Journal of Clinical and Nursing Research, 2026, Volume 10, Issue 1
http://ojs.bbwpublisher.com/index.php/JCNR
Online ISSN: 2208-3693
Print ISSN: 2208-3685

Composite Deep-Learning Model for 90-Day mRS Prediction in Post-Stroke Patients

Shihan Dong, Zhengwei Yao, Yuhang Chuai, Ran Li, Handong Zhang

Shanghai University of Engineering Science, Shanghai 201620, China

Abstract: To counteract small sample size, severe class imbalance and high feature redundancy in 90-day mRS prediction after stroke, this study proposes a four-stage pipeline— "ADASYN re-sampling → clinical + statistical feature screening → dimensionality reduction → 5-fold cross-validation" —and benchmark composite deep-learning architectures. ADASYN first balances the minority classes in the original feature space. Next, a tri-level filter (clinical domain knowledge, variance threshold, mutual information) removes clinically meaningless or redundant variables, after which PCA compresses the remaining features while preserving critical neurological signatures (e.g., brain-herniation history). Four hybrid CNN–RNN models are trained and compared under strict 5-fold cross-validation; the optimal ensemble yields stable, clinically interpretable probabilities that can support individualized rehabilitation planning.

Keywords: Stroke; 90-day mRS; Composite deep learning; ADASYN; 5-fold cross-validation

Online publication: Feb 11, 2026

1. Introduction

Spontaneous intracerebral hemorrhage (ICH), a non-traumatic rupture of intraparenchymal vessels, accounts for 10–15% of all stroke admissions. Its abrupt onset, rapid progression and poor prognosis produce an acute-phase case-fatality rate of 45–50%, and roughly 80% of survivors suffer permanent neurological disability, imposing heavy medical and socio-economic burdens [1]. The 90-day modified Rankin Scale (mRS), an ordinal outcome ranging from 0 (no symptoms) to 6 (death), is the standard metric for post-stroke functional independence and rehabilitation potential. Accurate pre-discharge prediction of this score can guide early, individualized neuro-rehabilitation programs and rational allocation of intensive-care resources, ultimately reducing long-term disability [2].

Nevertheless, data-driven prediction of 90-day mRS in ICH patients faces three practical obstacles:

(1) Small sample size

High-quality datasets must integrate personal history, vascular risk factors, treatment protocols and serial neuro-imaging, all of which are costly to collect.

(2) Class imbalance

Favorable outcomes (mRS 0–2) predominate, whereas severe disability (mRS 5) and death (mRS 6) are rare, biasing classifiers toward the majority.

(3) High-dimensional redundancy

Predictive variables span demographic covariates, laboratory values, and hundreds of radiomic descriptors (volume, location, shape, texture) that often contain mutual information.

Recent advances in medical artificial intelligence have introduced machine-learning (ML) and deep-learning (DL) solutions. Machine learning can effectively handle non-linear relationships and interactions between complex features, while also being adaptable to high-dimensional and large-scale datasets [3]. Some researchers have combined modern medical diagnosis with TCM syndrome elements as predictive variables to identify key factors influencing the risk of stroke, thereby improving the accuracy of predictive models [4,5]. Cetinoglu et al. developed a convolutional neural network-based deep learning model for vascular classification and stroke detection [6]. Liu T et al. established a hybrid machine learning model based on deep neural networks to predict the onset of stroke, targeting imbalanced stroke datasets [7]. Zhu et al. employed random-forest regression, gradient-boosting machines and multilayer perceptrons to quantify ICH prognosis [8]. Mao and colleagues reviewed DL-based hemorrhage segmentation and detection, highlighting bottlenecks and remedies [9]. Wang et al. demonstrated that ML models outperform traditional regression on RMSE and MAE when both clinical and imaging features are leveraged [10]. Meng focused on ischemic stroke, offering cross-aisle insights for cerebrovascular diseases [11]. Leveraging outstanding temporal-dependency modeling, deep-learning-centered AI has become the mainstream tool for the re-analysis of medical sequential data [12]. The triad of SCLSTM, Sim-LSTM and CELSTM proposed by Liu Boda sequentially tackles auxiliary diagnosis, intervention recommendation and recovery prediction of post-stroke depression; experiments show diagnosis accuracy reaches 0.737, intervention recommendation 0.895, and recovery-prediction error drops to 4.770, surpassing existing benchmarks across the board [13]. Li Chenhao et al. presented a deep-learning fusion scheme to boost abnormal gait-recognition accuracy [14]. Zhang Yuyu et al. introduced an adaptive deep extreme learning machine that significantly raises minority-class classification accuracy [15].

Despite these efforts, the literature rarely explores composite DL, an ensemble of two or more deep architectures in ICH outcome prediction. Because ICH prognosis is simultaneously shaped by baseline clinical traits, spatial imaging patterns (hematoma shape, oedema location) and temporal dynamics (hematoma expansion, oedema progression), a single network may fail to capture all interactions. Composite models can merge complementary representational strengths, improving both discrimination and calibration.

Accordingly, this study applies composite DL to predict simplified 90-day mRS scores in 97 consecutive ICH patients. The original seven-level mRS was collapsed into four ordered categories to mitigate extreme imbalance. This study has designed a full-cycle pipeline, re-sampling, feature filtering, PCA denoising and 5-fold cross-validation, and compared four CNN–RNN hybrids that respectively specialize in spatial (CNN) and longitudinal (LSTM/GRU) modelling. The most stable ensemble was offered as a ready-to-use tool for precision neuro-rehabilitation.

2. Data processing

The dataset comprises complete clinical and imaging records of 97 stroke patients, totaling 210 variables. Clinical descriptors cover baseline demographics and medical history (age, sex, hypertension, diabetes, atrial fibrillation, previous stroke, coronary artery disease, smoking, alcohol intake, pre-ICH mRS, systolic/diastolic blood pressure, onset-to-first-imaging time, etc.). Imaging descriptors quantify the hematoma and peri-lesional tissue (total hematoma/oedema volume, region-of-interest proportions, 3-D shape indices, grey-level distribution metrics, brain-structure ratios, etc.). The target is the 90-day mRS score (0–6). To accommodate the limited sample size and skewed class frequencies, the original seven-level mRS was collapsed into four ordered categories following established prognostic conventions: Category 0: mRS 0–1 (no or minimal disability), Category 1: mRS 2–3 (moderate disability), Category 2: mRS 4–5 (severe disability), Category 3: mRS 6 (death).

A rigorous preprocessing pipeline was then applied. Missing values were imputed with domain-appropriate statistics (median for continuous, mode for categorical). All numeric features were Z-score standardized to remove scale differences. The cohort was split into training and validation sets using clinically accepted ratios.

To counteract the scarcity of category 3 (mRS = 6) cases, Adaptive Synthetic Sampling (ADASYN) was performed in the original 210-dimensional feature space with k = 5 nearest neighbors, generating 21 synthetic minority samples and expanding the training set to 89 cases while preserving local manifold structure.

Next, a two-tier feature-selection strategy was executed:

(1) Clinical filter

15 core variables (e.g., age, GCS, total hematoma + oedema volume, onset-to-imaging time, systolic BP, hypertension history) were retained on the basis of published prognostic evidence.

(2) Data-driven filter

Imaging features with variance < 0.1 were discarded; mutual information (MI) with the collapsed mRS label was computed for the remainder, and features with MI > 0.05 were kept. The procedure reduced the dimensionality from 210 to 82 informative variables.

Finally, principal-component analysis was applied to the 82 selected features. The first 16 components, explaining ≥ 85% cumulative variance, were retained as the modelling input. This last compression step mitigates the curse of dimensionality while conserving the dominant data variance, providing a compact, denoised representation for the subsequent composite deep-learning training.

To fully align with the multi-dimensional nature of ICH-outcome prediction, structured clinical descriptors, spatially correlated imaging biomarkers, and longitudinal evolution across serial scans. This study designed four composite deep-learning architectures that share a two-branch strategy but differ in their algorithmic backbone.

Branch-1 is dedicated to static information: demographic variables and summarized imaging metrics (e.g., total volumes, shape indices) are vectorized and processed by a module chosen for tabular/spatial data (CNN or Transformer).

Branch-2 targets temporal dynamics: time-series of hematoma/oedema volumes, ventricular ratios, etc., acquired at ≤ 4 time-points, are fed to a sequence-specialized unit (LSTM, GRU or Transformer encoder).

The latent representations produced by the two branches (128-D each) were concatenated, yielding a 256-D fused embedding that is batch-normalized, dropout-regularized ($p = 0.3$) and fed to a dense classification head with Soft-max to output probabilities for the four collapsed mRS classes (0–3).

Table 1 summarizes the algorithmic assignment of every composite model; all share the same fusion logic but exploit different inductive biases to maximize information extraction from each data modality.

Table 1. Composite-model architectures

Model name	Sub-model A (role)	Sub-model B (role)
CNN + LSTM	CNN (2Conv + MaxPool, Local Features)	LSTM (units = 32, sequential dependencies)
CNN + GRU	CNN (2Conv + MaxPool, Local Features)	GRU (units = 32, light-weight temporal modelling)
LSTM + Transformer	LSTM (units = 32，base-level sequential patterns)	Transformer (4-head self-attention)
CNN + Transformer	CNN (2Conv + MaxPool, Local Features)	Transformer (4-head self-attention)

Training and validation hyper-parameters were tuned specifically for the 4-class imbalance setting:

(1) Training

The model was optimized using the Adam optimizer with a learning rate of 0.001. Categorical cross-entropy was used as the loss function. Training was performed with a batch size of 8 for a maximum of 50 epochs, with early stopping applied (patience = 10).

(2) Validation

Model selection was carried out using 5-fold stratified cross-validation within the training set, yielding approximately 54–55 training samples and 13–14 validation samples per fold. Final generalization performance was assessed on an independent hold-out test set.

(3) Evaluation metrics

Performance was evaluated using accuracy, macro-averaged F1 score, and recall for class 3 (mRS = 6). The latter was prioritized due to its clinical importance and substantial under-representation in the dataset.

3. Results

After selection the feature space shrank from 210 → 82 (60.95% redundancy removed); PCA retained 85% variance with only 16 components, raising the first-component contribution from 15.93% to 22.67%.

3.1. Model comparison

Internal CV shows CNN-LSTM is the most stable; details are given in **Table 2**.

Table 2. Performance comparison of all models

Model name	Mean accuracy (%)	Accuracy SD (%)	Mean macro-F1 (%)
CNN + LSTM	70.36	2.85	67.82
CNN + GRU	67.14	3.52	64.05
LSTM + Transformer	62.50	4.17	59.38
CNN + Transformer	65.71	3.83	62.16

3.2. Generalization comparison

To benchmark the robustness and generalization boundary of each composite architecture, this study has conducted a lateral evaluation on the fully independent test set; results are consolidated in **Table 3**.

Table 3. Performance on the independent test set

Model name	Accuracy	Macro-F1	Class-3 recall	Class-3 precision
CNN + LSTM	65.52	71.36	83.33	75.00
CNN + GRU	62.07	67.54	75.00	70.00
LSTM + Transformer	55.17	60.28	66.67	62.50
CNN + Transformer	58.62	63.81	70.00	65.00

CNN-LSTM achieved the best overall metrics:

Class 0 accuracy 72.22%, Class 1 accuracy 73.68%, with zero cross-class misclassification, confirming stable discrimination of mild disability.

Class 2 mis-classification rate 22.22%, predominantly drifting to Class 1, indicating residual confusion for moderate outcomes.

Class 3 (extreme disability) recall reached 83.33% and precision 75%, eliminating the chronic under-detection reported by earlier models and demonstrating the perceptual gain of CNN-LSTM on extreme cases.

4. Discussion

4.1. Value of the proposed pipeline

By inserting re-sampling before any dimensionality reduction, this study synthesize minority samples in the original high-dimensional feature space, preserving critical clinical signatures (e.g., brain-herniation history) that would otherwise be blurred by early compression.

Feature de-noising further removes imaging-derived variables devoid of clinical meaning (e.g., duplicate pixel counts), purifying the input pool and simultaneously increasing both information density and interpretability of the subsequent PCA components.

Closed-loop cross-validation keeps the entire hyper-parameter search and performance estimation inside the training folds, cutting off "random-fit" noise; the CNN-LSTM ensemble consequently exhibits a 2.85% accuracy SD, corroborating its stability and reproducibility.

4.2. Performance analysis

The superiority of CNN+LSTM stems from the synergy between "local-pattern extraction" and "sequential dependency modelling": CNN detects local interactions among features (e.g., joint effect of hematoma volume and age), while LSTM captures latent temporal evolution of the selected principal components (e.g., imaging indices tracking clinical deterioration). Transformer-based variants, limited to only six time-steps, fail to exploit the full power of self-attention.

4.3. Limitations & future work

The restricted sample size caps absolute accuracy. Enlarging the cohort or incorporating richer intra-hospital time-series (vital-sign streams) is expected to boost generalizability.

5. Conclusion

The proposed pipeline—ADASYN re-sampling → dual-criteria feature selection → PCA → stratified cross-validation, addresses both small-sample and class-imbalance challenges in post-stroke mRS prediction. The CNN + LSTM ensemble achieves 65.52% accuracy and 71.36% Macro-F1 on 97 patients, with an 83.33% recall for the most critical category, offering a practical aid for clinical rehabilitation assessment.

Funding

Shanghai University of Engineering Science Undergraduate Innovation Training Program (Project No.: cx2521005)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Gao L, 2016, Expert Consensus on Integrated Chinese and Western Medicine Diagnosis and Treatment of Hypertensive Intracerebral Hemorrhage in the Acute Phase. Chinese General Practice, 19(30): 3641–3648.

[2] Jiang B, Chen C, Tan H, et al., 2024, Correlation Analysis of Common Blood Biochemical Indicators with Neurological Function and Prognosis After Intravenous Alteplase Thrombolysis in Patients with Acute Cerebral Infarction. Chinese Journal of Geriatric Heart Brain and Vessel Diseases, 26(1): 63–66.

[3] Obermeyer Z, Emanuel E, 2016, Predicting the Future: Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, 375(13): 1216–1219.

[4] Wang L, 2023, Development and Validation of an Early Warning Model for New-Onset Stroke in Chinese Populations at High Risk of Stroke, thesis, China Academy of Chinese Medical Sciences.

[5] Sun Z, Ji J, Ma Z, et al., 2023, Construction and Application of Stroke TCM Pattern Differentiation Model Based on Machine Learning. Journal of Hunan University of Chinese Medicine, 43(4): 694–699.

[6] Cetinoglu Y, Koska I, Uluc M, et al., 2021, Detection and Vascular Territorial Classification of Stroke on Diffusion-Weighted MRI by Deep Learning. European Journal of Radiology, 145: 8.

[7] Liu T, Fan W, Wu C, 2019, A Hybrid Machine Learning Approach to Cerebral Stroke Prediction Based on Imbalanced Medical Dataset. Artificial Intelligence in Medicine, 101: 101723.

[8] Zhu A, Zhu M, Li T, et al., 2025, Prognosis Prediction and Key-Factor Exploration for Patients with Hemorrhagic Stroke. Journal of Qilu University of Technology, 39(2): 61–71.

[9] Mao T, Li Y, Li M, et al., 2025, Review on Deep-Learning-Based Lesion Segmentation and Detection in Acute Ischemic Stroke. Computer Systems and Applications, 34(1): 11–25.

[10] Wang H, Guo J, 2024, Establishment of a Machine-Learning-Based Clinical Intelligent Diagnosis and Prediction Model for Hemorrhagic Stroke. Heilongjiang Science, 15(10): 129–132.

[11] Meng Y, Xu J, Qiu K, et al., 2022, Application of Deep Learning in Ischemic Stroke. Zhejiang Clinical Medical Journal, 24(12): 1892–1894.

[12] Xu P, Fan C, Zhu R, et al., 2021, Prediction of Hospital Inpatients Based on Combined Prophet-LSTM-PSO Model. Journal of University of Shanghai for Science and Technology, 43(1): 68–72.

[13] Liu B, 2024, Assisted Clinical Decision-Making for Post-Stroke Depression Based on Long Short-Term Memory Networks and Its Implementation, thesis, University of Electronic Science and Technology of China.

[14] Li C, Yang P, Feng C, et al., 2025, Abnormal Gait Recognition in Stroke Patients Using Deep-Learning Fusion. Journal of Medical Biomechanics, 40(4): 955–962.

[15] Zhang Y, Li F, Wang W, et al., 2025, Deep Extreme Learning Machine Model for Imbalanced Datasets. Journal of Chongqing University of Technology (Natural Science), 39(6): 108–116.