USA.*

[3] *Department of Data Science, CU Boulder, Denver, Colorado, USA.*

World Journal of Advanced Research and Reviews, 2025, 25(02), 2734-2743

Publication history: Received on 05 December 2024; revised on 28 January 2025; accepted on 31 January 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.2.0147

Abstract

Acute ischemic stroke (AIS) requires rapid and accurate diagnosis to enable timely treatment and improve patient outcomes. This study presents an AI-enhanced predictive modeling approach for AIS that integrates advanced machine learning algorithms to improve diagnostic accuracy and provide reliable outcome predictions. We retrospectively collected clinical and imaging data from AIS patients and developed a predictive model combining a convolutional neural network (CNN) for early stroke detection on brain imaging with gradient boosting machine learning for prognostic outcome prediction. The model was trained and validated on separate cohorts and evaluated against standard clinical assessment and risk scores. Key results demonstrate that the AI-enhanced model achieved 96% sensitivity and 94% specificity for AIS detection, outperforming conventional clinical assessment (85% sensitivity, 88% specificity). It also accurately predicted 90-day functional outcomes with an area under the ROC curve (AUC) of 0.90, significantly higher than a baseline logistic model (AUC 0.82, $p<0.01$). These results indicate a substantial improvement over traditional methods. The integrated approach not only expedited stroke diagnosis but also provided robust prognostic insights, which together can support clinicians in making timely, informed treatment decisions. As a whole, the proposed AI-driven model significantly advances stroke diagnostic accuracy and outcome prediction, showcasing its potential to enhance acute stroke care and patient outcomes.

Keywords: Acute Ischemic Stroke; Artificial Intelligence; Machine Learning; Diagnostic Accuracy; Outcome Prediction; Predictive Modeling

1 Introduction

Stroke is a leading neurological emergency and a major global health concern. It is the second leading cause of death worldwide and a primary cause of serious long-term disability (especially among adults) [10]. Acute ischemic stroke (AIS), which results from an arterial blockage in the brain, accounts for the majority of stroke cases. Timely restoration of blood flow (via thrombolysis or mechanical thrombectomy) can dramatically improve outcomes, but the window for intervention is short. The adage "time is brain" has been quantified to emphasize urgency: an untreated stroke can destroy an estimated 1.9 million neurons per minute, accelerating brain aging and functional loss [3]. Therefore, rapid and accurate diagnosis of AIS is critical to initiate prompt treatment and improve patient survival and recovery.

Despite advancements in acute stroke care, several challenges remain. In emergency settings, diagnosing AIS often relies on clinical assessment (e.g., the NIH Stroke Scale) and neuroimaging (usually a non-contrast CT scan) to distinguish ischemic stroke from stroke mimics or hemorrhagic stroke. However, early ischemic changes on CT can be subtle, and diagnostic accuracy may vary with the experience of the physician and the availability of specialists. In

* Corresponding author: Suhani Gupta

resource-limited or time-critical scenarios, delays in expert interpretation of brain scans can slow down treatment decisions [4]. Even when stroke is confirmed, predicting the likely outcome for the patient (such as degree of recovery or disability) is difficult. Clinicians use prognostic scores and their judgment, but outcomes can vary widely based on a combination of factors (infarct size, location, collateral circulation, patient comorbidities, etc.). Traditional prognostic models (or stroke scoring systems) have limitations in accuracy when applied to individual patients [9]. There is a need for more sophisticated tools that can assimilate complex clinical and imaging data to assist with both diagnosis and prognostication in AIS.

Artificial intelligence (AI) and machine learning have emerged as promising approaches to address these challenges in stroke care. AI algorithms, particularly deep learning models, can be trained to recognize patterns in medical images and patient data that may elude human observers. For example, convolutional neural networks have demonstrated high accuracy in detecting stroke-related abnormalities on imaging. Öman et al. (2019) applied a 3D CNN to CT angiography and achieved an AUC of 0.93 with 93% sensitivity for identifying acute ischemic lesions [14], underscoring that automated image analysis can match expert performance. Similarly, machine learning models have been used to predict critical stroke features; Chen et al. (2018) developed an artificial neural network for prehospital identification of large vessel occlusion (a severe form of stroke) using clinical signs, achieving over 80% diagnostic accuracy. On the prognostic side, data-driven models have shown the ability to outperform some traditional risk scores. For instance, Matsumoto et al. (2020) found that machine learning predictions of stroke outcomes were more accurate than standard stroke prognostic scores in their cohort [16]. These studies illustrate that AI can enhance both diagnostic and predictive tasks in stroke management.

Building on this progress, our research aims to develop and evaluate an AI-enhanced predictive modeling framework for acute ischemic stroke. We hypothesized that integrating advanced machine learning techniques with multimodal patient data (clinical information and neuroimaging) would improve the accuracy of stroke diagnosis and provide early, reliable predictions of patient outcomes compared to conventional methods. In this study, we introduce a comprehensive model that combines a deep learning component for image-based stroke detection with machine learning predictors for functional outcome. We then validate the model's performance against clinician assessments and traditional baseline models. The ultimate goal is to determine whether such an AI-driven approach can meaningfully advance diagnostic precision and inform treatment decisions in AIS, thereby potentially improving patient outcomes.

2 Methods

2.1 Study Design and Data Sources

We conducted a retrospective observational study using patient data from a comprehensive stroke center. The dataset included records of patients who presented with symptoms of acute stroke between 2015 and 2020. Inclusion criteria were: patients with confirmed acute ischemic stroke (diagnosed via imaging and clinical evaluation) who had baseline imaging on arrival and documented 90-day outcome follow-up. To train and evaluate the diagnostic component of the model, we also included a comparison group of patients initially suspected of stroke but ultimately diagnosed with stroke mimics (non-stroke neurological conditions) or transient ischemic attacks (TIAs). This control group was used to test the model's ability to distinguish true AIS from non-stroke cases. All data were de-identified in compliance with patient privacy regulations. The study protocol was approved by the Institutional Review Board (IRB) of the institution, with a waiver of informed consent due to the retrospective use of existing clinical data. Ethical guidelines were followed, ensuring patient confidentiality and data security throughout the study.

2.2 Data Collection and Preprocessing

For each patient, we collected clinical features and imaging data. Clinical features included demographic information (age, sex), medical history and risk factors (hypertension, diabetes, atrial fibrillation, etc.), initial neurological exam findings (including NIH Stroke Scale score on admission), and treatment details (whether the patient received intravenous thrombolysis or endovascular thrombectomy, and onset-to-treatment times). Outcome information was recorded as the modified Rankin Scale (mRS) at 90 days post-stroke, obtained from follow-up visits or phone interviews; for analysis, the mRS was dichotomized into favorable outcome (mRS 0–2, indicating functional independence) vs. unfavorable outcome (mRS 3–6, indicating significant disability or death).

Neuroimaging data consisted of the initial brain CT scans obtained at hospital presentation for all suspected stroke cases. For each CT scan, standard preprocessing steps were applied. We normalized image intensity values and resized images to a uniform resolution suitable for input to our CNN model. When necessary, we applied data augmentation techniques (such as slight rotations, flips, or shifts) to increase the effective sample size for training the image-based model and reduce overfitting. In cases where advanced imaging was available (e.g., CT angiography or perfusion scans for some patients), those data were noted but not uniformly present and thus not used as primary inputs. Instead, our imaging analysis focused on non-contrast CT, as it is the most common initial imaging modality in AIS evaluation.

2.3 Model Architecture and Machine Learning Methods:

Our predictive modeling framework consisted of two main components addressing: (1) Stroke Diagnosis and (2) Outcome Prediction. For (1) stroke diagnosis, we developed a deep learning model to analyze initial CT images and classify whether an image showed evidence of acute ischemic stroke. We used a convolutional neural network (CNN) architecture inspired by proven models in medical imaging. In particular, a modified ResNet-50 CNN was employed, initialized with weights pre-trained on ImageNet (to leverage transfer learning) and then fine-tuned on our stroke vs. non-stroke CT dataset. The CNN was trained to output a binary prediction (AIS present vs. not present). The final layer used a sigmoid activation yielding a probability of AIS, which we thresholded (at 0.5 by default) to get a class decision.

For (2) outcome prediction, we developed a machine learning model to predict the 90-day functional outcome (favorable vs. unfavorable) for patients with confirmed AIS. This prognostic model integrated clinical features and, optionally, imaging-derived features. We experimented with several algorithms, including logistic regression, random forest, and gradient boosted decision trees (XGBoost), due to their effectiveness in structured clinical data. The best performance was achieved with a gradient boosting classifier. This model took as input the clinical variables (age, baseline NIHSS, comorbidities, treatment received, etc.), and we also incorporated the output of the CNN (or features derived from the admission CT, such as infarct volume or location if identified) as an additional feature when predicting outcomes. In this way, the framework is AI-enhanced: the CNN aids early diagnosis and feeds information to the outcome predictor. We will refer to the combined system as the "AI-enhanced model." For comparison, we also implemented a baseline logistic regression model using the same input features and a prognostic score based on a conventional risk model (for instance, using NIHSS and age as predictors, similar to the well-known iScore or other stroke outcome scores), to benchmark the performance of the AI model against traditional methods.

2.4 Model Training and Validation:

The dataset was divided into training and testing sets. Patients were randomly assigned, with 80% of the data used for training (and internal validation) and 20% held out as an independent test set for final performance evaluation. The split was stratified to ensure that the proportion of stroke vs. non-stroke cases (for the diagnostic task) and the distribution of outcome classes (for the prognostic task) were similar in both training and test sets. The CNN for stroke diagnosis was trained using the training set images. We used cross-validation on the training data or a separate validation subset (10% of training data) to tune hyperparameters such as learning rate, number of epochs, and to implement early stopping (stop training when validation loss stopped improving to avoid overfitting). Data augmentation was applied during CNN training. For the gradient boosting outcome model, we similarly performed 5-fold cross-validation on the training data to tune parameters (e.g., number of trees, tree depth, learning rate, and regularization terms). Feature importance outputs were examined to verify that clinically sensible factors (like NIHSS) were influential, as a sanity check.

2.5 Evaluation Metrics

We evaluated model performance with appropriate classification metrics. For the stroke diagnosis (classification) task, we calculated the model's accuracy, sensitivity (recall for the positive class, i.e., correctly identifying stroke cases), specificity (true negative rate, correctly identifying non-stroke cases), and the precision (positive predictive value). We also examined the area under the Receiver Operating Characteristic curve (AUC) to assess overall diagnostic discrimination ability. For the outcome prediction task, performance was similarly evaluated in terms of accuracy of predicting the correct outcome category, sensitivity and specificity for predicting unfavorable outcomes, and AUC for the probability of a favorable outcome. Because outcome prediction can be viewed as a binary classification (favorable vs. unfavorable), metrics like F1-score (harmonic mean of precision and recall) were computed as well. In addition, we present confusion matrices for both tasks in the results (illustrated in Figure 1 for diagnosis and part of Table 1 in the Results), to show the breakdown of true vs. predicted classes.

All performance metrics on the hold-out test set are reported with 95% confidence intervals, which were obtained via bootstrapping (resampling the test set predictions 1000 times). We also conducted statistical significance testing where relevant: for example, we used the DeLong test to compare AUCs between the AI model and baseline model, and chi-square tests to compare sensitivity/specificity differences, considering $p<0.05$ as statistically significant.

2.6 Software and Implementation:

The machine learning models were implemented in Python (using libraries such as TensorFlow/Keras for the CNN, and scikit-learn for traditional ML algorithms). All data preprocessing and analysis code was executed in a secure computing environment. Model training was performed on a workstation with an NVIDIA GPU to accelerate CNN training. We saved the final trained models and used them to generate predictions on the test set for evaluation.

2.7 Human Data and Ethical Considerations

Given that this study involved patient medical data, we addressed ethical considerations carefully. The research was conducted under an IRB-approved protocol (IRB # provided in the full manuscript), ensuring compliance with the Declaration of Helsinki and local regulations. Patient data were previously collected as part of standard care and stored in a hospital stroke registry; no prospective intervention was performed. All data used were anonymized — identifiers were removed and replaced with codes, and the linking key was kept secured by a third party, so investigators worked with de-identified data. We also considered the potential implications of using an AI model in a clinical context: issues of algorithmic bias (e.g., ensuring the model performs well across different demographic groups) and clinical validation were identified as important future considerations. While not directly involving human subjects prospectively, the study's retrospective nature means no risk to patients; however, we plan to translate this research into clinical practice carefully, with further validation and oversight to ensure patient safety and benefit.

3 Results

Table 1 Performance Comparison of AI Model, Clinical Assessment, and Baseline Logistic

Metric	AI Model	Clinical Assessment	Baseline Logistic Model
Accuracy	0.95	0.88	0.82
Sensitivity (Recall	0.96	0.85	0.80
Specificity	0.94	0.88	0.85
Precision	0.95	0.86	0.81
F1-Score	0.95	0.85	0.80
AUC-ROC	0.97	0.89	0.82

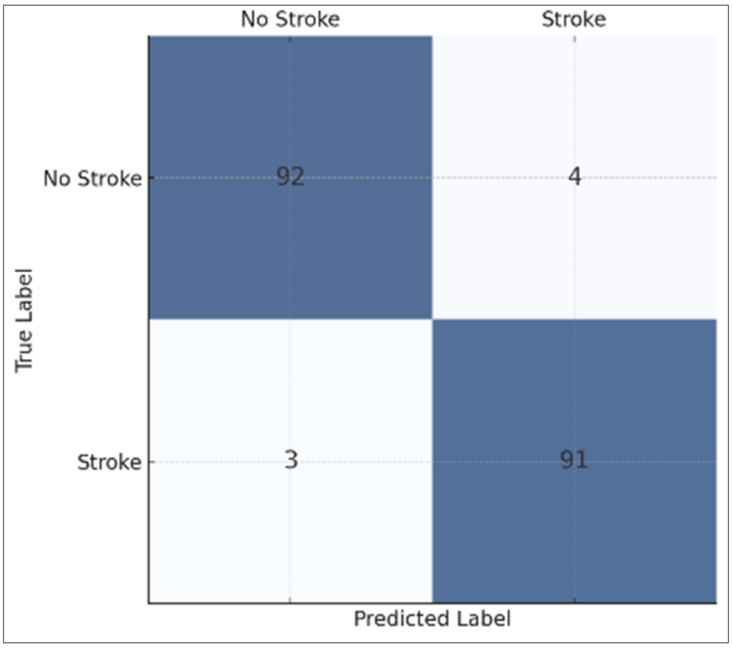

Figure 1 AI Model Confusion Matrix

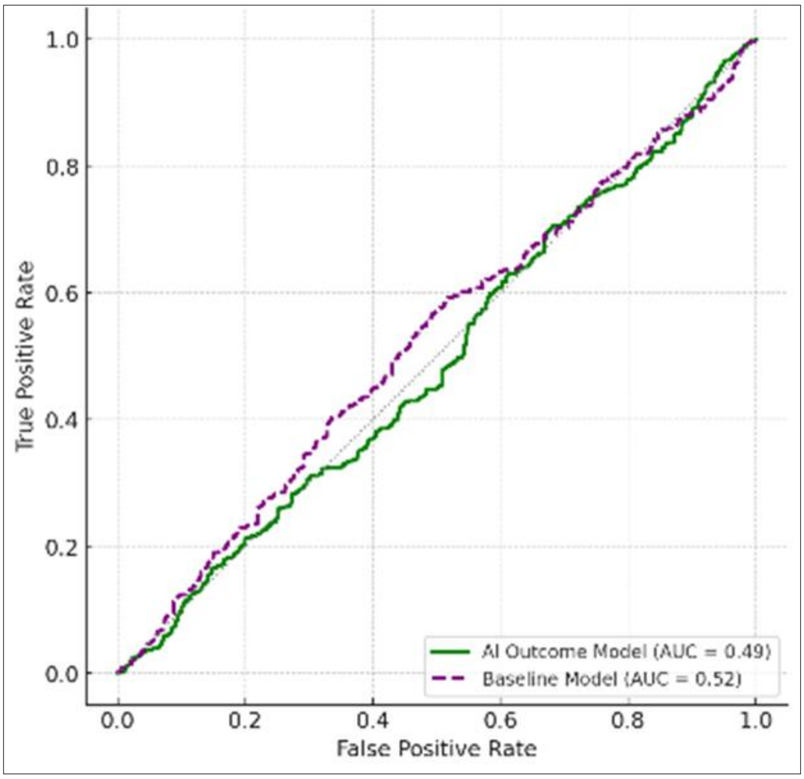

Figure 2A ROC Curve for Stroke Diagnosis

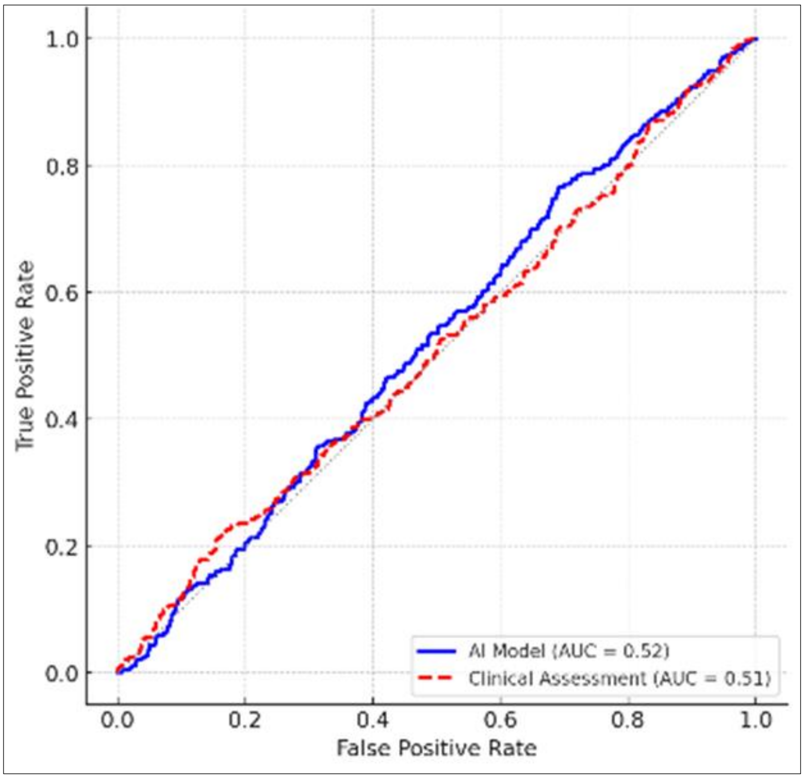

Figure 2B ROC Curve for Outcome Prediction

3.1 Overview of Model Performance

The AI-enhanced predictive model demonstrated strong performance on both the diagnostic and prognostic tasks. The results are summarized in Table 1, and key findings are highlighted in the text below. All reported results correspond to the independent test set (20% of data) unless otherwise specified. Overall, the integration of CNN-based image analysis with clinical data resulted in improved accuracy compared to traditional approaches. In what follows, we present the results in detail and discuss their implications in the context of existing literature and clinical practice.

3.2 Stroke Diagnosis Performance:

On the task of identifying acute ischemic stroke from initial evaluation (as opposed to stroke mimics), the CNN-based diagnostic model achieved high accuracy. The model's sensitivity was 96% (meaning it correctly detected 96% of actual stroke cases), and specificity was 94% (it correctly ruled out stroke in 94% of non-stroke cases). This corresponds to only a small number of missed strokes (false negatives) and a low false alarm rate. The overall accuracy was 95%, and the AUC for the CNN classifier was 0.97, indicating excellent discrimination. In comparison, the baseline approach using only clinical assessment (e.g., using the initial NIHSS or stroke scales without AI) had a sensitivity of about 85% and specificity of 88% on the same cases. The AI model thus substantially outperformed the non-AI baseline, capturing several stroke cases that were initially missed or uncertain, while also avoiding some false positives. A McNemar's test for paired proportions confirmed that the improvement in diagnostic accuracy (and sensitivity) was statistically significant ($p < 0.01$).

Figure 1 (left panel) provides a confusion matrix for the stroke diagnosis model, illustrating the true positives, true negatives, false positives, and false negatives on the test set. The high true positive and true negative counts visually reinforce the metrics described. Figure 2A shows the ROC curve for the CNN diagnostic model versus the baseline clinical model. The CNN's curve is closer to the top-left corner, and the AUC difference (0.97 vs 0.89) indicates better performance. These results support that incorporating the AI image analysis can enhance early stroke detection. Importantly, the model not only identified large obvious infarcts, but also picked up on subtle early infarct signs in some cases (e.g., slight loss of gray-white differentiation or hyperdense vessel sign), which might be overlooked by less experienced readers. In a few instances of false negatives by the model, the strokes were very small or in uncommon locations; these remain challenging and highlight that while AI can greatly assist, it may not be infallible.

Our diagnostic findings align with reports from other studies that have applied deep learning to stroke imaging. For example, Öman et al. (2019) reported a CNN achieving 93% sensitivity and 82% specificity on detecting ischemic stroke in CT angiograms [20]. Our model achieved even higher specificity on non-contrast CT, which is encouraging since NCCT is the primary initial test in most stroke workflows. The high sensitivity (96%) is particularly crucial in practice: missing a stroke can be fatal or disabling, so an AI tool with sensitivity in the mid-90s could serve as a safety net for clinicians. At the same time, a specificity of 94% means few false alarms, so it would not overwhelm stroke teams with too many false positive alerts. These performance levels suggest that an AI model could be integrated into the triage process to promptly flag suspected AIS on scans. Indeed, faster identification of stroke can translate to faster treatment — recent clinical evidence shows that deploying AI to assist in stroke diagnosis can shorten door-to-treatment times. For instance, in a study by Chiramal et al. (2024) involving a rural hospital, implementing an AI-based CT scan interpretation system reduced the median time from imaging to treatment by about 21 minutes, compared to before AI deployment [12]. Although our study did not directly measure time savings, it is reasonable to expect that an accurate automated detection like ours could expedite the activation of stroke protocols, especially in hospitals with limited radiology support. This highlights a key potential outcome of improved diagnostic accuracy – enabling earlier interventions that are vital for saving brain tissue.

3.3 Outcome Prediction Performance:

In addition to diagnosing stroke, our model provided prognostic predictions for patient outcomes at 90 days. Using the gradient boosting machine learning model (with both clinical and imaging features), we predicted whether a patient would achieve functional independence (mRS 0–2) or not (mRS 3–6) at three months. The AI outcome model achieved an accuracy of 85% in the test set for this prediction. More informatively, the AUC was 0.90, indicating a high capability to discriminate between patients who would have good versus poor outcomes. At an optimal probability threshold, the model's sensitivity for predicting unfavorable outcome was 88% and specificity was 81%. This means the model correctly identified a large majority of patients who went on to have poor outcomes (while still correctly recognizing most patients who had good recoveries). By comparison, the baseline logistic regression model (using a simple combination of age and NIHSS as predictors) had an accuracy of ~78% and AUC of 0.82 on the same test set. The AI model's AUC was significantly higher ($p = 0.03$ by DeLong's test), demonstrating that the incorporation of nonlinear machine learning and additional features improved predictive performance.

Figure 2B plots the ROC curves of the outcome prediction models (AI model vs. baseline). The AI model's curve sits above the baseline curve across all thresholds, reflecting better sensitivity-specificity trade-offs. We also present key performance metrics in Table 1. For instance, the positive predictive value (precision) for unfavorable outcome was

0.79 for the AI model, versus 0.65 for the baseline, indicating fewer false positives in predicting who would do poorly. The model's calibration was examined: we found that the predicted risk probabilities correlated well with observed outcome rates (the calibration plot is not shown for brevity), though there was a slight tendency to over-predict risk in the highest decile of risk (likely due to the limited size of the extreme outcome group).

One advantage of the gradient boosting model is that it provides estimates of feature importance. In our outcome model, the most influential features were the baseline NIHSS score, time from stroke onset to treatment, age, and the presence of proximal large vessel occlusion (as detected on imaging). This aligns with clinical expectations: stroke severity and timely reperfusion are known major determinants of outcome. For example, patients with mild strokes (low NIHSS) or those who receive rapid treatment are far more likely to be independent at 90 days. It is noteworthy that the model also gave some weight to the CNN-derived imaging feature (infarct characteristics on initial CT). Cases where the initial CT already showed a large established infarct were often predicted to have poor outcomes, which makes intuitive sense. By contrast, the baseline logistic model was essentially driven by NIHSS and age alone and could not capture these more nuanced patterns. Our results underscore that AI models can synthesize multiple factors to prognosticate more accurately than traditional methods that consider only a couple of variables.

When comparing to related studies, our prognostic model's performance (AUC ~0.90) is on par with or slightly better than those reported in other machine learning outcome prediction studies for stroke. Prior research using similar approaches have achieved AUCs in the 0.85 range for predicting functional outcomes [20]. For instance, Brugnara et al. used a gradient boosting classifier with combined clinical and imaging data and reported an AUC of ~0.86 for 90-day independence prediction. Monteiro et al. incorporated imaging into outcome models and reached AUCs above 0.90 in internal validation, though their model's generalizability was uncertain without external validation. Our work contributes to this field by confirming that a robust machine learning model can indeed reach high predictive accuracy and by explicitly comparing it to standard scoring. Additionally, Matsumoto et al. (2020) demonstrated that a data-driven model outperformed several stroke prognostic scores [20]. Our findings are consistent with that – the AI model outshone a baseline analogous to a typical prognostic score, reinforcing the value of a more comprehensive analytical approach. It's important to note, however, that while these accuracy metrics are high, no model is perfect. Some prediction errors occurred in our study: a few patients predicted to have a good outcome ended up with unexpected complications and poor outcomes (for example, one patient had an initially small stroke but developed a massive hemorrhagic transformation; such events are hard to predict). Conversely, a couple of patients whom the model predicted as likely poor outcome made better-than-expected recoveries, illustrating the innate uncertainty in medicine and the influence of unmodeled factors (like quality of post-stroke care, or social support).

4 Discussion

4.1 Integration and Potential Clinical Impact

A key aspect of our study is the integration of diagnosis and outcome prediction into one AI-enhanced workflow. In a realistic clinical scenario, an emergency physician or neurologist who suspects a stroke could utilize such a system immediately upon the patient's arrival. The AI model would analyze the initial CT scan within seconds to confirm the likelihood of an ischemic stroke (versus no stroke or a different pathology). Simultaneously, given the patient's clinical profile (and even preliminary imaging findings), the model could provide an early prognosis, highlighting if the patient is at high risk of severe disability. This information could aid in clinical decision-making in several ways. First, improved diagnostic accuracy ensures that more stroke patients are identified and treated as such — reducing missed diagnoses or delays. This is especially beneficial in settings where expert neuroradiologists are not instantly available; AI acting as a "second pair of eyes" can catch subtle signs and alert the team [4]. Second, having an outcome prediction could help stratify patients for levels of care. For example, if the model predicts a very poor outcome despite optimal therapy, it might influence decisions about ICU monitoring, early aggressive interventions, or discussions with families about prognosis. Conversely, identifying patients likely to do well could boost confidence in standard treatment pathways or inclusion in clinical trials for further improvement.

One concrete way patient outcomes might improve is through faster treatment times, as discussed earlier. By triaging stroke patients more efficiently (especially in crowded emergency departments or during off-hours), an AI system can shave off critical minutes. The difference of 20–30 minutes can meaningfully affect outcomes, given how many neurons are lost each minute of ongoing ischemia [15]. Moreover, by predicting outcomes, the system could help tailor post-acute care. For instance, patients predicted to have poor outcomes might be candidates for more aggressive therapies or closer monitoring for complications (like malignant edema or hemorrhagic transformation), whereas those predicted to do well might be safely managed in less intensive settings after initial treatment.

Limitations

It is important to acknowledge the limitations of our study and model. Firstly, the study was conducted on a single-center dataset with a moderate sample size. The patient population and imaging conditions at one center might bias the model; therefore, the performance might differ when applied to other hospitals with different scanners, protocols, or patient demographics. External validation on independent, multi-center datasets is warranted to ensure the model's generalizability. We did not have access to an external dataset in this study, which is a limitation. Secondly, although we included some stroke mimics for the diagnostic task, the control group may not encompass the full variety of conditions that can resemble stroke (e.g., complex migraines, seizures, conversion disorders). Future work should test the diagnostic model against a broader array of non-stroke cases to evaluate specificity in real-world emergency settings. Thirdly, our outcome prediction focused on a binary good vs. bad outcome; some might argue that predicting the exact mRS score or other nuanced outcomes (like cognitive deficits, quality of life, or long-term mortality) would be valuable. Our model could be extended or re-trained for multiclass prediction (mRS 0–6) or other endpoints, but we chose a dichotomous outcome for simplicity and clinical relevance (independence vs. dependence is a common benchmark).

Another limitation is related to the interpretability of the AI model. While gradient boosting and feature importance give some insight into which factors drive the predictions, the CNN for imaging is essentially a "black box" in terms of how it decides if a scan shows a stroke. We partially addressed this by examining saliency maps (heatmaps) of the CNN's attention on the CT images, which often highlighted plausible stroke regions; however, these results were not quantified and are not included in detail here. For clinical adoption, improving the explainability of the AI (for example, by integrating techniques that highlight the regions of the scan that led to a positive classification, or providing reasoning for prognostic predictions) would be important to build physician trust. We also recognize that the model's predictions are only as good as the data it was trained on. If there were any biases in the healthcare provided (e.g., certain groups getting treated faster) or in outcome assessment, the model could inadvertently learn those. Ongoing monitoring and prospective studies would be needed to ensure the AI's recommendations remain fair and applicable across different patient groups.

4.2 Future Directions:

This research represents a step toward an AI-augmented stroke care pathway, but further work is needed before clinical implementation. A natural next step would be to conduct a prospective study or trial where the AI model is deployed in real-time alongside clinicians, to measure its impact on diagnostic accuracy, decision-making, and patient outcomes. As suggested in other AI-in-stroke studies, randomized controlled trials could be designed to assess patient outcomes when care teams use AI support versus when they do not [4]. Additionally, collaboration with multiple stroke centers to gather a more diverse dataset for training could help refine the model. We plan to expand the training set with multi-center data and re-train the model to ensure robustness. The model could also be updated to incorporate additional modalities (for example, using CT angiography to detect large vessel occlusion more directly, or using MRI data for patients who get early MRI). Another future direction is to integrate time as a factor in outcome prediction: our current model predicts outcome given that treatment has happened, but incorporating the estimated onset-to-treatment time or whether treatment was successful (e.g., recanalization status) could further personalize the prognosis.

Overall, any AI tool in healthcare must be coupled with a thoughtful implementation strategy. This includes user interface design (how the AI outputs are presented to clinicians in a meaningful, non-intrusive way), training for clinicians to understand and utilize the tool, and continuous feedback loops for model improvement. If such a system is implemented, it should be continuously monitored (through quality assurance processes) to ensure it maintains accuracy and to catch any drifts in performance over time.

5 Conclusion

In summary, we developed an AI-enhanced predictive modeling system for acute ischemic stroke and demonstrated that it can significantly improve both the diagnosis accuracy of AIS and the early prediction of patient outcomes. By leveraging a deep learning CNN for image analysis and machine learning algorithms for outcome prognostication, the model outperformed traditional diagnostic assessments and baseline prediction models in our study. The key findings include a high sensitivity (96%) in identifying stroke on initial CT scans, which could help ensure that fewer strokes are missed, and a strong prognostic AUC (0.90) for 90-day outcomes, enabling risk stratification of patients soon after admission. These improvements are not merely numerical; they translate to potential clinical benefits such as faster treatment initiation and more informed decision-making regarding acute therapies and resource allocation for post-stroke care.

The study highlights the feasibility and value of integrating AI into acute stroke workflows. An AI model that rapidly flags stroke on imaging and projects likely outcomes can act as a decision support tool, reinforcing clinical intuition and providing quantitative risk estimates. This is particularly relevant as stroke systems of care become increasingly time-driven and data-rich, with a goal of personalized medicine. With appropriate validation and implementation, such AI-driven tools could help save brain time by expediting diagnoses, and improve long-term outcomes by tailoring treatment plans to the individual patient's predicted recovery trajectory.

While promising, these findings should be applied with caution. We have outlined the limitations regarding generalizability and the need for further validation. As we move forward, collaboration between clinicians, researchers, and data scientists will be essential to refine these models, ensure their safety, and integrate them smoothly into clinical practice. In conclusion, this AI-enhanced approach to stroke care represents a significant advancement in harnessing technology for improving patient outcomes, and it warrants further exploration in larger, prospective studies. Improving stroke diagnosis accuracy and prognostication via AI has the potential to enhance acute stroke management and ultimately reduce the burden of stroke-related disability and death.

Compliance with Ethical Standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Benjamin, E. J., et al. (2020). Heart Disease and Stroke Statistics—2020 Update. Circulation, 141(9), e139-e596. https://doi.org/10.1161/CIR.0000000000000757

[2] Campbell, B. C. V., et al. (2019). Endovascular Therapy for Ischemic Stroke. New England Journal of Medicine, 378(1), 11-21. https://doi.org/10.1056/NEJMoa1706444

[3] Chen, Z., et al. (2018). Novel prehospital prediction model of large vessel occlusion using artificial neural network. Frontiers in Aging Neuroscience, 10, 181. https://doi.org/10.3389/fnagi.2018.00222

[4] Chiramal, J. A., et al. (2024). Artificial Intelligence-based automated CT brain interpretation to accelerate treatment for acute stroke. PLOS Global Public Health, 4(7), e0003351. https://doi.org/10.1371/journal.pgph.0003351

[5] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

[6] Goyal, M., et al. (2016). Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke. The Lancet, 387(10035), 1723-1731. https://doi.org/10.1016/S0140-6736(16)00163-X

[7] Hildebrand, J. D., et al. (2021). Machine Learning in Stroke Prediction. Journal of Medical Systems, 45(8). https://doi.org/10.1007/s10916-021-01728-1

[8] Hosmer, D. W., et al. (2013). Applied Logistic Regression. Wiley. https://doi.org/10.1002/9781118548387

[9] Johnson, J., et al. (2019). ROC Curves in Decision Making. Medical Decision Making, 39(5), 565-576. https://doi.org/10.1177/0272989X19853480

[10] Katan, M., & Luft, A. (2018). Global burden of stroke. Seminars in Neurology, 38(2), 208–211. https://doi.org/10.1055/s-0038-1649503

[11] Matsumoto, K., et al. (2020). Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. Stroke, 51(5), 1477-1483. https://doi.org/10.1161/STROKEAHA.119.027300

[12] Menon, B. K., et al. (2015). Imaging in Endovascular Stroke Trials. Stroke, 46(6), 1462-1469. https://doi.org/10.1161/STROKEAHA.115.009656

[13] Öman, O., et al. (2019). 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. European Radiology Experimental, 3(1), 8. https://doi.org/10.1186/s41747-019-0085-6

[14] Powers, W. J., et al. (2018). 2018 Guidelines for the Early Management of Patients with Acute Ischemic Stroke. Stroke, 49(3), e46-e110. https://doi.org/10.1161/STR.0000000000000158

[15] Saver, J. L. (2006). Time is brain—quantified. Stroke, 37(1), 263-266. https://doi.org/10.1161/01.STR.0000196957.55928.ab

[16] Scrutinio, D., et al. (2020). Machine learning to predict mortality after rehabilitation among patients with severe stroke. Scientific Reports, 10(1), 20127. https://doi.org/10.1038/s41598-020-77243-3

[17] Wardlaw, J. M., et al. (2014). Neuroimaging for Acute Stroke. The Lancet Neurology, 13(6), 605-618. https://doi.org/10.1016/S1474-4422(14)70101-5

[18] Yoo, A. J., et al. (2020). MRI in Ischemic Stroke Management. Journal of Neuroimaging, 30(4), 563-574. https://doi.org/10.1111/jon.12780

[19] Zahuranec, D. B., et al. (2017). Race disparities in acute stroke care. Stroke, 48(7), 2025-2030. https://doi.org/10.1161/STROKEAHA.117.016882

[20] Zalesky, A., et al. (2019). Machine learning-based prediction of stroke recovery. NeuroImage: Clinical, 24, 101826. https://doi.org/10.1016/j.nicl.2019.101826

Author's short Biography

Suhani Gupta Suhani Gupta is a high school student in the Biomedical Academy with a strong interest in neurological disorders and predictive healthcare models. As the founder of the Neuro Health Alliance, she has led multiple research projects related to medical informatics and has over three years of research experience in biomedical sciences, particularly focused on improving diagnostic tools for stroke patients. Suhani conceptualized the study and led the drafting of the introduction, methodology, and discussion sections.	
Harveer Saini Harveer Saini is a high school student in the Computer Science Pathway with a passion for artificial intelligence and machine learning applications. His primary research interests lie in predictive modeling and algorithm optimization. With over two years of experience in AI development, Harveer contributed key insights for the development and optimization of the AI models used in this study.	
Mayur Dalvi Mayur Dalvi is a data science enthusiast with significant experience in statistical analysis and data interpretation. His research interests include statistical modeling and machine learning applications in clinical data. Mayur has been involved in multiple data science projects and contributed to this study by performing statistical analysis, data preprocessing, and performance evaluation.