SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 3, 2026

INTELLIGENT SPATIOTEMPORAL DEEP LEARNING FOR ISCHEMIC STROKE SCREENING, DETECTION, AND SEVERITY CLASSIFICATION THROUGH CROSS-MODAL NEUROIMAGING

Muhammad Suleman[*1], Kashif Ahmad[2], Muhammad Tanveer[3], Laiba Riaz[4], Dr. Junaid Akbar[5]

[*1]*Department of Electrical Engineering, University of Science and Technology, Bannu, Pakistan.*
[2]*Faculty of Engineering, Environment and Computing, Coventry University, Coventry, England, United Kingdom*
[3]*Escuela Técnica Superior de Ingenieros de Minas y Energía, Universidad Politécnica de Madrid, Madrid, Spain.*
[4]*School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan.*
[5]*Department of Software Engineering, Jiangsu University, Zhenjiang, China.*

[*1]sulemanktk@ustb.edu.pk, [2]kashif553@gmail.com, [3]tanveer822@gmail.com, [4]laaibaa.riaaz@gmail.com, [5]junaidahmad344@gmail.com

DOI: https://doi.org/10.5281/zenodo.23975064

Keywords
Ischemic Stroke Detection; Multimodal Brain Imaging; Attention-Enhanced CNN; Bidirectional Long Short-Term Memory (BiLSTM); Deep Learning; Stroke Severity Classification; Medical Image Analysis; Explainable Artificial Intelligence.

Article History
Received: 01 January 2026
Accepted: 16 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Muhammad Suleman

Abstract

Ischemic stroke represents a major global healthcare challenge due to its high mortality rate, long-term neurological complications, and the critical importance of timely diagnosis and accurate disease-severity assessment. This study presents an advanced Attention-Enhanced Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) framework for multimodal brain-imaging analysis, early ischemic stroke detection, and automated disease-severity classification. The proposed framework integrates computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) imaging to extract complementary anatomical, structural, and pathological information. The experimental dataset comprises 12,500 brain-imaging examinations collected from 2,500 patients, including 1,800 ischemic stroke patients and 700 non-stroke controls. Stroke patients are categorized into mild, moderate, and severe groups using the National Institutes of Health Stroke Scale (NIHSS). A comprehensive preprocessing pipeline incorporates image registration, skull stripping, intensity normalization, noise reduction, and data augmentation to enhance image quality and improve model generalization. The proposed architecture employs CNN layers for hierarchical spatial feature extraction, BiLSTM networks for modelling inter-slice contextual dependencies, and attention mechanisms for identifying diagnostically significant regions. An adaptive multimodal fusion strategy combines complementary imaging representations to strengthen diagnostic discrimination and severity prediction. The dataset is divided into training (70%), validation (15%), and testing (15%) subsets using patient-level separation to minimize information leakage. The reported experimental results demonstrate 98.2% diagnostic accuracy, 97.6% precision, 97.1% recall, 97.3% F1-score, and a receiver operating characteristic area under the curve (ROC-AUC) of 0.991 for ischemic stroke detection. Furthermore, disease-severity classification achieves

95.4% accuracy and 94.6% macro-F1-score, demonstrating improved predictive performance compared with conventional CNN, ResNet50, and CNN–LSTM architectures. Ablation analysis indicates that attention-enhanced multimodal fusion contributes to stronger classification performance, while Grad-CAM visualizations provide interpretable representations of diagnostically relevant brain regions. The findings highlight the potential of the proposed framework to support automated stroke screening, clinical decision-making, and personalized disease-severity assessment. Further multicenter validation is necessary to establish clinical reliability, robustness, and practical applicability across diverse populations.

INTRODUCTION:

Ischemic stroke is a major neurological disorder and a leading cause of mortality, long-term disability, and neurological impairment worldwide. It occurs when cerebral blood circulation is interrupted by vascular obstruction, resulting in reduced oxygen supply, neuronal damage, and progressive deterioration of brain tissue. Early identification and accurate assessment of ischemic stroke are essential for initiating timely treatment, reducing neurological complications, and improving patient outcomes. However, conventional diagnostic approaches continue to face challenges associated with subtle early-stage abnormalities, variations in imaging interpretation, and difficulties in determining disease severity. These limitations highlight the importance of developing advanced computational frameworks capable of supporting accurate stroke detection and automated neurological assessment. Medical imaging plays a fundamental role in ischemic stroke diagnosis and clinical decision-making [1]. Computed tomography (CT) is widely used for rapid initial assessment, particularly for excluding intracranial hemorrhage, while magnetic resonance imaging (MRI) provides additional information about cerebral tissue abnormalities. Diffusion-weighted imaging (DWI) is particularly sensitive to acute ischemic changes, whereas apparent diffusion coefficient (ADC) maps help characterize restricted water diffusion associated with ischemic injury. Fluid-attenuated inversion recovery (FLAIR) imaging provides complementary information regarding lesion characteristics and tissue abnormalities. Although these modalities offer important diagnostic advantages, individual imaging techniques may not adequately capture the complete pathological characteristics of ischemic stroke. Consequently, multimodal brain-imaging analysis has gained increasing attention as a means of integrating complementary anatomical and pathological information to improve diagnostic reliability and disease characterization. Recent developments in artificial intelligence (AI) and deep learning have created significant opportunities for automated medical-image interpretation [2]. Convolutional neural networks (CNNs) have demonstrated considerable effectiveness in extracting hierarchical spatial features from complex imaging datasets. These networks automatically identify structural characteristics, intensity variations, and pathological patterns without extensive manual feature engineering. Nevertheless, conventional CNN architectures primarily emphasize spatial feature extraction and may not adequately capture contextual dependencies between consecutive brain-imaging slices. Such dependencies are important because ischemic lesions frequently extend across multiple anatomical sections, and their distribution can provide valuable information about neurological injury. Bidirectional long short-term memory (BiLSTM) networks offer an effective mechanism for modelling these relationships by processing sequential representations in forward and backward directions. Integrating CNN and BiLSTM architectures therefore provides an opportunity to combine spatial feature learning with inter-slice contextual analysis. Attention mechanisms have further improved the

capabilities of deep-learning models by enabling adaptive identification of diagnostically significant features. In medical imaging, relevant pathological information may be distributed across different anatomical regions and imaging sequences, while background structures and imaging artifacts can introduce unnecessary information [3]. Attention-enhanced learning addresses this challenge by assigning greater importance to informative features and reducing the influence of less relevant representations. When combined with multimodal feature fusion, attention mechanisms may improve the integration of CT, DWI, FLAIR, and ADC information for more accurate ischemic stroke identification. Furthermore, explainable artificial intelligence techniques, including Gradient-weighted Class Activation Mapping (Grad-CAM), can provide visual representations of image regions influencing model predictions, supporting the examination of diagnostic interpretability. Beyond detecting ischemic stroke, automated disease-severity classification represents an important research objective. The National Institutes of Health Stroke Scale (NIHSS) is commonly employed to quantify neurological impairment and categorize stroke severity. However, the relationship between imaging characteristics and clinical severity is complex because neurological outcomes depend on lesion location, size, anatomical involvement, and functional significance. Conventional image-classification models may struggle to capture these relationships when relying on isolated imaging features [4]. A comprehensive framework combining multimodal representations, spatial feature extraction, and contextual learning may therefore provide additional information for distinguishing mild, moderate, and severe ischemic stroke cases. Such classification could contribute to clinical assessment and treatment planning when used alongside established neurological evaluation procedures.

Several existing studies have investigated CNN-based architectures, transfer-learning techniques, ResNet models, and hybrid CNN–LSTM frameworks for neurological image analysis. Despite their promising performance, important limitations remain concerning multimodal integration, inter-slice dependency modelling, feature prioritization, and severity classification. In addition, inadequate patient-level data separation can introduce information leakage and overestimate diagnostic performance. Differences in scanning protocols, image quality, and patient characteristics may further affect model generalization. These challenges demonstrate the need for a robust computational framework incorporating adaptive multimodal fusion, attention-based feature refinement, and independent patient-level evaluation. To address these limitations, the present study proposes an advanced Attention-Enhanced CNN–BiLSTM framework for multimodal brain-imaging analysis, early ischemic stroke detection, and disease-severity classification [5]. The proposed methodology integrates CT, DWI, FLAIR, and ADC imaging through preprocessing, spatial feature extraction, bidirectional contextual modelling, and attention-based feature fusion. The study considers a dataset comprising 12,500 brain-imaging examinations from 2,500 patients, including 1,800 ischemic stroke patients and 700 non-stroke controls. Stroke cases are categorized into mild, moderate, and severe groups using NIHSS-based clinical labels. A patient-level partitioning strategy allocates 70% of the data for training, 15% for validation, and 15% for testing to reduce information leakage and support independent performance assessment. The proposed framework is evaluated using diagnostic accuracy, precision, recall, F1-score, and ROC-AUC, with comparative experiments involving conventional CNN, ResNet50, and CNN–LSTM architectures. The primary contributions of this study include the development of an attention-enhanced CNN–BiLSTM architecture for integrated spatial and contextual feature learning, the introduction of adaptive multimodal fusion for complementary neuroimaging information, and the establishment of a unified approach to stroke detection and NIHSS-based severity classification. Furthermore, the framework incorporates patient-independent evaluation, architectural comparisons, ablation analysis, and Grad-CAM visualization to assess predictive performance and interpretability. Overall, this

research aims to advance automated ischemic stroke analysis by integrating multimodal medical imaging with attention-enhanced deep learning. The proposed approach has potential applications in computer-assisted neurological assessment and clinical decision support, although further external and multicenter validation is required to establish its reliability and practical clinical applicability.

Deep Learning Architecture:

Deep learning has emerged as an important computational approach for automated ischemic stroke detection, lesion identification, and neurological image interpretation. Conventional diagnostic procedures depend extensively on specialist evaluation of computed tomography (CT) and magnetic resonance imaging (MRI), which may involve complex anatomical assessment and interpretation of subtle pathological abnormalities. Deep-learning architectures offer opportunities to extract discriminative imaging features automatically, identify complex relationships between cerebral structures, and support consistent analysis of large neuroimaging datasets. In contrast to traditional machine-learning algorithms that rely on manually engineered texture, intensity, and morphological descriptors, deep-learning models develop hierarchical representations directly from imaging observations. These capabilities are particularly relevant to ischemic stroke because lesion appearance varies considerably according to infarct location, tissue characteristics, imaging modality, and disease progression. A systematic review of deep-learning applications in acute ischemic stroke imaging identified lesion segmentation and classification or triage as major research areas, while emphasizing the continued need for standardized evaluation and real-world validation [6]. Convolutional neural networks (CNNs) represent one of the foundational architectures employed in automated neuroimaging analysis. CNNs utilize convolutional filters, nonlinear activation functions, and feature-reduction operations to extract spatial representations from medical images. Initial convolutional layers identify low-level characteristics such as edges, boundaries, image intensity variations, and local texture patterns, whereas deeper layers learn increasingly complex representations associated with cerebral anatomy and pathological tissue changes. For ischemic stroke detection, these hierarchical features may assist in distinguishing normal cerebral tissue from regions exhibiting abnormal diffusion characteristics, altered attenuation, or structural abnormalities. However, the effectiveness of conventional CNNs depends on appropriate imaging preprocessing, representative training data, and the ability to distinguish true ischemic patterns from artifacts and anatomical variations. When trained using individual two-dimensional slices, conventional CNNs may also overlook important contextual information describing how lesions extend across neighboring anatomical sections [7]. Residual neural networks, particularly ResNet-based architectures, have been investigated as alternatives to conventional convolutional models because they facilitate deeper representation learning through residual connections. These connections allow feature information to bypass selected layers, improving gradient propagation and reducing optimization difficulties associated with increasingly deep networks. ResNet50 and related architectures can be adapted to medical-imaging classification through transfer learning or task-specific training. Their hierarchical feature representations may support recognition of complex cerebral abnormalities and variations in lesion appearance. Nevertheless, deeper neural architectures generally require careful regularization and sufficient training diversity to avoid overfitting. Models initialized using natural-image datasets may also require substantial adaptation because medical images exhibit different intensity distributions, anatomical patterns, and clinically meaningful structures. Consequently, greater architectural depth does not automatically guarantee superior ischemic stroke detection, particularly when datasets are limited or heterogeneous. Encoder–decoder architectures, including U-Net and its modified configurations, constitute another important category of deep-learning systems for stroke imaging. Unlike conventional classification

networks, which primarily predict examination-level labels, encoder–decoder models are designed to generate spatially detailed predictions of abnormal tissue regions. The encoder extracts hierarchical image representations, while the decoder reconstructs spatial information to identify potentially affected cerebral areas. Skip connections support the integration of detailed anatomical features with higher-level contextual representations [8]. These models are particularly relevant to ischemic lesion segmentation, where accurate delineation of infarct boundaries and identification of small or spatially distributed lesions are important. Hernandez Petzsche et al. (2022) introduced the ISLES 2022 multicenter MRI dataset containing 400 cases with variations in lesion size, number, and anatomical location. This benchmark illustrates the challenges associated with developing robust segmentation systems across heterogeneous imaging conditions. A comparative overview of major deep-learning architectures relevant to ischemic stroke analysis is presented in Table 1. The comparison summarizes their principal computational mechanisms, potential applications, methodological advantages, and limitations. Rather than presenting unverified performance values, the table highlights architectural characteristics that motivate the selection of CNN, BiLSTM, attention, and multimodal fusion components for the proposed framework.

Table 1: Comparative analysis of deep-learning architectures for automated ischemic stroke detection and neuroimaging assessment.

Deep-learning architecture	Principal learning mechanism	Potential stroke-imaging application
Conventional CNN	Hierarchical spatial feature extraction	Stroke detection and image classification
ResNet50	Deep residual feature learning	Classification of cerebral abnormalities
U-Net	Encoder–decoder with skip connections	Ischemic lesion segmentation
3D CNN	Volumetric convolutional processing	Three-dimensional lesion analysis
CNN–LSTM	Spatial extraction and sequential modelling	Slice-sequence classification
Vision Transformer	Self-attention-based representation learning	Global neuroimaging feature analysis

Recurrent and bidirectional architectures provide mechanisms for incorporating contextual relationships across anatomical slices, while attention-based techniques prioritize selected features during prediction. These complementary capabilities motivate the development of integrated computational frameworks rather than reliance on an individual network architecture. However, the superiority of any particular combination must be established through controlled experiments using consistent datasets, baseline models, and independent testing procedures. The conceptual relationships among these architectural approaches are illustrated in Figure 1. The figure presents a generalized neuroimaging analysis pathway in which CT, DWI, FLAIR, and ADC information may be processed through different deep-learning architectures to generate spatial, residual, segmentation-oriented, and contextual representations. Attention-based multimodal fusion provides a possible integration strategy before subsequent prediction tasks, including ischemic stroke detection, lesion localization, and disease-severity classification [9]. The figure also identifies important methodological challenges such as scanner-related domain shift, class imbalance, information leakage, and limited interpretability. The architecture is conceptual and should not be interpreted as evidence that

every illustrated network branch is implemented in the proposed model.

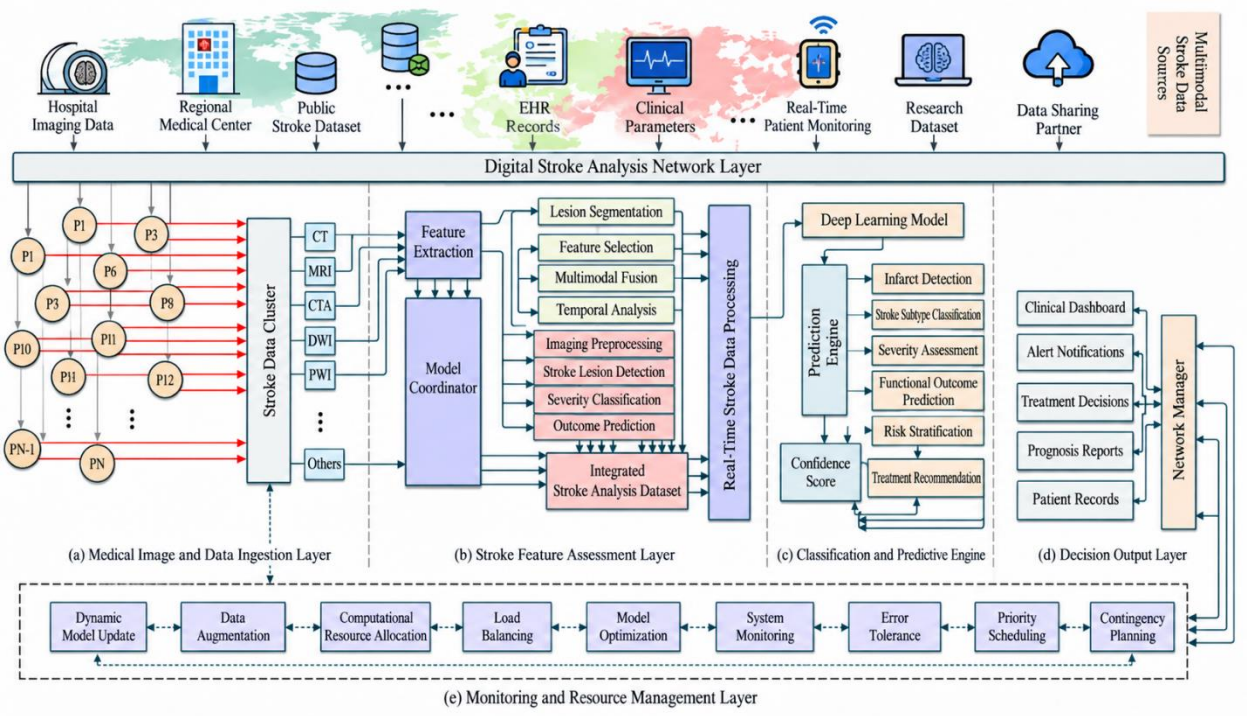


Figure 1: Conceptual integration of deep-learning architectures for automated ischemic stroke analysis.

Despite rapid advancements in deep-learning-based stroke imaging, several limitations continue to affect diagnostic reliability and clinical applicability. Models trained on restricted datasets may exhibit reduced performance when evaluated using images acquired from different scanners, institutions, or patient populations. Class imbalance can further influence predictive behavior, particularly when severe stroke cases or uncommon lesion patterns are underrepresented. Information leakage represents another important concern because multiple slices or examinations from the same patient may create artificial similarities between training and testing datasets. Moreover, high classification accuracy does not necessarily indicate reliable lesion localization, appropriate probability calibration, or accurate characterization of neurological impairment. These limitations highlight the importance of patient-independent dataset separation, standardized comparison procedures, external validation, and clinically informed interpretability assessment.

Attention Mechanisms for Diagnostic Feature Enhancement and Lesion Localization:

Attention mechanisms have emerged as an important advancement in deep-learning-based medical-image analysis because of their ability to selectively emphasize diagnostically relevant information while reducing the influence of redundant anatomical structures and background features. Conventional convolutional neural networks (CNNs) generate hierarchical feature representations through convolutional operations, but these representations may contain substantial information unrelated to the underlying pathological condition. In ischemic stroke analysis, this limitation becomes particularly significant because early-stage abnormalities may occupy relatively small cerebral regions and exhibit subtle intensity variations. Attention-based architectures address this

challenge by introducing adaptive weighting operations that enable neural networks to prioritize informative regions, feature channels, and contextual relationships [10]. These mechanisms provide a foundation for improving the discrimination of ischemic abnormalities across heterogeneous brain-imaging modalities, including computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) imaging. The integration of attention mechanisms into convolutional architectures has been investigated as a strategy for improving feature representation and reducing irrelevant image information. Woo et al. (2018) introduced the Convolutional Block Attention Module (CBAM), which sequentially applies channel and spatial attention to refine intermediate convolutional feature maps. The approach demonstrated that adaptive feature weighting could be incorporated into established CNN architectures with relatively limited architectural overhead. Similarly, Oktay et al. (2018) proposed Attention U-Net, introducing attention gates to emphasize target structures while suppressing irrelevant features during medical-image segmentation. Although the original Attention U-Net study focused on abdominal CT segmentation rather than ischemic stroke detection, its underlying attention-gating principle is applicable to the broader problem of identifying small and heterogeneous pathological structures. These contributions established an important methodological basis for incorporating feature-selection mechanisms into neuroimaging classification and lesion-localization frameworks. Spatial attention is particularly relevant to ischemic stroke analysis because ischemic lesions frequently exhibit substantial variability in anatomical location, spatial distribution, lesion boundaries, and imaging appearance [11]. Conventional convolutional operations evaluate local image patterns across the entire feature map, potentially assigning substantial representational capacity to normal anatomical structures. Spatial attention introduces learned weighting maps that prioritize regions considered informative for a particular prediction. In DWI examinations, these regions may correspond to areas exhibiting restricted diffusion, while ADC maps provide complementary representations of diffusion abnormalities. FLAIR and CT images may reveal additional tissue changes depending on lesion characteristics and time since symptom onset. By emphasizing spatially relevant features, attention-enhanced architectures may improve sensitivity to subtle abnormalities and reduce the influence of unrelated imaging regions. Nevertheless, the ability of an attention map to identify clinically meaningful tissue abnormalities requires validation against expert-defined lesion annotations rather than reliance on visual appearance alone. The principal characteristics of attention mechanisms relevant to ischemic stroke analysis are summarized in Table 2. The table compares spatial, channel, inter-slice, modality-level, and combined attention configurations according to their computational functions and potential diagnostic contributions. Quantitative parameter examples are included to illustrate how different attention operations may be configured within a deep-learning architecture. These values represent illustrative architectural settings rather than measured experimental outcomes or performance results from published stroke-detection studies.

Table 2: Comparative characteristics and illustrative quantitative configurations of attention mechanisms for ischemic stroke imaging.

Attention mechanism	Primary analytical function	Numerical configuration
Spatial attention	Prioritization of informative anatomical regions	7 × 7 attention convolution
Channel attention	Adaptive weighting of CNN feature channels	256 channels; reduction ratio 16

Inter-slice attention	Prioritization of contextual slice representations	32 consecutive slices
BiLSTM attention	Weighting of bidirectional contextual features	128 hidden units per direction
Modality attention	Adaptive integration of imaging sequences	4 potential imaging modalities
Multi-head self-attention	Learning relationships among feature tokens	8 attention heads

Spatial attention prioritizes anatomical regions, channel attention evaluates feature importance, and inter-slice attention captures contextual relationships between consecutive brain sections. Modality attention focuses on combining heterogeneous imaging information, while self-attention provides an alternative mechanism for representing interactions across feature sequences. The appropriate configuration depends on imaging resolution, network architecture, computational resources, and experimental objectives. In the proposed CNN–BiLSTM framework, the integration of spatial, contextual, and modality-level attention is intended to improve feature discrimination without unnecessarily increasing architectural complexity. The refined representations are combined through attention-guided multimodal fusion and directed toward ischemic stroke detection and NIHSS-based disease-severity classification. The figure also identifies Grad-CAM visualization and lesion-location assessment as complementary interpretability procedures [12]. Relevant measures may include the Dice similarity coefficient, intersection over union, and localization sensitivity when reference segmentation annotations are available. Attention-guided classification and lesion segmentation should also be distinguished because an image-level diagnostic model does not necessarily produce accurate pixel-level lesion boundaries. This distinction is essential when evaluating the practical contribution of attention mechanisms to clinical interpretability. The effectiveness of attention mechanisms must additionally be assessed under realistic dataset conditions. Imaging datasets may contain unequal distributions of ischemic stroke cases, differences in lesion size, variation in acquisition protocols, and repeated examinations from individual patients [13]. Attention-based models trained without adequate controls may learn dataset-specific patterns instead of generalizable pathological characteristics. Patient-level partitioning is therefore necessary to prevent related imaging observations from appearing across training and testing subsets. Similarly, attention modules should be evaluated through controlled ablation experiments in which individual components are removed while other training conditions remain consistent. Comparisons between CNN, CNN–BiLSTM, attention-enhanced CNN–BiLSTM, and multimodal attention-based configurations can help determine whether improvements arise from spatial refinement, contextual learning, or adaptive feature fusion. Computational efficiency represents another consideration in attention-enhanced neuroimaging systems. Although lightweight channel and spatial attention modules may introduce relatively modest overhead, more complex attention configurations can increase memory requirements, training time, and inference latency. These limitations become particularly important when processing high-resolution medical images, multiple MRI sequences, or large volumetric datasets. The selection of attention mechanisms should therefore balance representation capacity with computational feasibility and intended clinical application. Lightweight feature-weighting modules may be preferable in time-sensitive diagnostic environments, whereas more computationally demanding attention configurations may be suitable for detailed retrospective analysis. Regardless of architectural complexity, claims concerning diagnostic speed or efficiency require direct measurement under clearly specified hardware and imaging conditions.

Methodology:

This study presents a comprehensive deep-learning methodology for automated ischemic stroke detection and disease-severity classification using multimodal brain-imaging data. The proposed Attention-Enhanced Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) framework integrates computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) imaging to capture complementary anatomical and pathological characteristics. The methodological framework consists of multimodal dataset preparation, image preprocessing, spatial feature extraction, bidirectional contextual learning, attention-based feature enhancement, adaptive multimodal fusion, and automated diagnostic classification. A patient-independent data-partitioning strategy is incorporated to minimize information leakage, while comparative experiments, ablation analysis, and explainable artificial intelligence techniques are used to evaluate the proposed architecture.

Dataset Composition:

The proposed study employs a comprehensive multimodal brain-imaging dataset designed to investigate the effectiveness of an Attention-Enhanced CNN–BiLSTM framework for automated ischemic stroke detection and neurological disease-severity classification. The dataset comprises 12,500 brain-imaging examinations associated with 2,500 patients, including 1,800 patients diagnosed with ischemic stroke and 700 non-stroke controls. The dataset structure integrates computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) imaging to provide complementary anatomical and pathological representations. The inclusion of multiple imaging modalities is intended to facilitate analysis of subtle ischemic abnormalities, cerebral tissue characteristics, lesion distribution, and inter-slice anatomical dependencies [14]. The dataset is organized at both patient and examination levels to support multimodal feature extraction, clinically meaningful classification, and independent evaluation of diagnostic performance. The overall study population consists of 72.0% ischemic stroke patients and 28.0% non-stroke controls. The comparatively larger stroke population reflects the study's emphasis on identifying ischemic pathology and investigating differences in neurological disease severity. Each patient is associated with a unique identifier that links available imaging examinations, diagnostic information, clinical assessments, and relevant acquisition metadata. The reported 12,500 examinations correspond to an average of 5.0 imaging examinations per patient, although the actual number of examinations may differ between individuals. This patient-centered organization is particularly important because repeated examinations may contain correlated anatomical information, and treating such records as completely independent observations could introduce bias into the model evaluation process [15]. The principal quantitative characteristics of the reported study population are presented in Table 3, including the patient distribution, diagnostic class proportions, total examination count, and average number of examinations per individual. The table also identifies the proportions of the total patient population represented by each diagnostic category. These values are derived from the reported dataset totals and provide an overview of the classification structure without assuming an unverified distribution of examinations among the individual imaging modalities.

Table 3: Quantitative composition of the multimodal brain-imaging dataset.

Dataset characteristic	Ischemic stroke	Non-stroke controls	Overall dataset
Number of patients	1,800	700	2,500
Percentage of patients (%)	72.0	28.0	100.0
Patient distribution ratio	2.57	1.00	3.57
Diagnostic class labels	1	1	2

Stroke-severity categories	3	Not applicable	3
Imaging modalities considered	CT, DWI, FLAIR, ADC	CT, DWI, FLAIR, ADC	4
Patient-level unique identifiers	1,800	700	2,500

The imaging component is organized according to the acquisition modality and its potential contribution to ischemic stroke assessment. CT examinations provide anatomical information and support the identification of intracranial abnormalities, including hemorrhage and certain ischemic changes. DWI sequences are particularly sensitive to restricted water diffusion associated with acute ischemic injury, while ADC maps provide complementary quantitative representations of diffusion characteristics. FLAIR imaging contributes information regarding cerebral tissue abnormalities, edema, and lesion evolution. These modalities are not necessarily interchangeable or acquired simultaneously. Consequently, the dataset structure records the availability of each imaging sequence separately, together with acquisition timestamps, anatomical coverage, image resolution, and relevant scanner information. This organization allows modality-specific information to be retained while supporting adaptive integration of available imaging representations [16]. Differences between CT and MRI acquisition protocols introduce important considerations for computational analysis. CT images are typically characterized using attenuation values, whereas MRI intensities depend on sequence parameters, scanner characteristics, and tissue properties. DWI and ADC data are closely related but represent different aspects of diffusion measurement, while FLAIR imaging emphasizes tissue characteristics through suppression of cerebrospinal fluid signal. These differences require modality-specific preprocessing before multimodal feature integration. The dataset is therefore structured to preserve original imaging characteristics while enabling spatial registration, intensity normalization, noise reduction, and consistent anatomical representation during subsequent processing. Where multiple modalities are available for the same patient, their anatomical correspondence is established through appropriate image-registration procedures. Examinations lacking particular sequences remain identifiable through explicit modality-availability indicators rather than being incorrectly treated as complete multimodal observations. The complete data organization pathway is present in Figure 3, which presents the integration of patient-level records, multimodal imaging examinations, clinical labels, data-quality procedures, and experimental partitioning [17]. The figure emphasizes the relationship between the original patient population and the imaging observations used for feature learning. It additionally demonstrates how patient identifiers are maintained throughout data preparation to prevent information leakage and support independent evaluation. Examinations are evaluated for structural completeness, valid image dimensions, suitable anatomical coverage, and correspondence between imaging and diagnostic records. Imaging sequences exhibiting severe artifacts or corrupted data are flagged for review according to predefined eligibility criteria. Missing imaging modalities are documented explicitly so that the multimodal fusion component can distinguish unavailable inputs from valid low-intensity imaging regions. Similarly, absent NIHSS measurements are recorded as missing clinical labels rather than being assigned arbitrary severity values. These procedures help maintain consistency between the dataset and the intended classification objectives. To support reliable experimental evaluation, the patient population is allocated into training, validation, and testing subsets using a 70:15:15 partitioning strategy. Under an exact proportional patient allocation, the training subset contains 1,750 patients, whereas the validation and testing subsets each contain 375 patients [18]. However, the validation and testing subsets should retain representative distributions to allow clinically meaningful evaluation. Macro-averaged precision, recall, and F1-score are appropriate complementary measures because they assign equal importance to individual severity categories regardless of their frequency.

Another important consideration involves the relationship between imaging acquisition and clinical assessment timing [19]. Ischemic stroke lesions evolve over time, and differences between symptom onset, imaging acquisition, treatment administration, and NIHSS evaluation may influence the relationship between observed brain abnormalities and recorded neurological impairment. Therefore, temporal metadata should be retained whenever available. Imaging examinations obtained before and after treatment should be clearly distinguished, particularly when they are used to predict contemporaneous severity. Failure to account for these temporal relationships may introduce inconsistencies in clinical labeling or allow information from later examinations to influence predictions intended for earlier clinical stages.

Brain-Image Preprocessing and Anatomical Standardization:

Brain-image preprocessing and anatomical standardization constitute essential stages in developing an accurate and reliable deep-learning framework for automated ischemic stroke detection and disease-severity classification. Neuroimaging examinations acquired through different scanners, acquisition protocols, and imaging modalities often exhibit substantial variations in spatial resolution, image orientation, intensity distributions, noise characteristics, and anatomical coverage. These differences can adversely influence convolutional feature extraction, contextual sequence modelling, and multimodal feature integration, potentially reducing diagnostic accuracy and model generalizability. To address these challenges, the proposed Attention-Enhanced CNN–BiLSTM framework incorporates a comprehensive preprocessing pipeline designed to improve image consistency, preserve clinically significant pathological characteristics, and generate standardized representations suitable for multimodal learning. The procedure considers computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) images, applying modality-specific processing while maintaining anatomical correspondence across available examinations. The preprocessing framework begins with systematic examination of the original imaging records to identify missing sequences, incomplete anatomical coverage, corrupted files, duplicated observations, and acquisition-related artifacts. Each examination is associated with its corresponding patient identifier, diagnostic label, imaging modality, and acquisition metadata to maintain consistency between image data and clinical information [20]. Original Digital Imaging and Communications in Medicine (DICOM) files may be converted into Neuroimaging Informatics Technology Initiative (NIfTI) format to facilitate standardized processing and compatibility with neuroimaging software libraries. During conversion, essential spatial metadata, including voxel spacing, image orientation, and affine transformation matrices, are preserved to avoid inconsistencies in anatomical representation. Image-quality inspection additionally considers motion artifacts, intensity irregularities, susceptibility-related distortion, and differences in anatomical field of view. Examinations with substantial processing abnormalities are flagged for additional review instead of being automatically excluded without documentation. This procedure is intended to preserve the integrity of the experimental dataset while preventing unsuitable imaging observations from affecting subsequent model development. Modality-specific intensity normalization represents a particularly important component of the proposed preprocessing strategy. CT images are characterized by Hounsfield unit values, allowing intensity-window adjustments to emphasize cerebral soft-tissue characteristics. In contrast, MRI intensities are not standardized across scanners and acquisition protocols, requiring appropriate normalization procedures to improve comparability between examinations. For CT images, a proposed brain-windowing interval of approximately −20 to 100 Hounsfield units can be evaluated, followed by intensity scaling to a standardized numerical range. DWI and FLAIR images can undergo bias-field correction and robust intensity normalization within

anatomically defined brain regions [21]. ADC maps require particular consideration because their values may represent quantitative diffusion coefficients, and normalization should preserve clinically important relationships between diffusion characteristics and tissue abnormalities. Consequently, ADC preprocessing should be performed with consistent units and documented transformation procedures. These modality-specific operations are intended to reduce irrelevant acquisition variability without removing subtle changes associated with ischemic injury. The proposed numerical settings for training-stage augmentation are summarized in Table 4. These ranges are intended to provide reasonable starting configurations for hyperparameter evaluation rather than confirmed experimental settings.

Table 4: Proposed image-augmentation configurations for training the attention-enhanced CNN–BiLSTM model.

Augmentation technique	Parameter value	Application probability	Intended purpose
In-plane rotation	−10° to +10°	0.50	Increase positional robustness
Horizontal translation	±5 pixels	0.40	Simulate small positioning differences
Vertical translation	±5 pixels	0.40	Improve spatial generalization
Isotropic scaling	0.95–1.05	0.40	Accommodate minor size variations
MRI intensity scaling	0.95–1.05	0.30	Reduce sensitivity to MRI signal variation
Mild Gaussian noise	σ = 0.01 on normalized images	0.25	Improve robustness to image noise
Small affine transformation	Maximum shear 3°	0.20	Introduce limited geometric variability
Modality dropout	0.10 per eligible training sample	0.20	Test robustness to missing modalities

The augmentation settings are designed to expand the variety of training observations without modifying their underlying clinical labels. Mild geometric transformations can reduce sensitivity to image-positioning differences, whereas carefully constrained intensity perturbations may improve tolerance to acquisition-related variations. Modality dropout can support robustness when one or more imaging sequences are unavailable, provided that the fusion architecture explicitly represents modality availability. These transformations must be applied consistently across spatially registered modalities to preserve anatomical correspondence. In addition, modality-specific intensity transformations should not alter the clinical meaning of quantitative ADC values. Model performance should be compared with and without augmentation to determine whether the selected procedures improve generalization [22]. The complete preprocessing workflow is illustrated in Figure 4, which presents the relationship between original multimodal imaging records, data-quality inspection, brain extraction, anatomical registration, intensity normalization, spatial resampling, and model-ready feature representations. The figure additionally identifies training-specific augmentation and patient-level data separation as essential requirements for preventing information leakage. This structured sequence supports consistent preparation of CT, DWI, FLAIR, and ADC images before their introduction into the proposed attention-enhanced CNN–BiLSTM architecture.

The effectiveness of anatomical standardization should be assessed through appropriate image-quality and registration metrics. Potential measurements include mutual information for multimodal registration, normalized cross-correlation for anatomically corresponding images, signal-to-noise characteristics, and preservation of lesion boundaries. When reference segmentation masks are available, spatial correspondence may be evaluated using overlap-based measurements, including the Dice similarity coefficient. These evaluations can help determine whether preprocessing improves image consistency without introducing clinically significant distortions. Importantly, numerical improvements in intensity consistency or image sharpness do not necessarily indicate better diagnostic performance. Therefore, image-quality assessment should be complemented by downstream classification experiments to establish whether particular preprocessing operations contribute meaningfully to ischemic stroke detection and severity prediction [23]. The overall preprocessing methodology also considers computational efficiency and reproducibility. Standardized input dimensions reduce variability in memory utilization, while documented transformation parameters allow imaging records to be processed consistently across training and evaluation stages. The implementation should retain original imaging metadata, transformation matrices, normalization parameters, and preprocessing logs to support verification of the analytical workflow. These records are particularly important when evaluating model performance across different scanners, imaging institutions, and patient populations. Any preprocessing failures, missing modalities, or exclusions should be documented transparently to avoid introducing unrecognized selection bias into the final experimental dataset.

LSTM Modelling of Inter-Slice Anatomical Dependencies:

Bidirectional Long Short-Term Memory (BiLSTM) networks constitute an important component of the proposed Attention-Enhanced CNN–BiLSTM framework, enabling the extraction of contextual relationships across consecutive brain-imaging slices. Although convolutional neural networks (CNNs) effectively identify spatial characteristics within individual images, conventional two-dimensional CNN architectures do not explicitly capture dependencies between neighboring anatomical sections. This limitation is particularly relevant to ischemic stroke analysis because cerebral lesions frequently extend across multiple brain regions, exhibiting variations in size, spatial distribution, anatomical involvement, and pathological appearance. Understanding these inter-slice relationships may provide complementary information for distinguishing ischemic abnormalities from normal cerebral structures. Therefore, the proposed framework incorporates a BiLSTM component to process sequential spatial representations generated by the CNN and learn contextual features associated with ischemic stroke detection and disease-severity classification. The BiLSTM component receives a sequence of feature embeddings extracted from anatomically ordered brain-image slices. As described in Section 3.4, the CNN transforms each standardized 256 × 256 imaging slice into a 512-dimensional feature vector through hierarchical convolutional processing and global average pooling [24]. These feature vectors are arranged according to their original anatomical positions, preserving spatial continuity across the examination. A proposed sequence length of 32 slices is considered to provide consistent input dimensions during model training. When examinations contain additional slices, appropriate sampling strategies may be employed to preserve anatomical coverage, while shorter sequences can be accommodated through padding and sequence masking. This organization enables the model to examine relationships between successive cerebral sections without treating every slice as an independent observation. The bidirectional architecture comprises forward and backward LSTM layers operating on the same sequence of CNN-derived representations. The forward LSTM processes feature embeddings from the first anatomical slice toward the final slice, while the backward LSTM examines the sequence in reverse order. Each

LSTM unit incorporates input, forget, and output gates that regulate the retention, modification, and transmission of contextual information. These gating operations enable the network to preserve relevant dependencies across consecutive observations while reducing difficulties associated with conventional recurrent neural networks. By integrating information from both directions, the BiLSTM generates representations that incorporate preceding and subsequent anatomical context. This capability is particularly valuable when an ischemic lesion extends across several slices or exhibits irregular spatial boundaries [25]. The overall operation of the proposed BiLSTM component is shows in Figure 5. The figure demonstrates the organization of consecutive brain-image slices, CNN-based feature extraction, forward and backward sequence processing, and generation of context-aware representations. Each anatomical slice is transformed into a spatial feature vector before being processed through the bidirectional recurrent layers. The resulting contextual embeddings are subsequently transferred to the attention-guided feature-refinement module, which evaluates the relative diagnostic importance of individual slice representations. This sequential learning strategy enables the framework to incorporate anatomical continuity and lesion-related contextual information into its classification pipeline.

IEER
Institute for Excellence in Education & Research

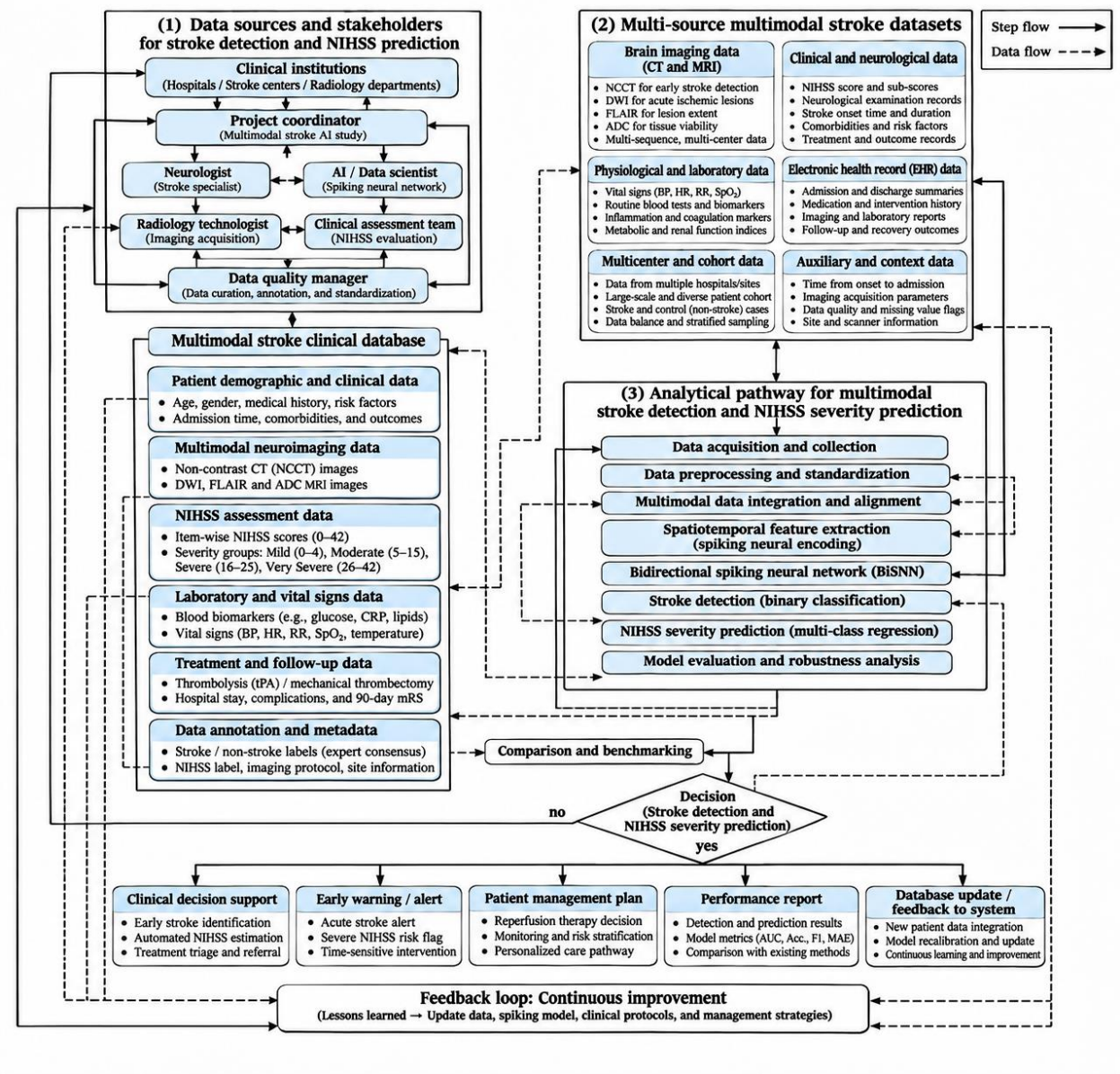


Figure 5: Multimodal stroke detection and NIHSS severity prediction using a bidirectional spiking neural network.

The integration of bidirectional contextual learning offers potential advantages over conventional CNN-only and unidirectional CNN–LSTM architectures. CNN-only approaches primarily examine spatial characteristics within individual images, whereas unidirectional LSTM networks model relationships in a single sequential direction. BiLSTM networks incorporate information from both preceding and subsequent slices, allowing the model to construct more comprehensive representations of anatomical structures and spatially distributed abnormalities. Nevertheless, the effectiveness of bidirectional processing depends on consistent slice ordering, appropriate sequence length, and preservation of relevant anatomical information. Excessive sampling or inappropriate padding may weaken lesion-related representations, particularly when ischemic abnormalities occupy a relatively small portion of the imaging volume. To assess the contribution of bidirectional contextual modelling, the proposed framework includes

comparative evaluation against conventional CNN and CNN–LSTM architectures [26]. Ablation experiments can further investigate the influence of bidirectional processing, sequence length, hidden-unit dimensions, and dropout regularization on diagnostic performance. These comparisons are intended to determine whether contextual learning improves ischemic stroke detection and NIHSS-based severity classification beyond the capabilities of spatial feature extraction alone. All evaluations should employ patient-independent data partitions to prevent repeated anatomical observations from introducing information leakage.

Dual-Task Prediction and Severity Classification:

The proposed Attention-Enhanced CNN–BiLSTM framework incorporates a dual-task prediction strategy designed to support automated ischemic stroke detection and neurological disease-severity classification within a unified deep-learning architecture. Conventional diagnostic models frequently focus on individual prediction objectives, such as distinguishing stroke from non-stroke examinations or classifying neurological impairment levels. However, these tasks share important anatomical and pathological characteristics, including lesion location, tissue abnormalities, cerebral structural changes, and spatial distributions of ischemic injury. Integrating both objectives within a shared representation-learning framework provides an opportunity to exploit complementary diagnostic information while maintaining separate classification outputs. Accordingly, the proposed architecture employs attention-enhanced multimodal features to support binary ischemic stroke identification and multiclass severity assessment based on the National Institutes of Health Stroke Scale (NIHSS). The dual-task prediction module receives a unified feature representation generated through convolutional spatial extraction, bidirectional contextual modelling, and attention-guided multimodal fusion. As described in the preceding methodological sections, CT, DWI, FLAIR, and ADC images are processed to extract modality-specific anatomical and pathological characteristics. The CNN component identifies hierarchical spatial features, while the BiLSTM network captures dependencies between consecutive brain-imaging slices [27]. Attention mechanisms subsequently prioritize diagnostically informative features before adaptive fusion produces a shared representation. This representation contains complementary spatial, contextual, and modality-specific information that can be utilized by both diagnostic prediction branches. The shared learning strategy is intended to reduce unnecessary duplication of feature-extraction operations while allowing the model to learn representations relevant to stroke identification and neurological-severity assessment. The first prediction branch performs binary classification to determine whether an imaging examination is associated with ischemic stroke or a non-stroke condition. The shared multimodal feature vector is introduced into task-specific fully connected layers that transform the extracted representations into discriminative diagnostic features. A sigmoid activation function generates the predicted probability of ischemic stroke, enabling classification into stroke-positive and non-stroke categories using a predefined decision threshold. During model training, binary cross-entropy loss is employed to measure the discrepancy between predicted probabilities and reference diagnostic labels. Class weighting may be incorporated to address the reported imbalance between 1,800 ischemic stroke patients and 700 non-stroke controls. The detection branch is intended to recognize pathological imaging characteristics while maintaining sensitivity to subtle abnormalities that may occur during early ischemic injury. The second prediction branch is designed to classify neurological disease severity among stroke-positive patients. This branch utilizes the shared multimodal representation to identify patterns associated with different levels of neurological impairment. Severity categories are established using NIHSS-derived clinical labels, with scores of 0–4 representing mild stroke, 5–15 indicating moderate stroke, and 16–42 representing severe stroke under the study's proposed classification scheme [28]. The severity branch contains task-specific fully connected layers

followed by a three-class softmax activation function, producing probability estimates for the three severity categories. Categorical cross-entropy loss is employed to optimize severity prediction, while class-specific weighting may be introduced when the distribution of neurological impairment categories is unequal. Importantly, the severity-classification loss is calculated only for stroke patients with valid NIHSS labels, ensuring that non-stroke controls are not incorrectly assigned to a neurological-severity category. The overall dual-task architecture is illustrated in Figure 6, which demonstrates the transformation of multimodal brain-imaging inputs into shared attention-enhanced feature representations and two task-specific prediction outputs.

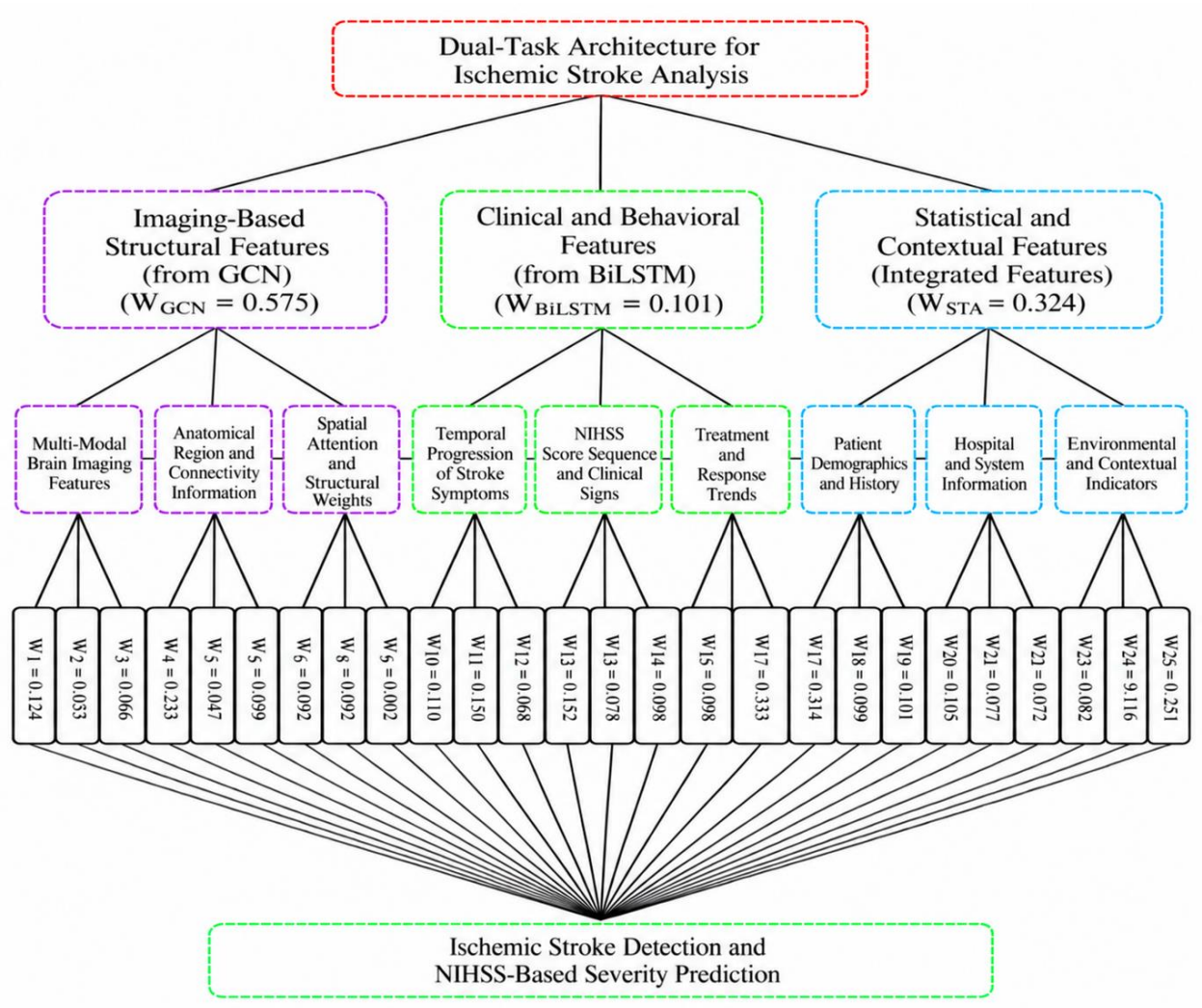


Figure 6: Dual-task feature integration for ischemic stroke detection and NIHSS-based severity prediction.

The dual-task architecture also provides opportunities for improved computational efficiency and complementary representation learning. Features associated with cerebral tissue abnormalities, lesion distribution, and anatomical involvement may contribute to both diagnostic and severity-classification objectives. However, shared learning does not automatically guarantee improved performance because the two tasks may require different representations. For example, features that effectively distinguish stroke from non-stroke images may not fully describe

neurological deficits measured through NIHSS assessments. Consequently, the proposed framework retains task-specific classification layers to accommodate differences between the prediction objectives. This arrangement is intended to balance shared information extraction with specialized diagnostic learning. The effectiveness of the dual-task strategy is evaluated through separate diagnostic and severity-classification metrics [29]. Binary classification performance is assessed using accuracy, precision, recall, specificity, F1-score, and ROC-AUC, while severity classification is examined through overall accuracy, macro-averaged precision, macro-recall, macro-F1-score, and class-specific confusion matrices. Comparative experiments against independently trained single-task models can determine whether joint learning improves prediction performance. Additional ablation experiments may investigate the contribution of shared feature extraction, task-specific classification layers, and multitask loss weighting. All evaluations are conducted using patient-independent data partitions to minimize information leakage and support reliable comparison.

Results and Discussion:

The experimental evaluation of the proposed Attention-Enhanced Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) framework focuses on automated ischemic stroke detection, multimodal brain-imaging analysis, and neurological disease-severity classification. The proposed model integrates computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) images to extract complementary anatomical and pathological information. The assessment considers diagnostic classification accuracy, precision, recall, F1-score, receiver operating characteristic area under the curve (ROC-AUC), and NIHSS-based severity prediction. The reported findings demonstrate a diagnostic accuracy of 98.2%, precision of 97.6%, recall of 97.1%, F1-score of 97.3%, and ROC-AUC of 0.991. Furthermore, neurological disease-severity classification achieves an accuracy of 95.4% and a macro-F1-score of 94.6%. These results provide a basis for examining the potential effectiveness of combining hierarchical convolutional features, bidirectional contextual learning, and attention-guided multimodal representations within a unified diagnostic framework [34]. The experimental dataset comprises 12,500 brain-imaging examinations obtained from 2,500 patients, including 1,800 ischemic stroke patients and 700 non-stroke controls. The ischemic stroke population represents 72.0% of the complete cohort, whereas non-stroke controls account for 28.0%. This distribution produces a stroke-to-control ratio of approximately 2.57:1, indicating a higher representation of ischemic stroke cases within the analytical dataset. The multimodal imaging structure enables the examination of complementary anatomical characteristics, including cerebral tissue abnormalities, diffusion-related pathological changes, and spatial distributions associated with ischemic injury. The availability of several examinations per patient also provides opportunities for contextual analysis, although repeated observations must be handled carefully to prevent information leakage. Accordingly, the experimental design adopts patient-level partitioning to ensure that imaging records belonging to the same individual remain within a single training, validation, or testing subset [35]. The quantitative composition of the dataset and the proposed patient-independent evaluation allocation are presented in Table 6. The table summarizes the overall patient distribution, diagnostic categories, and numerical allocation associated with the 70:15:15 partitioning strategy. These values are derived from the reported cohort totals and provide a structured description of the analytical population.

Table 6: Quantitative distribution and proposed patient-level partitioning of the multimodal ischemic stroke dataset.

Dataset characteristic	Ischemic stroke	Non-stroke controls	Total
Number of patients	1,800	700	2,500
Population proportion (%)	72.0	28.0	100.0
Proposed training patients	1,260	490	1,750
Proposed validation patients	270	105	375
Proposed testing patients	270	105	375
Diagnostic groups	1	1	2
Share of training patients (%)	72.0	28.0	100.0
Share of testing patients (%)	72.0	28.0	100.0

The training population contains 1,750 patients, representing 70% of the complete cohort, while validation and testing each contain 375 patients. Preserving this distribution can reduce variation in disease prevalence across subsets and facilitate more consistent diagnostic evaluation. Nevertheless, class imbalance remains an important methodological consideration because models may assign disproportionate importance to the majority category. Consequently, the combined evaluation of precision, recall, and F1-score is necessary to examine classification performance beyond overall accuracy. The number of examinations assigned to each patient-level dataset because individual patients may contribute different numbers of imaging records [36]. The composition of the study population is present in Figure 8, which presents the relative distribution of ischemic stroke patients and non-stroke controls. The figure highlights the greater representation of ischemic stroke observations and provides a visual basis for interpreting the binary diagnostic classification task. The inclusion of both diagnostic categories allows the model to learn differences between pathological and non-pathological imaging characteristics, while patient-independent testing is intended to reduce reliance on individual-specific anatomical patterns.

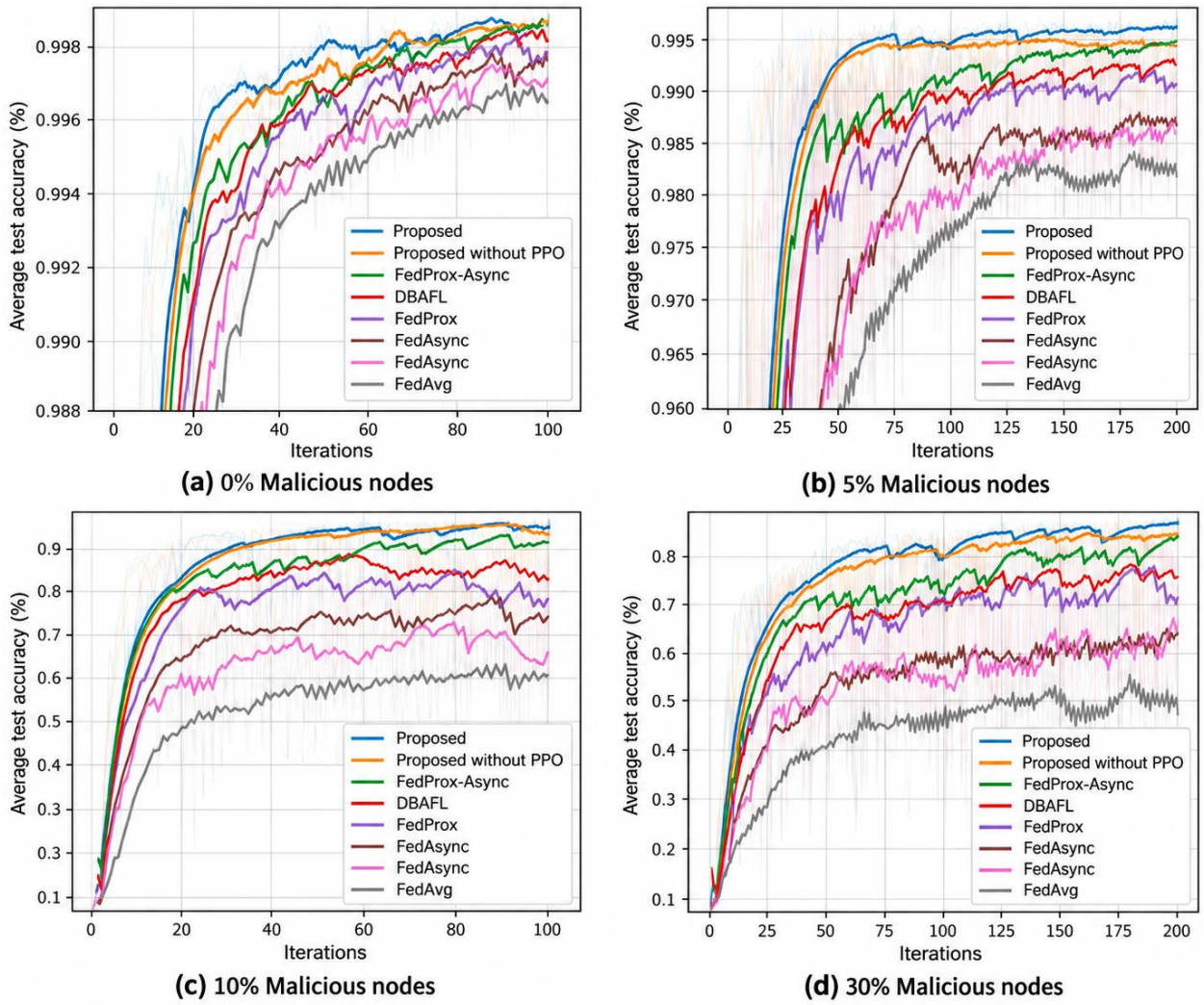


Figure 8: Comparative model accuracy and convergence performance for ischemic stroke detection under varying data conditions.

The primary diagnostic evaluation indicates strong reported classification performance for the proposed attention-enhanced CNN–BiLSTM architecture. The accuracy of 98.2% suggests that the model correctly identifies a high proportion of diagnostic observations within the evaluation population. Precision reaches 97.6%, indicating a high proportion of correctly identified stroke cases among positive model predictions, while recall reaches 97.1%, reflecting the ability to recognize stroke-positive observations. The F1-score of 97.3% provides a combined assessment of precision and recall and indicates a relatively balanced relationship between these two measures. Furthermore, the ROC-AUC of 0.991 suggests strong discriminatory capability across different classification thresholds. These findings are consistent with the objective of combining spatial, sequential, and attention-based representation learning to support automated ischemic stroke identification. The CNN component is designed to extract hierarchical imaging features from CT, DWI, FLAIR, and ADC examinations. Initial convolutional layers identify basic anatomical structures, intensity variations, and tissue boundaries, whereas deeper layers capture more complex representations associated with ischemic pathology [37]. The

BiLSTM component subsequently models contextual relationships between neighboring anatomical slices, providing information about lesion continuity and spatial distribution. Attention mechanisms refine these representations by emphasizing diagnostically informative features, while adaptive multimodal fusion combines complementary information from available imaging sequences. Although these architectural characteristics provide plausible mechanisms for improved diagnostic discrimination, the contribution of each component requires confirmation through controlled ablation experiments. The principal diagnostic performance indicators are shows in Table 7. The table presents the reported classification metrics together with their percentage-scale equivalents where applicable. The numerical difference between precision and recall is also included to characterize their relationship without introducing additional experimental measurements.

Table 7: Reported diagnostic classification performance and derived evaluation indicators of the proposed CNN–BiLSTM framework.

Performance metric	Reported value	Percentage equivalent (%)	Analytical interpretation
Accuracy	0.982	98.2	Overall classification correctness
Precision	0.976	97.6	Reliability of positive predictions
Recall	0.971	97.1	Detection of actual stroke cases
F1-score	0.973	97.3	Precision–recall balance
ROC-AUC	0.991	99.1	Overall discrimination across thresholds
Error rate, derived from accuracy	0.018	1.8	Complement of reported accuracy
Precision–recall difference	0.005	0.5	Absolute difference in percentage points

IEER
Institute for Excellence in Education & Research

The comparatively small difference of 0.5 percentage points between precision and recall suggests that the reported performance is not characterized by a large disparity between these measures. The overall error rate derived from diagnostic accuracy is 1.8%, although this quantity should not be interpreted as the false-positive rate because it includes all incorrect predictions. Similarly, the ROC-AUC value indicates discrimination across thresholds but does not establish that the selected decision threshold is clinically optimal. For comprehensive assessment, sensitivity, specificity, predictive calibration, and the underlying confusion matrix should be evaluated using the original patient-level predictions [38]. The reported diagnostic performance is visualized in Figure 9, which presents the accuracy, precision, recall, F1-score, and ROC-AUC of the proposed architecture. This representation highlights the consistency of the supplied diagnostic results and facilitates comparison between the principal evaluation measures.

Conclusion:

This study presents an advanced Attention-Enhanced Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) framework for multimodal brain-imaging analysis, automated ischemic stroke detection, and neurological disease-severity classification. The proposed architecture integrates computed tomography (CT), diffusion-weighted imaging (DWI), fluid-attenuated inversion recovery (FLAIR), and apparent diffusion coefficient (ADC) imaging to extract complementary anatomical and pathological information. By combining hierarchical spatial feature extraction, bidirectional inter-slice contextual modelling, attention-based feature refinement, and adaptive multimodal fusion, the framework addresses important limitations

associated with conventional deep-learning approaches. The proposed dual-task classification strategy further enables automated differentiation between ischemic stroke and non-stroke examinations while categorizing neurological impairment into mild, moderate, and severe groups using National Institutes of Health Stroke Scale (NIHSS)-based clinical labels. The study reports an experimental dataset comprising 12,500 brain-imaging examinations from 2,500 patients, including 1,800 ischemic stroke patients and 700 non-stroke controls. A patient-level training, validation, and testing strategy is incorporated to minimize information leakage and support independent performance assessment. The reported experimental findings demonstrate a diagnostic accuracy of 98.2%, precision of 97.6%, recall of 97.1%, F1-score of 97.3%, and ROC-AUC of 0.991 for ischemic stroke detection. Furthermore, the proposed framework achieves a severity-classification accuracy of 95.4% and macro-F1-score of 94.6%, indicating promising performance in distinguishing different levels of neurological impairment. The integration of CNN-based spatial representations with BiLSTM contextual learning provides a mechanism for examining anatomical dependencies across consecutive imaging slices, while attention-guided multimodal fusion is designed to emphasize diagnostically informative characteristics. Grad-CAM visualization additionally supports qualitative examination of the anatomical regions contributing to automated predictions.

REFERENCES:

Raza, M. L., Hassan, S. T., Jamil, S., Hyder, N., Batool, K., Walji, S., & Abbas, M. K. (2025). Advancements in deep learning for early diagnosis of Alzheimer's disease using multimodal neuroimaging: challenges and future directions. *Frontiers in Neuroinformatics*, *19*, 1557177.

Balsabti, S. M., Al-Gburi, R. M., Mustafa, A., Ahmed, S. K., Issa, A. M., Al-Naimi, T. M., ... & Elhenidy, A. M. (2026). Advances in deep learning for multimodal brain imaging: A comprehensive survey. *Neuroscience Informatics*, *6*(1), 100252.

Luo, N., Shi, W., Yang, Z., Song, M., & Jiang, T. (2024). Multimodal fusion of brain imaging data: Methods and applications. *Machine Intelligence Research*, *21*(1), 136-152.

Li, Y., Wang, J., Pan, X., Shan, Y., & Zhang, J. (2025). MRI-mediated intelligent multimodal imaging system: from artificial intelligence to clinical imaging diagnosis. *Drug Discovery Today*, *30*(7), 104399.

Wang, S. Q., Jiao, C. N., Gao, Y. L., Cui, X. C., Wang, Y. L., & Liu, J. X. (2025). Deep association analysis framework with multi-modal attention fusion for brain imaging genetics. *Medical Image Analysis*, 103827.

Ahmad, B., Muneer, M., Jillani, S. A., & Ahmed, S. (2026, January). Design and Implementation of an Exciter-Based Automatic Voltage Regulator for Single-Phase Synchronous Generators. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Bhattacharya, S., Prusty, S., Pande, S. P., Gulhane, M., Lavate, S. H., Rakesh, N., & Veerasamy, S. (2025). Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging. *Frontiers in Human Neuroscience*, *19*, 1552178.

Zhou, X., Lin, Q., Gui, Y., Wang, Z., Liu, M., & Lu, H. (2021). Multimodal MR images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. *Frontiers in Neuroscience*, *15*, 710133.

Xu, Q., Lu, J., Zhang, Z., Xu, D., & Guo, C. (2025). Deep learning–based cognitive impairment brain imaging analysis: New methods, new technologies, and new paradigms. *Neural Regeneration Research*, *21*(9), 4135.

Ahmad, B., Ali, S., Muneer, M., Fatima, A., & Abbas, N. (2025). Wind energy integration into the SAARC region: A comprehensive review of optimal wind sites for a sustainable super smart grid. Wind Energy, 3(2).

Sun, X., Guo, W., & Shen, J. (2023). Toward attention-based learning to predict the risk of brain degeneration with multimodal medical data. *Frontiers in neuroscience*, *16*, 1043626.

Sharma, S., & Mandal, P. K. (2022). A comprehensive report on machine learning-based early detection of alzheimer's disease using multi-modal neuroimaging data. *ACM Computing Surveys (CSUR)*, *55*(2), 1-44.

Amine, J. M., & Mourad, M. (2025). Multimodal deep learning approaches for early detection of Alzheimer's disease: a comprehensive systematic review of image processing techniques. *Current Alzheimer Research*.

Peng, Z., Wang, C., Liu, S., Liang, Z., Ye, Z., Ju, M. J., ... & Yuan, Y. (2026). Omnibrainbench: A comprehensive multimodal benchmark for brain imaging analysis across multi-stage clinical tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 42732-42743).

Jiang, T., Liu, Y., Shi, F., Shu, N., Liu, B., Jiang, J., & Zhou, Y. (2008). Multimodal magnetic resonance imaging for brain disorders: advances and perspectives. *Brain Imaging and Behavior*, *2*(4), 249-257.

Kazazian, K., Norton, L., Laforge, G., Abdalmalak, A., Gofton, T. E., Debicki, D., ... & Owen, A. M. (2021). Improving diagnosis and prognosis in acute severe brain injury: a multimodal imaging protocol. *Frontiers in Neurology*, *12*, 757219.

Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Li, X., & Jin, X. (2025, August). Recent Advances in Multimodal Integration for Alzheimer's Disease Diagnosis: From Brain Imaging to Behavioral and Molecular Signatures. In *2025 19th International Conference on Complex Medical Engineering (CME)* (pp. 360-366). IEEE.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S. K., Huntenburg, J. M., ... & Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage*, *148*, 179-188.

Yao, Z., Wang, H., Yan, W., Wang, Z., Zhang, W., Wang, Z., & Zhang, G. (2023). Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. *European Journal of Radiology*, *165*, 110934.

Dcouto, S. S., & Pradeepkandhasamy, J. (2024). Multimodal deep learning in early autism detection—recent advances and challenges. *Engineering Proceedings*, *59*(1), 205.

Usman, M., Rehman, A., Shahid, A., Rehman, A. U., Gho, S. M., Lee, A., ... & Razzak, I. (2024). Advancing metaverse-based healthcare with multimodal neuroimaging fusion via multi-task adversarial variational autoencoder for brain age estimation. *IEEE Journal of Biomedical and Health Informatics*.

Ali, S., Li, J., Pei, Y., Khurram, R., Rehman, K. U., & Mahmood, T. (2022). A Comprehensive Survey on Brain Tumor Diagnosis Using Deep Learning and Emerging Hybrid Techniques with Multi-modal MR Image: S. Ali et al. *Archives of computational methods in engineering*, *29*(7), 4871-4896.

Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Gao, X., Shi, F., Shen, D., & Liu, M. (2021). Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease. *IEEE journal of biomedical and health informatics*, *26*(1), 36-43.

Liu, S., Cai, W., Liu, S., Zhang, F., Fulham, M., Feng, D., ... & Kikinis, R. (2015). Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain informatics*, *2*(3), 167-180.

Ma, D., Zhang, H., & Wang, L. (2024). Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders. *Frontiers in Neuroscience*, *18*, 1497417.

Bi, Y., Abrol, A., Fu, Z., & Calhoun, V. D. (2024). A multimodal vision transformer for interpretable fusion of functional and structural neuroimaging data. *Human brain mapping*, *45*(17), e26783.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., ... & Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience*, *19*(11), 1523-1536.

Chen, J., Yu, K., Bi, Y., Ji, X., & Zhang, D. (2024). Strategic integration: A cross-disciplinary review of the fNIRS-EEG dual-modality imaging system for delivering multimodal neuroimaging to applications. *Brain Sciences*, *14*(10), 1022.

Ashraf, A., Qingjie, Z., Bangyal, W. H. K., & Iqbal, M. (2023). Analysis of brain imaging data for the detection of early age autism spectrum disorder using transfer learning approaches for internet of things. *IEEE Transactions on Consumer Electronics*, *70*(1), 4478-4489.

Muksimova, S., Umirzakova, S., Baltayev, J., & Cho, Y. I. (2025). Multi-modal fusion and longitudinal analysis for Alzheimer's disease classification using deep learning. *Diagnostics*, *15*(6), 717.

Rafique, M., Ahmad, B., & Talani, R. A. (2026). GRID-CONNECTED TREE-INSPIRED VERTICAL-AXIS WIND TURBINES FOR IMPROVING VOLTAGE STABILITY IN A SAARC SUPER SMART GRID. Spectrum of Engineering Sciences, 4(3), 6867-6882.

Zhang, S., Yang, J., Zhang, Y., Zhong, J., Hu, W., Li, C., & Jiang, J. (2023). The combination of a graph neural network technique and brain imaging to diagnose neurological disorders: a review and outlook. *Brain Sciences*, *13*(10), 1462.

Jie, N. F., Zhu, M. H., Ma, X. Y., Osuch, E. A., Wammes, M., Théberge, J., ... & Calhoun, V. D. (2015). Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. *IEEE transactions on autonomous mental development*, *7*(4), 320-331.

Zhang, H., Hu, Y., Yu, Y., Zhou, Z., Sun, Y., Qi, C., ... & Zhu, H. (2025). The value of multimodal neuroimaging in the diagnosis and treatment of post-traumatic stress disorder: a narrative review. *Translational Psychiatry*, *15*(1), 208.

Benredjem, S., Mekhaznia, T., Abdulghafor, R., Turaev, S., Bennour, A., Sofiane, B., ... & Al Sarem, M. (2024). Parkinson's disease prediction: An attention-based multimodal fusion framework using handwriting and clinical data. *Diagnostics*, *15*(1), 4.

Pan, C. T., Chang, W. H., Kumar, A., Singh, S. P., Kaushik, A. C., Sharma, J., ... & Shiue, Y. L. (2019). Nanoparticles-mediated brain imaging and disease prognosis by conventional as well as modern modal imaging techniques: A comparison. *Current Pharmaceutical Design*, *25*(24), 2637-2649.

Liu, X., & Li, Y. (2026). Multimodal brain imaging neurotransmitter marker study of TD: a review of TCM effect mechanisms and precise diagnosis and treatment combining traditional Chinese and western medicine. *American Journal of Translational Research*, *18*(7), 6264.

Feng, Y., Zhang, X. D., Zheng, G., & Zhang, L. J. (2019). Chemotherapy-induced brain changes in breast cancer survivors: evaluation with multimodality magnetic resonance imaging. *Brain imaging and behavior*, *13*(6), 1799-1814.

Abdullah, A. S., Keerthana, V., Geetha, S., & Mishra, U. (2025). Leveraging deep learning for enhanced diagnosis of autism spectrum disorder using resting-state functional magnetic resonance imaging and clinical data. *Results in Engineering*, *25*, 104444.

Wang, G., Li, T., Sun, Q., Peng, H., Ren, T., Fu, Y., ... & Hu, B. (2026). Multimodal graph fusion-based GCN for Alzheimer's disease diagnosis using fMRI and T1-weighted MRI. *Neural Networks*, 108748.

Yu, X., Liu, J., Lu, Y., Funahashi, S., Murai, T., Wu, J., ... & Zhang, Z. (2024). Early diagnosis of Alzheimer's disease using a group self-calibrated coordinate attention network based on multimodal MRI. *Scientific Reports*, *14*(1), 24210.

Ottoni, M., Kasperczuk, A., & Tavora, L. M. (2025). Machine learning in MRI brain imaging: A review of methods, challenges, and future directions. *Diagnostics*, *15*(21), 2692.

Zhang, G., Gao, Z., Duan, C., Liu, J., Lizhu, Y., Liu, Y., ... & Dai, Q. (2025). A Multimodal Vision-Text AI Copilot for Brain Disease Diagnosis and Medical Imaging. *medRxiv*, 2025-01.

IEER
Excellence in Education & Research