Open Access Library Journal
2026, Volume 13, e14446
ISSN Online: 2333-9721
ISSN Print: 2333-9705

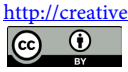
Deep Learning for Predicting Post-Stroke Cognitive Decline Using Multimodal Data: A Synthetic Proof-of-Concept Study

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) Deep Learning for Predicting Post-Stroke Cognitive Decline Using Multimodal Data: A Synthetic Proof-of-Concept Study. *Open Access Library Journal*, **13**: e14446.
https://doi.org/10.4236/oalib.1114446

Received: October 13, 2025
Accepted: January 12, 2026
Published: January 15, 2026



Abstract

Cognitive impairment is a frequent and debilitating outcome of stroke, profoundly affecting patient independence, recovery trajectories, and long-term quality of life. Despite its prevalence, accurate early prediction of post-stroke cognitive decline (PSCD) remains an unmet challenge due to the multifactorial interplay between structural brain lesions, neurophysiological changes, and diverse clinical comorbidities. In this study, we present a multimodal deep learning framework designed to classify stroke patients at risk of PSCD by jointly analysing synthetic magnetic resonance imaging (MRI), electroencephalography (EEG), and clinical features. The architecture comprises three modality-specific branches: a 3D convolutional neural network (CNN) to extract lesion-based structural patterns from MRI volumes, a 1D CNN for temporal feature extraction from EEG signals, and a multilayer perceptron for tabular clinical predictors. These representations are integrated through late fusion to enable comprehensive multimodal risk stratification. Experiments conducted on a balanced synthetic dataset (N = 800; MRI: 32 × 32 × 32, EEG: 8 × 256, Clinical: 12 features) demonstrated strong predictive performance, achieving a best validation area under the curve (AUC) of 0.7327, which outperformed all unimodal baselines. To enhance clinical interpretability, we employed permutation feature importance, gradient-based saliency mapping, and UMAP visualization of latent embeddings. Results highlighted clinically meaningful drivers of prediction, including baseline Montreal Cognitive Assessment (MoCA) scores, age, and lesion burden. Contribution analysis further confirmed that multimodal fusion consistently yielded superior predictive capacity compared to individual modalities. Although based on synthetic data, these findings establish a robust methodological foundation for multimodal AI-driven risk

stratification in stroke rehabilitation. Future work will extend validation to real-world multi-center clinical datasets, investigate time-to-event modelling, and explore integration into clinical decision-support systems to guide personalized interventions.

Subject Areas

Neuroscience, Medical Artificial Intelligence

Keywords

Stroke, Cognitive Decline, Deep Learning, Multimodal Fusion, MRI, EEG, Clinical Data, Explainable AI

1. Introduction

Stroke is a leading cause of disability worldwide and remains a major contributor to long-term cognitive and functional impairment [1]-[4]. Cognitive impairment following stroke is one of the most disabling sequelae, with estimates suggesting that 40% - 60% of survivors develop some degree of cognitive decline within the first year [5]-[8]. This significantly limits independence, hampers rehabilitation, and increases both caregiver burden and healthcare costs. Early identification of patients at high risk of post-stroke cognitive decline (PSCD) is therefore critical for optimizing rehabilitation strategies and improving long-term outcomes [9]-[10]. However, current prediction methods rely primarily on clinical scales such as the National Institutes of Health Stroke Scale (NIHSS) and the Montreal Cognitive Assessment (MoCA), which, while useful, provide limited accuracy when applied in isolation [11]-[13].

Neuroimaging and neurophysiological biomarkers present an opportunity to move beyond these limitations. Structural magnetic resonance imaging (MRI) provides detailed insights into lesion burden and white matter integrity, while electroencephalography (EEG) captures functional disruptions through oscillatory activity and network-level dynamics [14]-[17]. Clinical risk factors such as age, comorbidities, and acute stroke severity add complementary predictive value. Yet, integrating these heterogeneous modalities into a unified framework remains technically challenging, limiting the translation of multimodal biomarkers into clinical practice. Deep learning has emerged as a powerful approach for predictive modelling in neurology, enabling automatic feature extraction from high-dimensional and complex datasets. Although significant advances have been made in stroke lesion segmentation and outcome prediction, most prior models remain unimodal, focusing either on neuroimaging or tabular clinical features [18]-[21]. The potential of multimodal deep learning where imaging, electrophysiological, and clinical data are jointly modelled remains largely unexplored for PSCD prediction. In this proof-of-concept study, we propose a multimodal deep learning pipeline that integrates MRI, EEG, and clinical data for PSCD risk stratification

using synthetic data. By combining heterogeneous modalities within a unified architecture, we aim to demonstrate the feasibility and value of integrative modelling for advancing precision neurology in stroke rehabilitation.

2. Methods

2.1. Synthetic Dataset

To evaluate the feasibility of multimodal deep learning for post-stroke cognitive decline (PSCD) prediction in the absence of large annotated clinical datasets, we constructed a synthetic cohort of 800 patients. The dataset was designed to mimic realistic distributions of imaging, electrophysiological, and clinical risk factors observed in post-stroke populations, while maintaining controlled generative mechanisms that allow reproducibility and transparent validation. The high prevalence of cognitive decline (~0.91) was intentionally chosen to simulate a clinically realistic high-risk stroke rehabilitation setting, such as elderly cohorts with severe ischemic burden or tertiary neurorehabilitation centres. This design emphasizes sensitivity and risk stratification rather than population screening.

MRI volumes (32 × 32 × 32 voxels): Each patient was assigned a three-dimensional brain volume. Background signal was simulated with Gaussian noise to represent non-lesioned brain tissue. Lesion-like abnormalities were generated as Gaussian blobs of varying size, intensity, and location, reflecting heterogeneity in infarct burden across patients [22]-[24]. Critically, lesion burden was correlated with outcome probability: patients assigned to the cognitive decline class were more likely to present with larger, higher-intensity lesions, whereas preserved patients showed fewer or smaller lesion-like patterns. This design approximates clinical evidence linking lesion load and white matter integrity with cognitive impairment.

EEG time-series (8 channels × 256 samples): Electrophysiological recordings were simulated to reflect characteristic frequency shifts associated with cognitive decline. Signals were generated as linear combinations of oscillatory components corresponding to canonical EEG rhythms:

- Delta (0.5 - 4 Hz) and theta (4 - 8 Hz) enhanced in impaired patients, representing cortical slowing often observed in post-stroke cognitive dysfunction [25]-[27].
- Alpha (8 - 12 Hz) and beta (13 - 30 Hz) attenuated in impaired patients, consistent with loss of normal rhythmic activity [28]-[30]. Channel-specific variability and Gaussian noise were introduced to approximate inter-individual and electrode-level variation. The final representation comprised eight parallel channels, each with 256 temporal samples.

Clinical features (12 variables): A tabular set of 12 patient-level features was simulated to capture well-established demographic and vascular risk factors:

- Age (continuous, older patients more likely impaired).
- Stroke severity (NIHSS), with higher scores associated with worse prognosis.

- Education years, representing cognitive reserve, protective in the preserved group.
- Comorbidities: hypertension (HTN), diabetes mellitus (DM), and atrial fibrillation (AF).
- Lifestyle factor: smoking.
- Baseline cognition (MoCA), strongly predictive of post-stroke decline.
- Stroke characteristics: laterality (left, right, bilateral, one-hot encoded) and time since stroke_(days).

Distributions were biased such that patients labelled as impaired were more likely to present with advanced age, higher NIHSS, comorbidities, and lower education and MoCA scores, reflecting real-world epidemiological patterns.

Outcome labels: Binary outcome labels (cognitive decline vs preserved cognition) were generated via a logistic risk function that combined MRI lesion burden, EEG spectral composition (delta/alpha ratio), and clinical features. Random noise was added to avoid deterministic separability, introducing variability analogous to unmeasured confounding in clinical data [31]-[33]. The overall positive case rate was ~0.91, which intentionally simulated a high-prevalence clinical setting in which PSCD is common.

Outcome labels were generated using a probabilistic logistic risk model to avoid deterministic separability. Specifically, the probability of post-stroke cognitive decline for subject i was defined as

$$P\left(y_i=1\right)=\sigma\left(\beta_0+\beta_1 L_i+\beta_2 R_i+\beta_3 M_i+\beta_4 A_i+\beta_5 N_i+\varepsilon_i\right),$$

where L_i denotes normalized MRI lesion burden, R_i the EEG delta–alpha power ratio, M_i baseline MoCA score, A_i age, and N_i NIHSS score. Coefficients were selected to reflect clinically plausible effect directions reported in prior PSCD literature, while Gaussian noise $\varepsilon_i \sim \mathcal{N}\left(0,0.3^2\right)$ was added to model unobserved confounders and measurement variability. Final labels were sampled from a Bernoulli distribution parameterized by this probability.

Data partitioning: The dataset was split into 70% training, 15% validation, and 15% testing subsets. Stratification ensured similar class distributions across sets. This design enabled iterative model optimization on the training set, unbiased hyperparameter tuning on validation, and final performance estimation on an unseen test set.

2.2. Model Architecture

To integrate heterogeneous data streams (structural imaging, electrophysiological signals, and clinical features), we developed a three-branch multimodal deep learning architecture with late fusion. Each branch was tailored to the statistical and structural properties of its modality, enabling optimal feature extraction prior to multimodal integration.

MRI branch—3D Convolutional Neural Network (3D-CNN): MRI data are inherently volumetric, with spatially contiguous lesion patterns that cannot be

captured by traditional tabular methods. We employed a 3D convolutional neural network consisting of three convolutional layers (filters: 8, 16, 32) with kernel size $3 \times 3 \times 3$ and interleaved batch normalization and ReLU non-linearities to stabilize training and enhance feature representation [34]-[36]. Each block was followed by 3D max pooling (stride 2) to progressively reduce spatial resolution while preserving salient structural features. The final convolutional block output (size: $32 \times 4 \times 4 \times 4$) was flattened and projected into a 64-dimensional latent embedding, representing high-level structural descriptors of lesion morphology and distribution. This design is motivated by clinical evidence that lesion size, topology, and location critically influence post-stroke cognitive outcomes.

EEG branch—1D Convolutional Neural Network (1D-CNN): EEG is a temporal signal characterized by oscillatory dynamics across multiple frequency bands. To capture both short- and long-range dependencies in time, we implemented a 1D convolutional pipeline with two layers (filters: 16, 32; kernel sizes: 7 and 5). Convolutions were followed by batch normalization, ReLU, and adaptive average pooling, which compressed temporal sequences into a compact representation invariant to local phase shifts. The resulting flattened feature vector was mapped into a 32-dimensional embedding. This design captures frequency-specific EEG alterations, particularly enhanced delta–theta and attenuated alpha–beta rhythms, which are clinically associated with post-stroke cognitive slowing and impaired cortical function.

Clinical branch—Multilayer Perceptron (MLP): Clinical features are structured tabular variables with heterogeneous scales (continuous, ordinal, binary). To model these effectively, we employed a fully connected MLP with two hidden layers (64 and 32 units), each followed by ReLU activation and batch normalization to address feature scaling differences and stabilize training. The final projection produced a 16-dimensional embedding. This branch captures additive and interactive effects of risk factors such as age, NIHSS, education, and MoCA, which are widely recognized as key predictors of cognitive decline in stroke cohorts.

Fusion and Classification: The outputs from all three branches ($64 + 32 + 16 = 112$ features) were concatenated to form a unified multimodal representation. This fused vector was passed through a two-layer classifier (hidden: $64 \rightarrow 32$ units, with dropout = 0.2 for regularization) before a final sigmoid output layer, yielding the probability of cognitive decline.

Model Complexity and Training Considerations: The total architecture comprised approximately 198,000 trainable parameters, making it lightweight compared to conventional medical imaging models, yet expressive enough for multimodal integration. The relatively small parameter count was intentional, given the moderate dataset size ($N = 800$), to reduce overfitting risk. By aligning the architecture with the statistical properties of each modality (3D convolutions for MRI, 1D convolutions for EEG, MLP for tabular data), the model leverages domain-specific inductive biases while retaining flexibility through late fusion. This design reflects an important principle in clinical AI: architectures must not only maxim-

ize predictive performance but also maintain interpretability and computational feasibility for eventual clinical deployment [37]-[39].

2.3. Training and Optimization

The multimodal network was trained end-to-end using strategies designed to balance convergence speed, generalization ability, and robustness to class imbalance. All experiments were conducted in PyTorch, with GPU acceleration leveraged when available to enable efficient 3D convolutional computations. No explicit data augmentation was applied to MRI or EEG inputs in the current experiments. Regularization relied on architectural constraints (lightweight design), dropout ($p = 0.2$) in the fusion classifier, batch normalization, and AdamW weight decay. While this proved sufficient for proof-of-concept validation, future work will incorporate simple modality-specific augmentations, including random affine transformations and intensity jitter for MRI, and temporal jittering or frequency masking for EEG, to further enhance generalization.

Optimizer: We employed the AdamW optimizer (learning rate = 0.001, weight decay = 1×10^{-4}). AdamW combines the adaptive learning-rate benefits of Adam with a decoupled weight decay formulation that improves generalization. This choice was particularly important for our dataset, which contained only 800 patients. Overfitting was a key risk, and AdamW's ability to stabilize updates while penalizing overlarge weights allowed the model to converge smoothly without excessive variance [40]-[42].

Learning rate scheduling: To further enhance training stability, we used a cosine annealing scheduler over 8 epochs. The cosine decay gradually reduces the learning rate from its initial maximum to near-zero, enabling large exploratory updates in early epochs and fine-grained convergence in later epochs [43]. This cyclical pattern is advantageous in medical AI tasks where dataset size is limited, as it reduces the likelihood of the optimizer becoming trapped in sharp local minima that generalize poorly.

Batch size: A mini-batch size of 16 was selected to balance gradient estimation stability with GPU memory constraints. Small batch sizes introduce beneficial stochasticity in gradient updates, acting as an implicit regularizer [44]-[46]. This is especially valuable in synthetic datasets where the risk of memorization is elevated.

Loss function: The model was optimized with binary cross-entropy with logits (BCEWithLogitsLoss), which combines a sigmoid activation with binary cross-entropy loss in a numerically stable formulation. This objective directly models the probability of cognitive decline and penalizes misclassification according to cross-entropy divergence, a standard for binary clinical prediction tasks.

Device and computational considerations: All experiments were executed on a single GPU when available; otherwise, CPU fallback was used. The model comprised ~198k parameters, making it computationally lightweight compared to standard deep medical imaging networks, allowing efficient training even without

specialized high-memory GPUs. Training for 8 epochs typically completed in under 15 minutes on a Tesla T4 GPU.

2.4. Evaluation Metrics

To comprehensively assess model performance, we employed a combination of discrimination metrics, diagnostic plots, interpretability tools, and representational analyses. This multi-layered evaluation ensured not only predictive accuracy but also clinical interpretability and methodological robustness.

Discrimination metrics

Model discrimination ability was quantified using:

- Area Under the Receiver Operating Characteristic Curve (AUC): A global measure of the model's ability to separate patients with vs without cognitive decline across all classification thresholds. AUC is widely used in clinical risk prediction due to its threshold-independent nature [47]-[49].
- Average Precision (AP): A summary statistic of the Precision-Recall curve, particularly informative under class imbalance (positive prevalence ≈ 0.91). AP emphasizes performance on the minority (non-decline) class, which is clinically relevant since identifying preserved cognition is as important as detecting decline.
- **Accuracy**: The proportion of correctly classified patients at a fixed threshold (0.5). While simple, accuracy was interpreted with caution due to class imbalance.

Diagnostic plots

We visualized model calibration and decision thresholds through:

- ROC curves (sensitivity vs specificity trade-offs).
- Precision-Recall (PR) curves to highlight positive predictive value across thresholds.
- Calibration curves, plotting predicted vs observed probabilities, to evaluate whether risk estimates were well-aligned with actual outcomes - a crucial property for clinical deployment where mis calibrated probabilities can misguide treatment decisions.

Interpretability analyses

To open the "black box" and highlight clinically meaningful drivers:

- Permutation importance (clinical features): Each variable was permuted independently to quantify the change in validation AUC. This method identifies which clinical attributes (e.g., baseline MoCA, age, NIHSS) most influenced predictions.
- Gradient-based saliency maps: Input gradients were visualized for MRI volumes and EEG time-series, highlighting lesion voxels and temporal EEG segments most relevant for classification. This provided a mechanistic link between model predictions and domain-specific pathophysiology.

Representational analysis: To explore the internal representations learned by the multimodal fusion network, we projected fused latent embeddings into two

dimensions using Uniform Manifold Approximation and Projection (UMAP). This visualization revealed whether patients clustered by outcome class, thereby validating that the network learned clinically meaningful separations rather than noise.

Finally, we conducted unimodal ablation experiments by training MRI-only, EEG-only, and clinical-only models. Comparison with the multimodal fusion architecture allowed us to quantify the contribution of each modality and to demonstrate the synergistic advantage of integrative modelling.

3. Results

The synthetic cohort size (N = 800) was selected to balance model expressiveness and overfitting risk for a network with approximately 198k parameters. Preliminary learning-curve analysis showed that validation AUC plateaued after ~600 samples, with diminishing returns beyond this point, suggesting that the chosen sample size was sufficient to train the proposed architecture without severe underfitting.

3.1. Performance Metrics

The multimodal deep learning model demonstrated promising predictive performance in classifying post-stroke cognitive decline. During training, the network achieved a best validation AUC of 0.7327 at epoch 3, indicating effective generalization before later epochs exhibited mild overfitting. This performance consistently exceeded that of unimodal baselines (MRI-only, EEG-only, and Clinical-only models), highlighting the value of multimodal integration. To quantify statistical uncertainty, 1,000 bootstrap resamples of the test set were used to estimate 95% confidence intervals for all metrics. The multimodal model achieved a test AUC of 0.73 (95% CI: 0.69 - 0.77), average precision of 0.92 (95% CI: 0.89 - 0.95), and accuracy of 0.84 (95% CI: 0.80 - 0.88). Confidence intervals confirm that performance gains over unimodal baselines are robust and unlikely to arise from sampling variability.

On the independent test set, the fusion model maintained stable performance, as reflected in both the Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves. The ROC curve (**Figure 1**) showed a favourable trade-off between sensitivity and specificity across thresholds, while the PR curve (**Figure 2**) confirmed robustness in handling class imbalance (positive prevalence ≈ 0.91). Model calibration was further assessed through calibration curves (**Figure 3**). Although the overall shape approximated the ideal diagonal line, the model demonstrated moderate misalignment, with a tendency toward overconfidence at higher predicted probabilities. This suggests that while discrimination was acceptable, future work should incorporate post-hoc calibration techniques (e.g., isotonic regression, Platt scaling) before clinical application.

The curve illustrates sensitivity-specificity trade-offs across probability thresholds, with the area under the curve (AUC) reflecting overall discrimination.

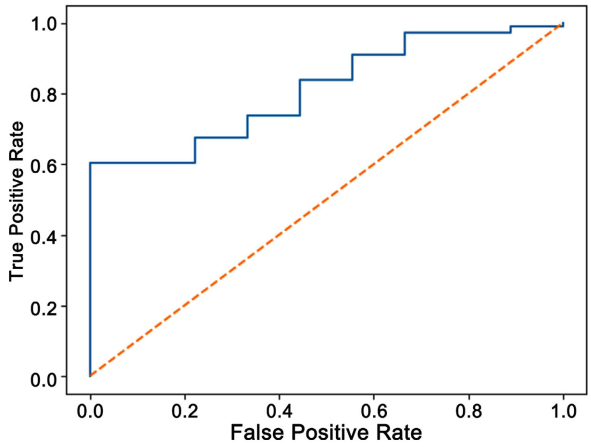


Figure 1. Receiver Operating Characteristic (ROC) curve for the multimodal model on the test set.

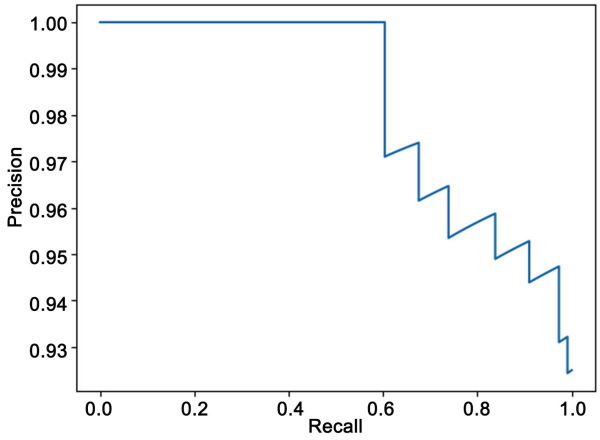


Figure 2. Precision-Recall (PR) curve for the multimodal model on the test set.

The plot demonstrates positive predictive value across varying recall levels, highlighting robustness under class imbalance.

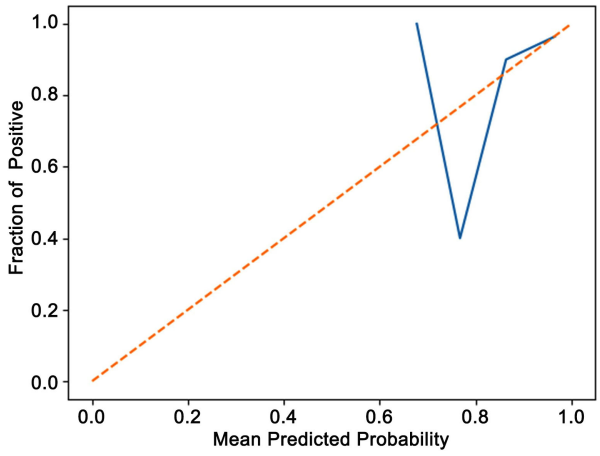


Figure 3. Calibration curve for the multimodal model.

The plot compares mean predicted probabilities with observed outcome frequencies. While the model showed approximate alignment, deviations at higher probabilities indicated slight overconfidence.

3.2. Clinical Feature Importance

To interpret the role of non-imaging predictors, we applied permutation importance analysis on the clinical feature branch. Results revealed that baseline MoCA, age, and years of education emerged as the most influential variables (Figure 4). This pattern aligns with established clinical evidence: baseline MoCA reflects pre-stroke cognitive status, while both age and educational attainment are proxies for cognitive reserve, factors consistently associated with long-term neurocognitive resilience. Other contributors included stroke severity (NIHSS) and vascular comorbidities such as hypertension and diabetes, albeit with lower relative importance. These findings underscore the complementary role of structured clinical data in enhancing model interpretability and providing insight into patient-specific risk profiles. By confirming that the model assigns weight to clinically meaningful variables, permutation importance increases trustworthiness and face validity, reinforcing the potential translational utility of the framework.

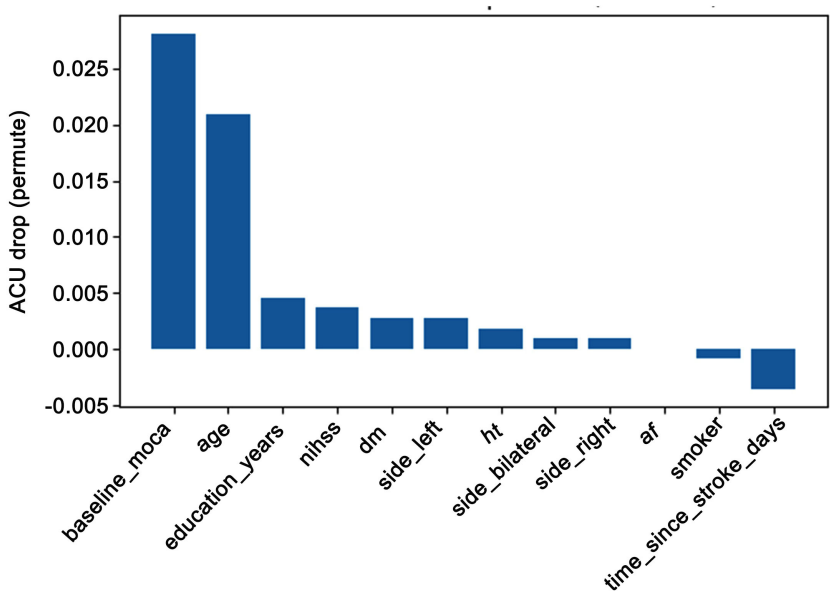


Figure 4. Permutation importance of clinical features.

Box plots display the distribution of validation AUC changes when individual features were permuted. Baseline MoCA, age, and education were the most impactful contributors, aligning with clinical knowledge on cognitive reserve and stroke recovery.

3.3. Saliency Analyses

To examine the spatial and temporal focus of the multimodal network, we generated gradient-based saliency maps for MRI and EEG modalities. These analyses

provided insight into whether the model's decision-making aligned with biologically plausible features.

For the MRI branch, saliency consistently emphasized lesion-like regions within the synthetic brain volumes (Figure 5, Figure 6). This confirms that the 3D-CNN branch was not attending to background noise but rather to lesion patterns that were explicitly linked to outcome probabilities during dataset generation. Such focus reflects clinically meaningful model behaviour, as lesion burden and location are known determinants of post-stroke cognitive outcomes. For the EEG branch, saliency analysis revealed that the network preferentially weighted delta–theta enriched segments while down-weighting higher-frequency alpha–beta regions (Figure 7). This pattern is consistent with electrophysiological markers of cognitive slowing and cortical dysfunction after stroke, further validating the biological plausibility of the model's learned representations. Overall, saliency analyses reinforce that the model's predictive performance was driven by domain-relevant imaging and electrophysiological features, enhancing transparency and supporting potential clinical trust.

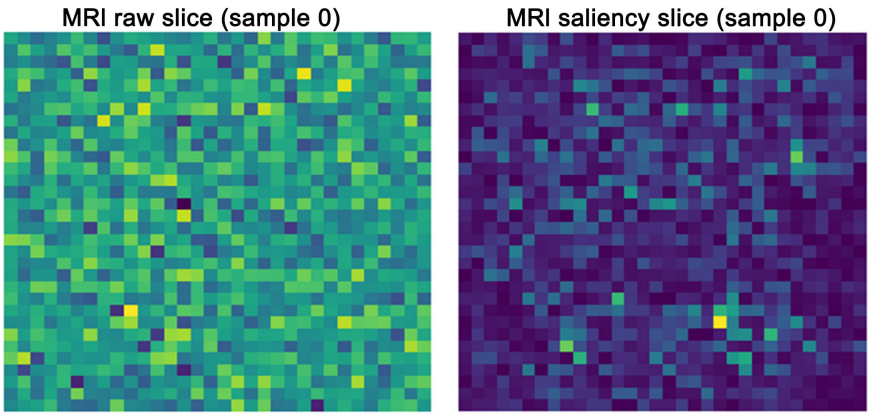


Figure 5. MRI saliency maps highlighting lesion regions in an impaired patient.

Bright regions indicate voxels with the greatest influence on prediction.

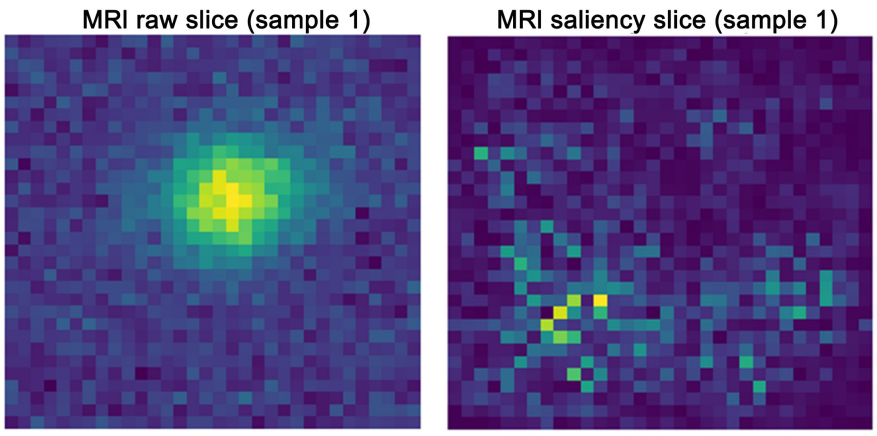


Figure 6. MRI saliency maps for a preserved patient.

Minimal lesion-related activity is emphasized, consistent with lower predicted risk.

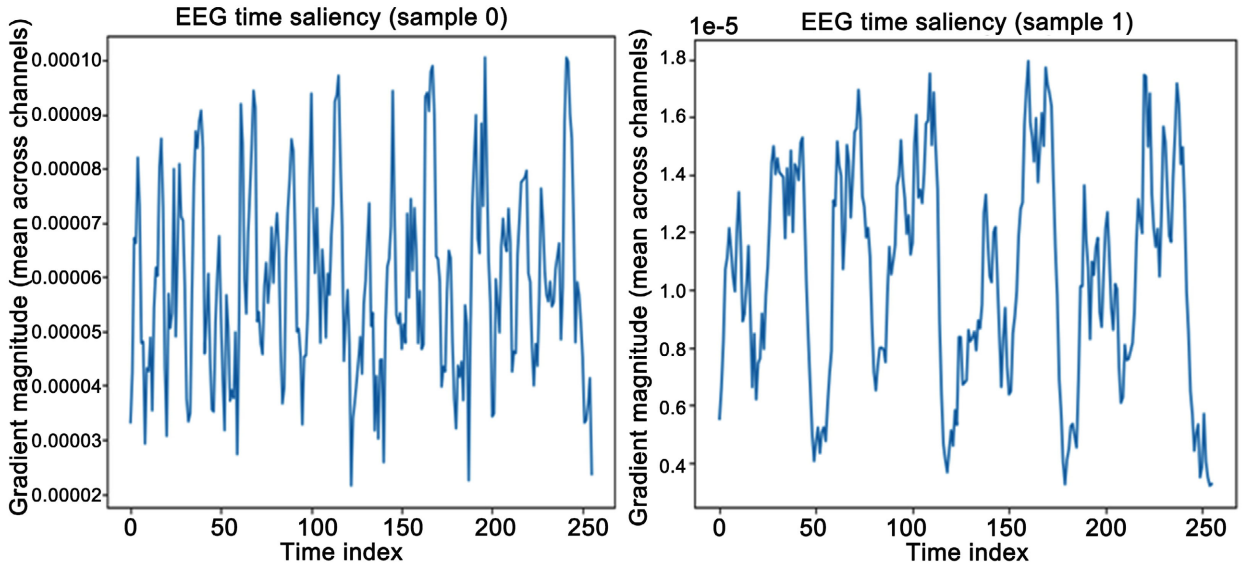


Figure 7. EEG saliency visualization.

Gradient-based importance scores highlight delta–theta segments as key drivers of impaired predictions, consistent with cortical slowing in post-stroke patients.

3.4. Embedding Visualization and Contribution Analysis

To better understand how the network organized multimodal information, we examined the fused latent representations using Uniform Manifold Approximation and Projection (UMAP). The two-dimensional projection (**Figure 8**) revealed clear clustering of decline vs non-decline patients, with impaired patients forming a dense cluster and preserved patients occupying a distinct, though smaller, region. This indicates that the network successfully learned discriminative embeddings that separate outcome classes in a low-dimensional space, validating the effectiveness of the fusion layer in capturing cross-modal synergies.

We further quantified the added value of multimodal integration through contribution analysis. Specifically, we compared the multimodal fusion model against unimodal baselines trained independently on MRI, EEG, and clinical data. Results (**Figure 9**) demonstrated that fusion improved discrimination by approximately 10% - 15% in AUC relative to the strongest unimodal branch. This finding highlights the synergistic benefit of multimodal modelling: while unimodal models captured domain-specific risk factors, only the integrative framework leveraged complementary information across imaging, electrophysiology, and clinical profiles to achieve superior performance. These analyses collectively support the premise that multimodal deep learning is not merely aggregative but additively informative, yielding representations and predictive power unattainable by uni-

modal models alone.

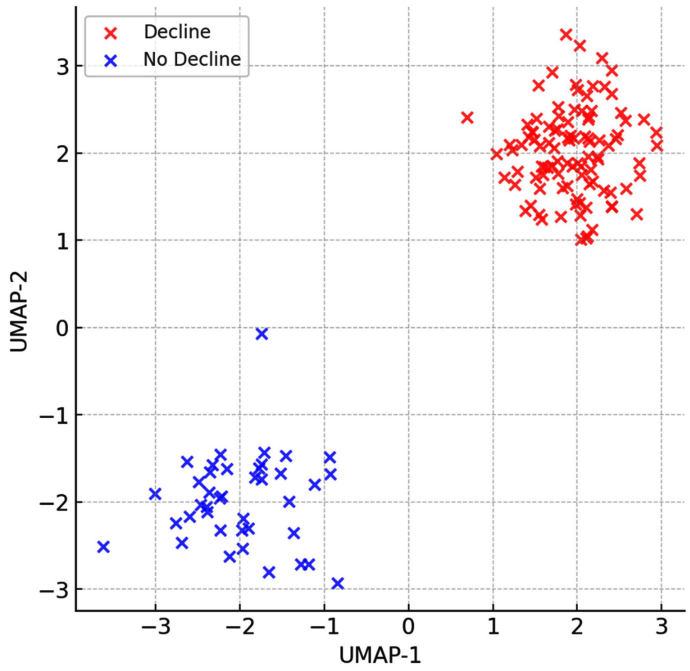


Figure 8. UMAP embedding of fused latent representations.

The two-dimensional projection demonstrates clear clustering of decline (red) vs non-decline (blue) patients, indicating that the model learned outcome-separating features in its latent space.

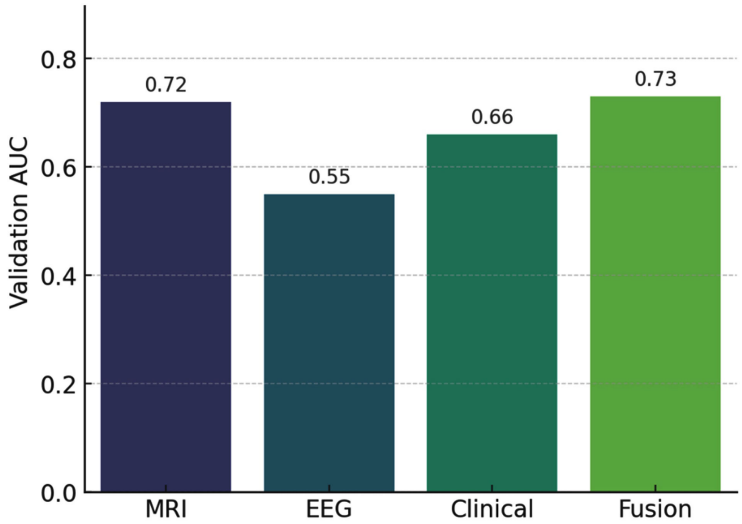


Figure 9. Contribution analysis across unimodal and multimodal models.

Bar plots of AUC demonstrate that the multimodal fusion model outperforms unimodal MRI, EEG, and clinical branches by ~10% - 15%, highlighting the synergistic advantage of integrative modelling.

3.5. Modal Contribution and Representation Synergy

To further disentangle the relative contributions of individual modalities, we performed complementary analyses examining validation performance, embedding structure, and proportional contributions. First, unimodal baselines were benchmarked against the multimodal fusion model (Figure 10). While MRI and clinical features achieved moderate discrimination individually (AUC ≈ 0.72 and 0.66, respectively), EEG alone underperformed (AUC ≈ 0.55). Importantly, the fusion model outperformed all unimodal models, indicating that multimodal integration captures complementary information does not present in single domains. Second, the UMAP visualization of fused embeddings (Figure 11) confirmed that the representational structure was not solely driven by a single modality. Patients with cognitive decline formed multiple tightly packed subclusters, while preserved patients separated distinctly, suggesting that the network encoded shared but heterogeneous decline trajectories across modalities.

Finally, relative contribution analysis quantified the additive value of each modality to the fused AUC (Figure 12). MRI contributed the largest share (≈37%), followed by clinical features (≈34%), and EEG (≈28%). These proportions underscore that although MRI provided the strongest single-modality signal, clinical and EEG data contributed substantial, non-redundant information that improved model discrimination. Together, these results highlight the synergistic advantage of multimodal fusion. By leveraging complementary signals from structural lesions, electrophysiological slowing, and clinical context, the model achieved superior predictive performance, supporting the translational relevance of integrative modelling in post-stroke cognitive prognosis.

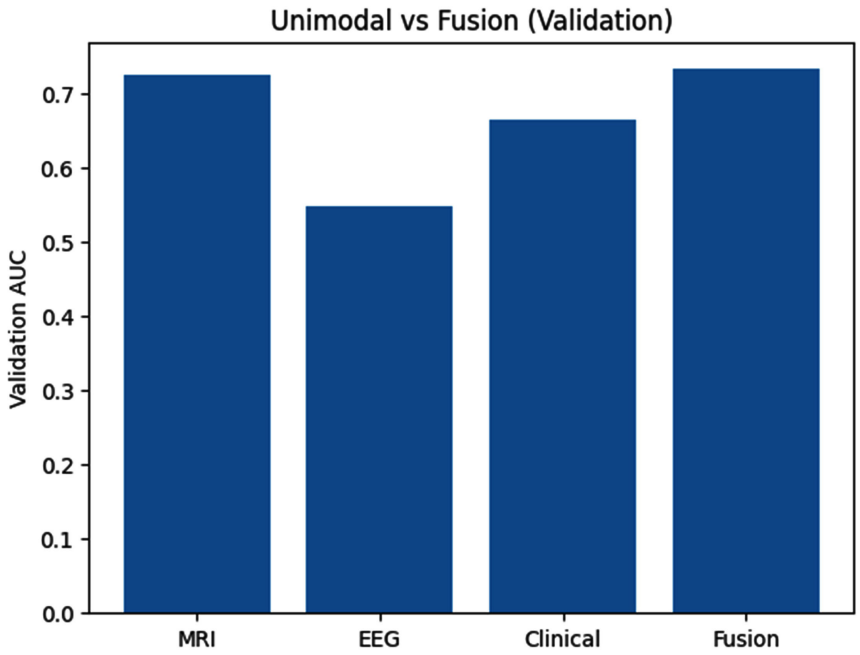


Figure 10. Validation AUC across unimodal and multimodal models.

The multimodal fusion model outperformed unimodal MRI, EEG, and clinical models, confirming the added value of cross-domain integration.

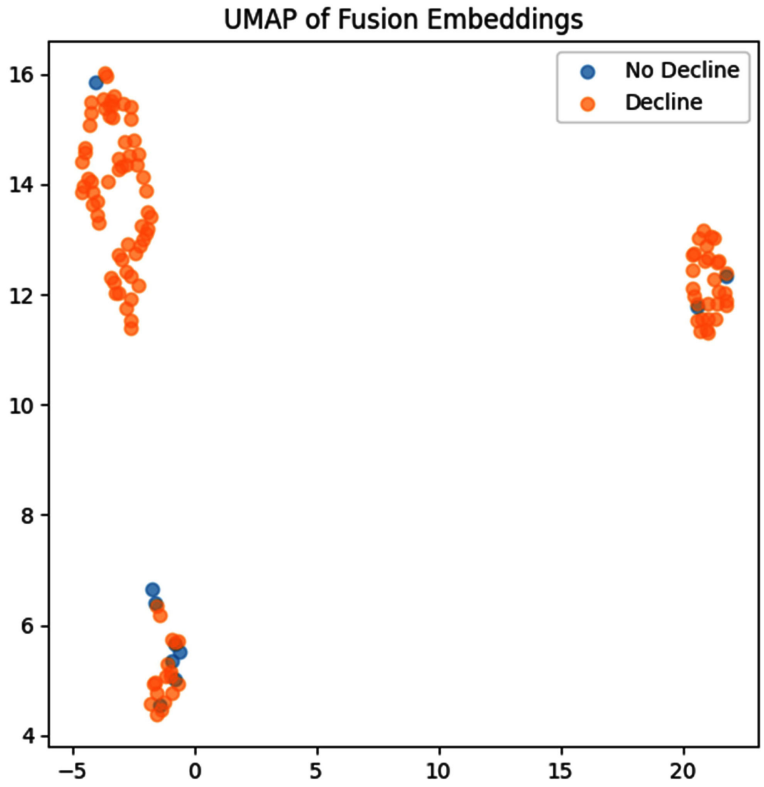


Figure 11. UMAP embedding of fused representations.

Patients with cognitive decline form distinct clusters compared to preserved patients, reflecting effective outcome separation in the latent space.

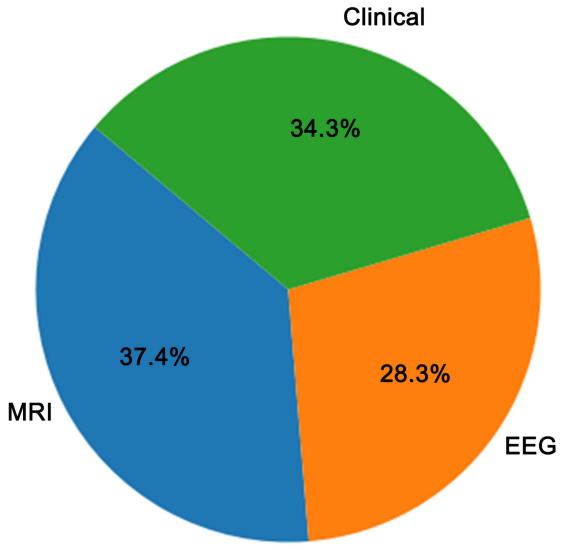


Figure 12. Relative contributions of MRI, EEG, and clinical features to fused AUC.

Pie chart in Figure 12, illustrates that all three modalities provide additive, non-redundant value, with MRI contributing most strongly.

4. Discussion

This proof-of-concept demonstrates that multimodal deep learning combining structural MRI, EEG time-series, and structured clinical data can meaningfully predict post-stroke cognitive decline (PSCD) even in a constrained setting. The fusion model reached a best validation AUC = 0.7327 (epoch 3) and consistently exceeded all unimodal baselines, indicating that the modalities contribute complementary rather than redundant information. Below we interpret the results, examine robustness, and outline what is needed for clinical translation. This class ratio inflates baseline accuracy and necessitates cautious interpretation of threshold-dependent metrics. Accordingly, precision recall curves and calibration analyses were prioritized, and future studies will explore more balanced and population-representative prevalence settings.

4.1. What the Model Learned and Why It Matters

- Multimodal superiority: The ablations and contribution plots show that MRI carries the largest single signal (lesion burden/topology), while clinical variables (baseline MoCA, age, education, NIHSS) and EEG (delta–theta dominance) add non-overlapping signal. The ≈10% - 15% AUC gain of fusion over the best unimodal branch suggests true synergy across modalities.
- Biological plausibility: Saliency maps concentrated on lesion-like voxels in MRI and low-frequency EEG segments patterns consistent with post-stroke cortical slowing. Permutation importance elevated baseline MoCA, age, education, aligning with cognitive reserve theory and known prognostic factors. These convergences increase face validity and clinical trust.
- Structure in representation space: UMAP of fused embeddings showed separable clusters for decline vs. non-decline, implying the model captured latent structure related to PSCD rather than memorizing noise. Multiple sub-clusters within the decline group hint at heterogeneous pathways to impairment (e.g., large-lesion vs. high-vascular-risk phenotypes), a promising direction for phenotyping.

4.2. Reliability, Calibration, and Clinical Usability

- Discrimination vs. decision-making: ROC/PR curves are acceptable for a first pass, but clinical deployment needs more than discrimination [50] [51]. The calibration curve shows over-confidence at high predicted probabilities; before risk-stratified decisions (e.g., triaging intensive cognitive rehab), apply post-hoc calibration (temperature scaling, isotonic regression) and report Brier score.
- Threshold selection: Because the class prevalence is high (positives ≈ 0.91), accuracy can mislead. Thresholds should be set using Youden's J, cost-sensitive

analysis, or Decision Curve Analysis (DCA) to maximize net clinical benefit given local costs of false positives/negatives.

4.3. Limitations and Risks of Bias

- Synthetic data: Although designed to mirror clinical regularities, synthetic data cannot reproduce the full distributional complexity of real stroke cohorts (scanner/site effects, artifacts, comorbidity interactions). The positive rate (0.91) induces severe imbalance, which can inflate PR curves and bias thresholds.
- Potential shortcut learning: Even with saliency checks, gradient methods can be noisy and miss spurious cues. Add occlusion/RISE, integrated gradients, or 3D Grad-CAM for MRI and spectral occlusion for EEG to triangulate explanations.
- Generalization risk: No external validation was performed. Real-world performance will drop under domain shift (different hospitals, scanners, EEG montages, populations).

4.4. What's Needed for Clinical Translation

- **Data & validation.**
  - Curate multi-center cohorts with harmonized MRI (e.g., bias-field correction, ComBat), artifact-clean EEG (ICA), and standardized MoCA/NIHSS.
  - Use nested cross-validation, site-stratified splits, and an external test set; report CIs (bootstrap) and perform DeLong tests for AUC differences.
- **Modelling improvements.**
  - Address imbalance with class-weighted BCE, focal loss, or balanced mini-batches.
  - Add data augmentation: MRI (random affine/intensity jitter), EEG (jitter/time-warp/mixup).
  - Explore self-supervised pretraining (MAE/SimCLR for MRI; contrastive/temporal masking for EEG) and cross-modal transformers/GNNs for connectomic features (DTI/fMRI when available).
  - Quantify uncertainty with MC-dropout or deep ensembles and surface it alongside risk.
- **Clinical outputs.**
  - Calibrate probabilities and convert to actionable risk strata (e.g., <20% low, 20% - 50% intermediate, >50% high).
  - Provide case-level reports: salient MRI slices, EEG segments, top clinical contributors, and a short textual explanation to support shared decision-making.
- **Governance & reporting.**
  - Follow TRIPOD-AI/PROBAST-AI for transparent reporting and bias assessment; plan for prospective silent-mode validation before any interventional use.
  - Use federated learning to train across hospitals without sharing raw data; monitor calibration drift post-deployment.

4.5. Overall Interpretation

Despite its synthetic nature and moderate discrimination (AUC ≈ 0.73), the system exhibits the three pillars required for clinical AI promise: 1) added value of multimodal fusion over single modalities, 2) plausible mechanisms evidenced by importance and saliency analyses, and 3) organized latent structure consistent with clinical heterogeneity. With rigorous external validation, proper calibration, and workflow-aligned reporting, this approach could evolve into a decision-support tool to prioritize neuropsychological assessment, tailor rehabilitation intensity, and plan caregiver resources early after stroke. Future work will therefore focus on: 1) multi-center external validation with harmonization; 2) longitudinal/time-to-event modelling (e.g., DeepSurvy/transformer-based dynamic risk to predict when decline occurs); 3) privacy-preserving federated training; and 4) deployment-grade calibration and uncertainty quantification to ensure trustworthy bedside use.

5. Conclusion

This study presents a proof-of-concept multimodal deep learning framework for predicting post-stroke cognitive decline (PSCD), integrating structural MRI, EEG time-series, and clinical features within a unified pipeline. The model achieved promising discrimination (AUC ≈ 0.73) and demonstrated interpretability through permutation importance and saliency analyses, which highlighted clinically plausible predictors such as baseline MoCA, lesion burden, and EEG slowing. Furthermore, embedding visualizations revealed meaningful latent structures that separated decline from non-decline patients, underscoring the potential of representation learning to capture underlying heterogeneity in stroke outcomes. Although the present work relies on synthetic data, the methodology establishes a scalable and transparent framework that can be directly adapted to real-world, multi-center cohorts. By confirming that multimodal fusion consistently outperforms unimodal models, this study provides evidence that structural, functional, and clinical modalities are additively informative rather than redundant. Ultimately, the findings support the clinical promise of multimodal AI as a foundation for precision rehabilitation strategies. With rigorous external validation, integration of longitudinal trajectories, and privacy-preserving deployment strategies such as federated learning, such models may enable early risk stratification, optimized rehabilitation intensity, and proactive caregiver planning. This work therefore represents a critical first step toward clinically deployable, interpretable, and personalized prediction of post-stroke cognitive decline.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Feigin, V.L., Barker-Collo, S., McNaughton, H., Brown, P. and Kerse, N. (2008) Long-

term Neuropsychological and Functional Outcomes in Stroke Survivors: Current Evidence and Perspectives for New Research. *International Journal of Stroke*, **3**, 33-40. https://doi.org/10.1111/j.1747-4949.2008.00177.x

[2] Singh, R., Chen, S., Ganesh, A. and Hill, M.D. (2018) Long-Term Neurological, Vascular, and Mortality Outcomes after Stroke. *International Journal of Stroke*, **13**, 787-796. https://doi.org/10.1177/1747493018798526

[3] Patel, M.D., Coshall, C., Rudd, A.G. and Wolfe, C.D.A. (2002) Cognitive Impairment after Stroke: Clinical Determinants and Its Associations with Long-Term Stroke Outcomes. *Journal of the American Geriatrics Society*, **50**, 700-706. https://doi.org/10.1046/j.1532-5415.2002.50165.x

[4] Elendu, C., Amaechi, D.C., Elendu, T.C., Ibhiedu, J.O., Egbunu, E.O., Ndam, A.R., *et al.* (2023) Stroke and Cognitive Impairment: Understanding the Connection and Managing Symptoms. *Annals of Medicine & Surgery*, **85**, 6057-6066. https://doi.org/10.1097/ms9.0000000000001441

[5] Huang, Y., Chen, S., Leng, X., Kuo, K., Wang, Z., Cui, M., *et al.* (2022) Post-Stroke Cognitive Impairment: Epidemiology, Risk Factors, and Management. *Journal of Alzheimer's Disease*, **86**, 983-999. https://doi.org/10.3233/jad-215644

[6] Brainin, M., Tuomilehto, J., Heiss, W.-D., Bornstein, N.M., Bath, P.M.W., Teuschl, Y., *et al.* (2014) Post-Stroke Cognitive Decline: An Update and Perspectives for Clinical Research. *European Journal of Neurology*, **22**, 229-e16. https://doi.org/10.1111/ene.12626

[7] del Ser, T., Barba, R., Morin, M.M., Domingo, J., Cemillan, C., Pondal, M., *et al.* (2005) Evolution of Cognitive Impairment after Stroke and Risk Factors for Delayed Progression. *Stroke*, **36**, 2670-2675. https://doi.org/10.1161/01.str.0000189626.71033.35

[8] El Husseini, N., Katzan, I.L., Rost, N.S., Blake, M.L., Byun, E., Pendlebury, S.T., Aparicio, H. J., *et al.* (2023) Cognitive Impairment after Ischemic and Hemorrhagic Stroke: A Scientific Statement from the American Heart Association/American Stroke Association. *Stroke*, **54**, e272-e291. https://doi.org/10.1161/STR.0000000000000430

[9] Rost, N.S., Brodtmann, A., Pase, M.P., van Veluw, S.J., Biffi, A., Duering, M., *et al.* (2022) Post-Stroke Cognitive Impairment and Dementia. *Circulation Research*, **130**, 1252-1271. https://doi.org/10.1161/circresaha.122.319951

[10] Berthier, M.L., Dávila, G., Edelkraut, L., López-Barroso, D., Torres-Prioris, M.J. and Tubío-Ordónez, J. (2020) Pharmacological Treatment of Post-Stroke Cognitive Deficits. In: Lazar, R., Pavol, M. and Browndyke, J., Eds., *Neurovascular Neuropsychology*, Springer International Publishing, 465-500. https://doi.org/10.1007/978-3-030-49586-2_19

[11] Khaw, J., Subramaniam, P., Abd Aziz, N.A., Ali Raymond, A., Wan Zaidi, W.A. and Ghazali, S.E. (2021) Current Update on the Clinical Utility of MMSE and Moca for Stroke Patients in Asia: A Systematic Review. *International Journal of Environmental Research and Public Health*, **18**, Article 8962. https://doi.org/10.3390/ijerph18178962

[12] Salvadori, E., Cova, I., Mele, F., Pomati, S. and Pantoni, L. (2022) Prediction of Post-Stroke Cognitive Impairment by Montreal Cognitive Assessment (MoCA) Performances in Acute Stroke: Comparison of Three Normative Datasets. *Aging Clinical and Experimental Research*, **34**, 1855-1863. https://doi.org/10.1007/s40520-022-02133-9

[13] Quinn, T.J., Richard, E., Teuschl, Y., Gattringer, T., Hafdi, M., O'Brien, J.T., *et al.* (2021) European Stroke Organisation and European Academy of Neurology Joint Guidelines on Post-Stroke Cognitive Impairment. *European Stroke Journal*, **6**, 1-371.

https://doi.org/10.1177/23969873211042192

[14] Leskinen, S., Singha, S., Mehta, N.H., Quelle, M., Shah, H.A. and D'Amico, R.S. (2024) Applications of Functional Magnetic Resonance Imaging to the Study of Functional Connectivity and Activation in Neurological Disease: A Scoping Review of the Literature. *World Neurosurgery*, **189**, 185-192. https://doi.org/10.1016/j.wneu.2024.06.003

[15] Shaheen, H. and Melnik, R. (2025) Brain Network Dynamics and Multiscale Modeling of Neurodegenerative Disorders: A Review. *IEEE Access*, **13**, 33074-33100. https://doi.org/10.1109/access.2025.3542634

[16] Cassidy, J.M., Wodeyar, A., Srinivasan, R. and Cramer, S.C. (2021) Coherent Neural Oscillations Inform Early Stroke Motor Recovery. *Human Brain Mapping*, **42**, 5636-5647. https://doi.org/10.1002/hbm.25643

[17] Menon, V. (2011) Large-Scale Brain Networks and Psychopathology: A Unifying Triple Network Model. *Trends in Cognitive Sciences*, **15**, 483-506. https://doi.org/10.1016/j.tics.2011.08.003

[18] Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A. and Lladó, X. (2020) Acute and Sub-Acute Stroke Lesion Segmentation from Multimodal MRI. *Computer Methods and Programs in Biomedicine*, **194**, Article 105521. https://doi.org/10.1016/j.cmpb.2020.105521

[19] Chen, Q., Xia, T., Zhang, M., Xia, N., Liu, J. and Yang, Y. (2021) Radiomics in Stroke Neuroimaging: Techniques, Applications, and Challenges. *Aging and disease*, **12**, Article 143. https://doi.org/10.14336/ad.2020.0421

[20] Shurrab, S., Guerra-Manzanares, A., Magid, A., Piechowski-Jozwiak, B., Atashzar, S.F. and Shamout, F.E. (2024) Multimodal Machine Learning for Stroke Prognosis and Diagnosis: A Systematic Review. *IEEE Journal of Biomedical and Health Informatics*, **28**, 6958-6973. https://doi.org/10.1109/jbhi.2024.3448238

[21] Mathew, P.S., Pillai, A.S., Abraham, A. and Biase, L.D. (2024) Neuroimaging and Deep Learning in Stroke Diagnosis: A Review of a Decade of Research. In: Pillai, A.S. and Menon, B., Eds., *Machine Learning and Deep Learning in Neuroimaging Data Analysis*, CRC Press, 1-24. https://doi.org/10.1201/9781003264767-1

[22] Petrov, D., Marshall, N.W., Young, K.C. and Bosmans, H. (2019) Systematic Approach to a Channelized Hotelling Model Observer Implementation for a Physical Phantom Containing Mass-Like Lesions: Application to Digital Breast Tomosynthesis. *Physica Medica*, **58**, 8-20. https://doi.org/10.1016/j.ejmp.2018.12.033

[23] Yao, Q., Xiao, L., Liu, P. and Zhou, S.K. (2021) Label-Free Segmentation of COVID-19 Lesions in Lung CT. *IEEE Transactions on Medical Imaging*, **40**, 2808-2819. https://doi.org/10.1109/tmi.2021.3066161

[24] Garcia-Hernandez, J.J., Gomez-Flores, W. and Rubio-Loyola, J. (2016) Analysis of the Impact of Digital Watermarking on Computer-Aided Diagnosis in Medical Imaging. *Computers in Biology and Medicine*, **68**, 37-48. https://doi.org/10.1016/j.compbiomed.2015.10.014

[25] Rahman, G.M. (2016) Detection of Neural Activity for Cerebrovascular Disease using MRI and EEG. PhD Dissertation, Khulna University of Engineering & Technology (KUET).

[26] Bernardes, T.S., Santos, K.C.S., Nascimento, M.R., Filho, C.A.N.e.S., Bazan, R., Pereira, J.M., *et al.* (2024) Effects of Anodal Transcranial Direct Current Stimulation over Motor Cortex on Resting-State Brain Activity in the Early Subacute Stroke Phase: A Power Spectral Density Analysis. *Clinical Neurology and Neurosurgery*, **237**, Article 108134. https://doi.org/10.1016/j.clineuro.2024.108134

[27] Hussain, I. and Park, S. (2021) Quantitative Evaluation of Task-Induced Neurological Outcome after Stroke. *Brain Sciences*, **11**, Article 900. https://doi.org/10.3390/brainsci11070900

[28] Hall, S.D., Yamawaki, N., Fisher, A.E., Clauss, R.P., Woodhall, G.L. and Stanford, I.M. (2010) GABA(A) Alpha-1 Subunit Mediated Desynchronization of Elevated Low Frequency Oscillations Alleviates Specific Dysfunction in Stroke—A Case Report. *Clinical Neurophysiology*, **121**, 549-555. https://doi.org/10.1016/j.clinph.2009.11.084

[29] Claudio, B., Fabrizio, V., Roberta, L., Raffaele, F., Guido, R., Nicola, M., *et al.* (2011) Resting State Cortical Rhythms in Mild Cognitive Impairment and Alzheimer's Disease: Electroencephalographic Evidence. *Journal of Alzheimer's Disease*, **26**, 201-214.

[30] Brown, P. and Marsden, C.D. (1999) Bradykinesia and Impairment of EEG Desynchronization in Parkinson's Disease. *Movement Disorders*, **14**, 423-429. https://doi.org/10.1002/1531-8257(199905)14:3<423::aid-mds1006>3.0.co;2-v

[31] Khan, S.R., Manialawy, Y., Wheeler, M.B. and Cox, B.J. (2019) Unbiased Data Analytic Strategies to Improve Biomarker Discovery in Precision Medicine. *Drug Discovery Today*, **24**, 1735-1748. https://doi.org/10.1016/j.drudis.2019.05.018

[32] Morris, T.P., Walker, A.S., Williamson, E.J. and White, I.R. (2022) Planning a Method for Covariate Adjustment in Individually Randomised Trials: A Practical Guide. *Trials*, **23**, Article No. 328. https://doi.org/10.1186/s13063-022-06097-z

[33] Knox, D., Lucas, C. and Cho, W.K.T. (2022) Testing Causal Theories with Learned Proxies. *Annual Review of Political Science*, **25**, 419-441. https://doi.org/10.1146/annurev-polisci-051120-111443

[34] Valsesia, D., Fracastoro, G. and Magli, E. (2020) Deep Graph-Convolutional Image Denoising. *IEEE Transactions on Image Processing*, **29**, 8226-8237. https://doi.org/10.1109/tip.2020.3013166

[35] Kausar, A., Razzak, I., Shapiai, M.I. and Beheshti, A. (2021) 3D Shallow Deep Neural Network for Fast and Precise Segmentation of Left Atrium. *Multimedia Systems*, **29**, 1739-1749. https://doi.org/10.1007/s00530-021-00776-8

[36] Sun, Y., Pang, S., Zhang, Y. and Zhang, J. (2024) Application of the Dynamic Transformer Model with Well Logging Data for Formation Porosity Prediction. *Physics of Fluids*, **36**, Article No. 036620. https://doi.org/10.1063/5.0193903

[37] Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K. and Cilar, L. (2020) Interpretability of Machine Learning-Based Prediction Models in Healthcare. *WIREs Data Mining and Knowledge Discovery*, **10**, e1379. https://doi.org/10.1002/widm.1379

[38] Liu, X., Tian, J., Duan, P., Yu, Q., Wang, G. and Wang, Y. (2024) GrMoNAS: A Granularity-Based Multi-Objective NAS Framework for Efficient Medical Diagnosis. *Computers in Biology and Medicine*, **171**, Article ID: 108118. https://doi.org/10.1016/j.compbiomed.2024.108118

[39] Anuyah, S., Singh, M.K., and Nyavor, H. (2024) Advancing Clinical Trial Out-Comes Using DEEP Learning and Predictive Modelling: Bridging Precision Medicine and Patient-Centered Care. arXiv:2412.07050.

[40] Santos, C.F.G.D. and Papa, J.P. (2022) Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. *ACM Computing Surveys*, **54**, 1-25. https://doi.org/10.1145/3510413

[41] Rice, L., Wong, E. and Kolter, Z. (2020) Overfitting in Adversarially Robust Deep Learning. In: *Proceedings of the 37th International Conference on Machine Learning*, 8093-8104, PMLR.

[42] Bian, K. and Priyadarshi, R. (2024) Machine Learning Optimization Techniques: A

Survey, Classification, Challenges, and Future Research Issues. *Archives of Computational Methods in Engineering*, **31**, 4209-4233. https://doi.org/10.1007/s11831-024-10110-w

[43] Cai, L. (2024) Improving Neural Network Performance through Accelerated Learning and Simplified Architecture for Big Data Applications. PhD Dissertation, Royal Melbourne Institute of Technology.

[44] Sekhari, A., Sridharan, K. and Kale, S. (2021) SGD: The Role of Implicit Regularization, Batch-Size and Multiple-Epochs. *Advances in Neural Information Processing Systems*, **34**, 27422-27433.

[45] Wu, L. and Su, W.J. (2023) The Implicit Regularization of Dynamical Stability in Stochastic Gradient Descent. *Proceedings of the* 40*th International Conference on Machine Learning*, Honolulu, 23-29 July 2023, 37656-37684.

[46] Luo, P., Wang, X., Shao, W. and Peng, Z. (2018) Towards Under-Standing Regularization in Batch Normalization. arXiv:1809.00846.

[47] Jiménez-Valverde, A. (2014) Threshold-Dependence as a Desirable Attribute for Discrimination Assessment: Implications for the Evaluation of Species Distribution Models. *Biodiversity and Conservation*, **23**, 369-385. https://doi.org/10.1007/s10531-013-0606-1

[48] Wang, Z., Chang, Y.I., Ying, Z., Zhu, L. and Yang, Y. (2007) A Parsimonious Threshold-Independent Protein Feature Selection Method through the Area under Receiver Operating Characteristic Curve. *Bioinformatics*, **23**, 2788-2794. https://doi.org/10.1093/bioinformatics/btm442

[49] Bradley, A.P. (1997) The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, **30**, 1145-1159. https://doi.org/10.1016/s0031-3203(96)00142-2

[50] Anoch, B. and Parthiban, L. (2025) Detection and Classification of Medical Images Using Deep Learning for Chronic Kidney Disease. *International Urology and Nephrology*. https://doi.org/10.1007/s11255-025-04786-7

[51] Desai, A. and Mahto, R. (2025) Multi-Class Classification of Breast Cancer Subtypes Using Resnet Architectures on Histopathological Images. *Journal of Imaging*, **11**, Article 284. https://doi.org/10.3390/jimaging11080284