Ingénierie des Systèmes d'Information
Vol. 31, No. 10, October, 2026, pp. 3597-3612
Journal homepage: http://iieta.org/journals/isi

# A Privacy-Aware Deep Learning Framework for Joint Magnetic Resonance Imaging-Based Brain Disorder Classification and Clinical Score Regression Across Heterogeneous Datasets

Padmapriya Krishna Murthy

Department of Computer Science and Engineering, R.M.D Engineering College, RSM Nagar, Chennai 601206, India

Corresponding Author Email: priya.gpjan18@gmail.com

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.311029

**Received:** 10 February 2026
**Revised:** 15 September 2026
**Accepted:** 29 September 2026
**Available online:** 31 October 2026

***Keywords:***
*magnetic resonance imaging, brain disorder classification, clinical score regression, deep learning, domain adaptation, differential privacy, medical data security*

**ABSTRACT**

Automated analysis of brain magnetic resonance imaging (MRI) can support the classification of neurological disorders, but differences between imaging datasets may limit model generalization, while the use of clinical data raises important privacy concerns. This study presents a privacy-aware deep learning (DL) framework for joint brain disorder classification and clinical score regression across heterogeneous MRI datasets. The framework combines a deep convolutional neural network encoder with a Modified Domain Adaptation Method (MDAM), which aligns latent state-transition structures between source and target domains. A shared representation is then used by parallel classification and regression branches to predict diagnostic categories and associated clinical scores. The framework was evaluated on 2,500 MRI images drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and a publicly available Kaggle brain MRI dataset. The four diagnostic categories were cognitively normal (CN), mild cognitive impairment (MCI), Alzheimer's disease (AD), and other brain disorders. Privacy safeguards included data de-identification, AES-256 encryption, role-based access control, audit logging, and differential privacy (DP) during model training. The proposed model achieved an accuracy of 94.48%, a precision of 94.12%, a recall of 93.86%, and an F1-score of 93.99%, outperforming the four comparison methods reported in the study. Ablation experiments further indicated that the feature encoder, domain adaptation module, and joint-learning components each contributed to classification performance. These findings suggest that combining domain alignment with joint prediction may improve MRI-based classification across heterogeneous datasets. The reported security evaluations, however, do not establish clinical deployment or comprehensive real-world cybersecurity performance.

## 1. INTRODUCTION

Brain disorders such as AD, MCI, Parkinson's disease (PD), brain tumors, and other neurodegenerative diseases are the most common causes of disability and mortality globally. A timely diagnosis is crucial to start treatment as early as possible, monitor disease progression, and enhance patients' quality of life [1]. MRI is one of the most widely used non-invasive imaging methods because it provides high-resolution structural information from brain tissue without exposing the patient to ionizing radiation [2]. Clinically, MRIs are often used alongside neurologic exams and cognitive assessment scores to aid diagnosis. However, manual MRI interpretation is subjective, time-consuming, and heavily reliant on radiologists' experience, and it can lead to inter-observer variability and delayed diagnosis [3].

In recent years, advancements in Artificial Intelligence (AI), specifically Machine Learning (ML) and DL, have greatly enhanced automated medical image analysis. Existing ML methods require manually designed features, such as texture descriptors, shape measurements, intensity statistics, etc., and tissue volume estimation before diagnosis. These approaches rely on expert-designed features, which are not always well suited to the anatomical variations of different neurological disorders [4]. Moreover, handcrafted features are generally not very robust when applied to heterogeneous datasets collected with different scanners, imaging protocols, and patient populations [5].

One popular alternative is DL, which learns hierarchical feature representations directly from MRI images. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders, and deep belief networks have been successful for disease diagnosis, lesion detection, and brain image segmentation [6]. These models remove the need for manual feature engineering and can learn discriminative representations from large-scale datasets, providing greater diagnostic accuracy than ever before. As a result, DL-based Computer-Aided Diagnosis (CAD) has emerged as an important field for enhancing Clinical Decision Support (CDS) systems [7].

Several crucial issues remain unaddressed. First, many studies classify brain diseases or predict a clinical score, but few combine both tasks into a single learning framework. Disease categories and clinical assessment scores are highly correlated, and treating them independently defeats the purpose of using the model, since it can obtain information

directly from one of the disease categories [8]. Second, MRI data from different hospitals may also be subject to domain shifts due to different scanner manufacturers, settings, image quality, and patient populations. All these differences affect the generalization capability of DL models trained on one dataset and limit their application in real-world clinical settings [9]. Third, current approaches focus on predictive performance and often neglect medical image analysis, even though such images are becoming more relevant for privacy-preserving analysis and are increasingly demanded for automated clinical decision-making. These limitations motivate the development of a more powerful, generalizable model that can handle disease diagnosis and clinical score diagnosis together while remaining reliable on heterogeneous data [10].

To overcome these research gaps, this paper introduces an Amalgamated DL model for Joint Regression and Diagnosis (ADL-JR&C) for secure clinical diagnosis via MRI images. In contrast to existing DL approaches optimized for a single task, the proposed framework is trained to perform two tasks: disease diagnosis and clinical score regression, allowing the model to learn more meaningful representations of features to capture disease severity and progression. Moreover, add a Modified Domain Adaptation Method (MDAM) to reduce differences between MRI datasets with different spatial distributions and acquisition characteristics. The proposed framework thus enhances the model's generalizability and its ability to diagnose diseases in different clinical contexts. The major contributions of this work are summarized as follows:

- Proposes a privacy-aware deep learning framework for joint MRI-based brain disorder classification and clinical score regression.
- Introduces MDAM to align latent state-transition structures across heterogeneous MRI datasets.
- Develops shared feature representations supporting simultaneous diagnostic classification and clinical score estimation.
- Integrates encryption, differential privacy, access control, and audit logging for enhanced patient data protection.
- Demonstrates improved classification performance through comparative evaluation and systematic ablation experiments.

## 2. LITERATURE REVIEW

DL has emerged as a leading paradigm for automated brain disease diagnosis because it can learn hierarchical feature representations directly from MRI images. This study proposed a joint transformer architecture for 3D MRI-based AD diagnosis. The method combines ViT-based feature extraction with time-series transformer modeling to capture inter-slice dependencies and long-range anatomical relationships. Using T1-weighted Alzheimer's Disease Neuroimaging Initiative (ADNI) MRI data, the framework performed binary and multiclass diagnosis. Study [11] demonstrated the potential of transformer-based representations for automated AD diagnosis. A multi-task joint learning framework was developed for simultaneous AD diagnosis and clinical-score diagnosis. The method incorporates adaptive patch pruning to remove information-poor image regions and combines fine-grained MRI features with non-image clinical information. A weighted multi-task loss enables subjects with incomplete clinical scores to contribute to training. This study [12] directly demonstrates the benefit of jointly modeling categorical diagnosis and clinical severity.

The ViT-AD framework employed a modified Vision Transformer for five-stage AD diagnosis using brain MRI. The architecture incorporated fine-tuned transformer parameters and additional layers to improve discrimination among cognitively normal, early MCI, late MCI, MCI, and AD groups. The study used 1,296 ADNI MRI images and demonstrated the applicability of transformer-based feature learning for multi-stage neurological disease diagnosis [13]. Study [14] developed a transformer-based unified multimodal framework, termed AD-Transformer, to integrate structural MRI, clinical information, and genetic data for AD assessment. Patch-CNN extracted discriminative information from 3D medical-image patches, while transformer modeling integrated imaging and non-imaging representations. The framework addressed both AD diagnosis and MCI conversion diagnosis, demonstrating the value of multimodal representation learning for clinical assessment.

The study demonstrated transfer learning for AD diagnosis under changing MRI technologies and acquisition protocols. The study addressed domain shifts from evolving MRI hardware and protocols and evaluated transfer-learning strategies to adapt models to new imaging conditions. The findings demonstrate that domain adaptation or transfer learning can improve diagnostic performance when MRI distributions change, supporting the need for robust cross-domain learning [15]. A multicentric prototype- and pseudo-labeling-based source-free domain adaptation method was proposed for MRI-based AD and MCI diagnosis. The framework combines CNN and transformer features with multicentric prototypes to generate robust pseudo-labels without requiring access to the original source data. This approach directly addresses domain shift and privacy-related restrictions in medical imaging, providing an important direction for generalizable MRI-based neurological disease diagnosis [16].

This study introduced a multimodal surface-based transformer for early AD diagnosis using T1-weighted MRI and Positron Emission Tomography (PET)-derived cortical features. Study [17] employed a middle-fusion strategy with cross-attention to integrate multimodal representations while reducing the computational burden of volumetric processing. Experiments using ADNI1–3 demonstrated strong diagnostic performance and supported surface-based multimodal transformers as an alternative to computationally intensive 3D approaches. Investigated AD Assessment Scale diagnosis from T1-weighted MRI using fine-tuning of a pretrained DL model. Study [18] addressed the comparatively limited research on direct MRI-based clinical-score diagnosis and employed an ensemble strategy, robust pre-processing, augmentation, and regularization. Evaluation on unseen external datasets demonstrated the potential of pretrained representations for quantitative clinical assessment from brain MRI [19].

Recent studies demonstrate substantial progress in transformer-based MRI diagnosis, multimodal learning, joint disease diagnosis and clinical-score diagnosis, and domain adaptation. Domain adaptation studies primarily focus on diagnosis, while clinical-score studies generally emphasize regression [20]. Therefore, there remains scope for a unified framework that simultaneously performs MRI-based brain disease diagnosis, clinical-score diagnosis, and latent feature

alignment across heterogeneous MRI domains. The proposed ADL-JR&C with MDAM addresses this combined research gap. The proposed framework outperforms previous methods in diagnosis and regression performance and is more robust to domain variations, yielding better diagnostic accuracy and more reliable CDS.

## 3. MATERIALS AND METHODS

Developed the proposed ADL-JR&C framework for MRI-based brain disease diagnosis and clinical score diagnosis using MRI scans and associated clinical information. The collected MRI modalities, including T1, T2, and Fluid-Attenuated Inversion Recovery (FLAIR) images, were initially preprocessed with skull stripping, bias-field correction, intensity normalization, image resizing, and data augmentation to improve image quality and reduce inter-scan variability. To protect sensitive patient information, the clinical and imaging data were de-identified before model development, while secure data storage was implemented using AES-256 encryption. Role-based access control restricted data access to authorized users, and audit logging maintained traceability of data-related activities. The pre-processed MRI images were then fed into a deep CNN encoder to automatically extract hierarchical spatial and anatomical features.

An MDAM was then applied to reduce distributional differences among MRI datasets obtained from different hospitals or scanners and improve cross-domain generalization. The resulting feature representations were supplied to the ADL-JR&C model, which jointly performs disease diagnosis and clinical score regression using shared representations. The diagnosis branch employs a softmax layer to determine disease categories, whereas the regression branch predicts the corresponding clinical assessment score. A state-transition probability modelling module was incorporated to represent disease progression patterns and provide additional progression-related information. The model was trained using a joint loss function combining diagnosis and regression objectives. Model performance was evaluated using diagnosis accuracy, precision, recall, F1-score, and clinical-score diagnosis error. The model presented the predicted disease category, clinical score, progression risk, and corresponding visualization to the clinician for review and feedback. If the diagnosis was unsatisfactory, the model was updated and retrained using additional data and performance feedback. Finally, the generated diagnostic results were compiled into clinical reports for integration with Hospital Information Systems (HIS) or Picture Archiving and Communication Systems (PACS).

### 3.1 Dataset description

Brain MRI images from the ADNI database and a publicly available Kaggle Brain MRI dataset were used to evaluate the proposed ADL-JR&C framework, as summarized in Table 1. The ADNI dataset comprises 1,469 T1-weighted structural MRI images representing CN, MCI, and AD categories and serves as the source domain. The Kaggle dataset contains 1,031 brain MRI images representing normal and other brain-disorder categories and serves as the target domain, providing additional domain diversity. The datasets contain 2,500 MRI images. Before feature extraction, both datasets underwent a unified pre-processing pipeline comprising skull stripping, denoising, intensity normalization, resizing to 256 × 256 pixels, and augmentation using rotation, flipping, and zooming.

Table 2 presents the class-wise distribution of the MRI dataset across four diagnostic categories: CN, MCI, AD, and other brain disorders. The dataset contains 2,500 MRI scans, with 700 CN, 650 MCI, 650 AD, and 500 other-disorder samples. A 70:15:15 patient-level split allocates 1,750 samples for training, 375 for validation, and 375 for testing, while maintaining representative class distributions across all subsets. This stratified distribution supports reliable model training, validation, and independent performance evaluation.

**Table 1.** Dataset description

| Parameter | ADNI Dataset | Kaggle Brain MRI Dataset | Combined Dataset |
|---|---|---|---|
| Data source | AD Neuroimaging Initiative | Public Kaggle brain MRI dataset | ADNI and Kaggle datasets |
| Domain role | Source domain | Target domain | Cross-domain adaptation |
| Patient population | CN, MCI, and AD research participants | Normal and brain-disorder subjects | Heterogeneous MRI populations |
| Number of subjects | 1,469 subjects | 1,031 subjects | 2,500 subjects |
| Total MRI images | 1,469 MRI images | 1,031 MRI images | 2,500 MRI images |
| MRI modality | T1-weighted structural MRI | Brain MRI images | Standardized MRI representation |
| Acquisition environment | Multi-centre research imaging environment | Independent public dataset environment | Different acquisition environments |
| Scanner information | Multiple scanner configurations across centres | Scanner metadata inconsistently available | Scanner variation contributes to domain shift |
| Imaging protocol | ADNI research imaging protocols | Dataset-specific imaging protocols | Protocol differences create domain variability |
| Diagnostic categories | CN, MCI, and AD | Normal and brain disorder | Harmonized reported diagnostic categories |
| Diagnostic criteria | ADNI-defined research diagnosis | Publicly provided dataset labels | Label differences considered during harmonization |
| Original format | NIfTI (.nii) | JPG/PNG | Standardized model-compatible format |
| Image resolution | 256 × 256 pixels | 256 × 256 pixels | 256 × 256 pixels |
| Preprocessing | Skull stripping, denoising, normalization, resizing, augmentation | Same preprocessing procedures | Unified preprocessing pipeline |
| Training set | 70% training partition | Adaptation training data | 70% training partition |
| Validation set | 15% validation partition | Validation data where applicable | 15% validation partition |
| Testing set | 15% testing partition | Independent target evaluation | 15% testing partition |

| | | | |
|---|---|---|---|
| MDAM role | Learns source-domain representations | Enables target-domain adaptation | Reduces source-target distribution differences |
| Domain difference | Research-cohort-specific imaging characteristics | Independent dataset characteristics | Represents cross-domain distribution shift |
| Security | De-identification and controlled data access | De-identification and controlled access | Unified privacy-preserving processing |

Note: ADNI = Alzheimer's Disease Neuroimaging Initiative; MRI= magnetic resonance imaging; CN = cognitively normal; MCI = mild cognitive impairment; AD = Alzheimer's disease; MDAM = Modified Domain Adaptation Method.

**Table 2.** Patient-level splitting and data leakage prevention

| Disease Class | Total Subjects | Total Magnetic Resonance Imaging (MRI) Scans | Training (70%) | Validation (15%) | Testing (15%) |
|---|---|---|---|---|---|
| Cognitively normal (CN) | 700 | 700 | 490 | 105 | 105 |
| Mild cognitive impairment (MCI) | 650 | 650 | 455 | 98 | 97 |
| Alzheimer's disease (AD) | 650 | 650 | 455 | 98 | 97 |
| Other brain disorders | 500 | 500 | 350 | 74 | 76 |
| Total | 2,500 | 2,500 | 1,750 | 375 | 375 |

### 3.2 Security-integrated ADL-JR&C flow

As depicted in Figure 1, the proposed ADL-JR&C framework provides a systematic pipeline for MRI-based brain disease diagnosis and clinical score diagnosis with an integrated security module. The process begins by collecting MRI scans and associated clinical information from authorized repositories such as ADNI and public MRI datasets. Before model training and analysis, the security module protects sensitive patient information through data de-identification, AES-256 encryption, Role-Based Access Control (RBAC), and audit logging. Patient identifiers are replaced with pseudonymous identifiers using a one-way hashing operation:

$$ID_{Pseudo} = H\left(ID_{patient} || S\right) \tag{1}$$

where, $H(\cdot)$ denotes a cryptographic hash function and $S$ represents a randomly generated salt. The de-identified clinical and MRI data are then encrypted using AES-256 before storage or transmission:

$$C = AES_{256}(K, M) \tag{2}$$

where, $M$ represents the protected medical data, $K$ is the encryption key, and $C$ is the resulting ciphertext. Access to the stored data is controlled using RBAC, in which access is granted only when the user's assigned role possesses the required permission:

$$A(u,p) = \begin{cases} 1, & (\text{Role}(u), p) \in P \\ 0, & \text{Otherwise} \end{cases} \tag{3}$$

where, $u$ denotes the user, $p$ represents the requested operation, and $P$ is the authorized role-permission set. Each data-access event is recorded in an audit log containing the user role, timestamp, operation, and resource identifier, while a chained hash can be used to maintain log integrity:

$$L_t = H(E_t || L_{t-1}) \tag{4}$$

where, $E_t$ is the current audit event and $L_{t-1}$ is the hash of the preceding log entry. After security processing, the authorized and de-identified MRI data undergo skull stripping, denoising, intensity normalization, resizing, and augmentation. The processed images are then supplied to the deep convolutional neural network (DCNN) encoder for hierarchical anatomical feature extraction. The extracted features are aligned using MDAM, which reduces distribution differences among heterogeneous MRI datasets through state-transition probability modelling. The aligned features are subsequently passed to ADL-JR&C, where shared fully connected layers generate a common representation for simultaneous disease diagnosis and clinical score regression. The diagnosis branch applies Softmax to determine the disease category, whereas the regression branch predicts the clinical assessment score. The model optimizes the joint diagnosis, regression, and domain-adaptation objectives during training. Finally, the predicted outputs are presented to clinicians for review and integrated into the clinical decision-support workflow.

### 3.3 Reports and magnetic resonance imaging image collection

The first stage of the proposed ADL-JR&C framework involves collecting clinical reports and MRI scans from authorized medical repositories, including the ADNI database and publicly available MRI datasets. Each patient record consists of structural MRI images (T1, T2, or FLAIR-weighted scans) together with associated clinical information, including demographic characteristics, neurological examination findings, and cognitive assessment scores. Let the collected dataset be represented as:

$$\mathcal{D} = \{(I_i, Y_i, C_i)\}_{i=1}^{N} \tag{5}$$

where, $I_i$ denotes the MRI image of the ith subject, $C_i$ represents the corresponding clinical report, $Y_i$ is the disease label (e.g., CN, MCI, or AD, and $N$ is the total number of subjects. The collected MRI images are standardized to a common format before further processing. The image acquisition process can be expressed as:

$$I_i = f_{acq}(S_i, P_i) \tag{6}$$

where, $S_i$ denotes the MRI scanner characteristics, $P_i$ represents the acquisition protocol (T1, T2, or FLAIR), and $f_{acq}(.)$ denotes the image acquisition function. The corresponding clinical information is represented as:

$$C_i = [a_i, g_i, s_i, m_i] \tag{7}$$

where, $a_i$ is the patient's age, $g_i$ denotes gender, $s_i$ represents clinical assessment scores, and $m_i$ contains additional medical history. The integrated dataset $(I_i, C_i)$ forms the input to the

subsequent preprocessing stage, ensuring that both imaging biomarkers and clinical information are available for feature learning and joint diagnosis in the proposed ADL-JR&C framework. This comprehensive data collection process improves diagnostic reliability while providing sufficient information for accurate brain disease diagnosis and clinical score diagnosis.

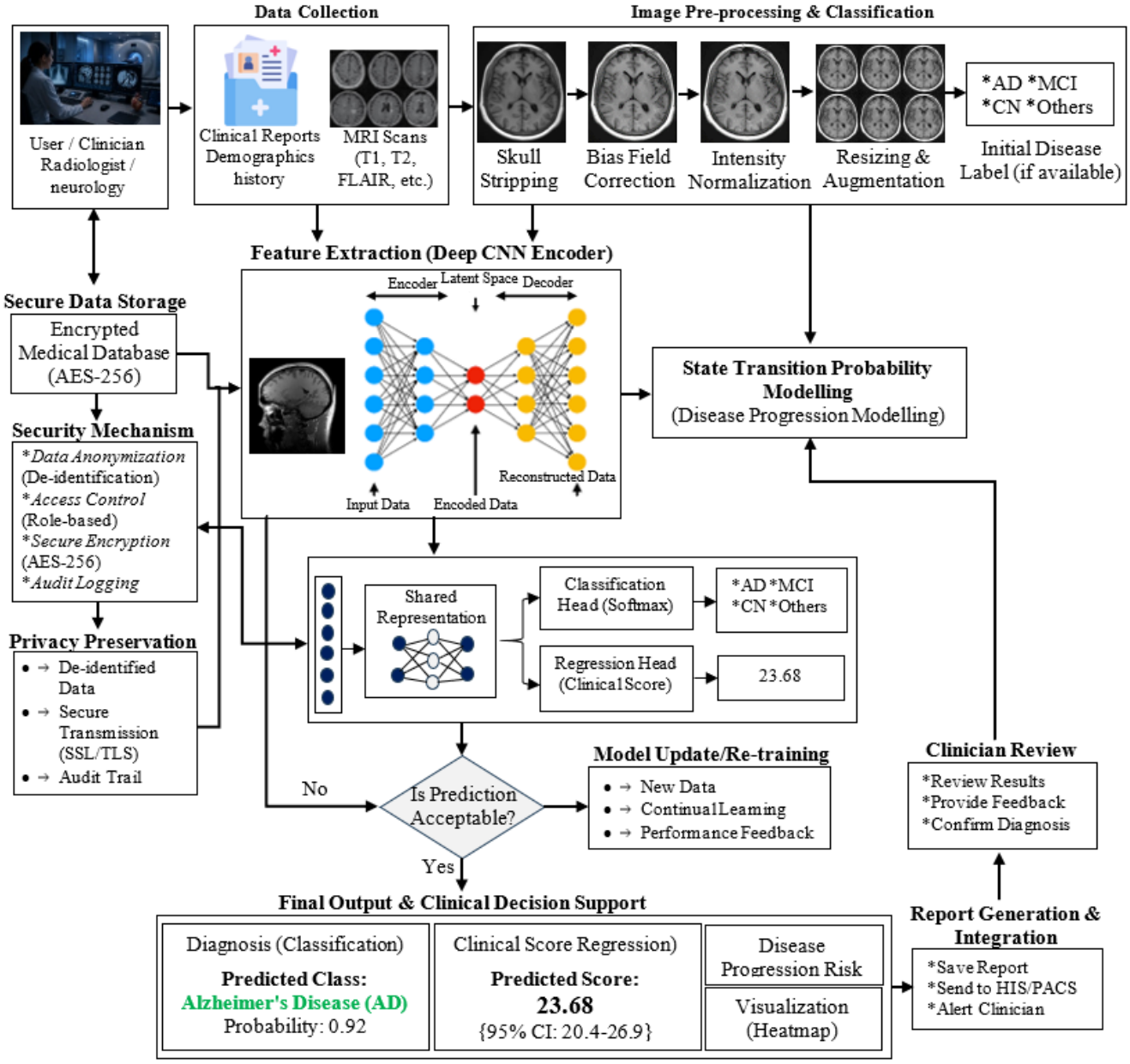



**Figure 1.** Proposed architecture

### 3.4 Image preprocessing

Image preprocessing is a crucial stage in the proposed ADL-JR&C framework, as it improves MRI image quality and ensures the extracted features are robust and consistent across datasets. Initially, the acquired MRI image $I(x, y)$ undergoes skull stripping to remove non-brain tissues such as:

$$I_b(x, y) = I(x, y) \times M(x, y) \tag{8}$$

where, $M(x, y)$ is the binary brain mask, and $I_b(x, y)$ is the extracted brain region. Next, denoise the image using a Gaussian filter to suppress imaging artifacts while preserving anatomical structures. The filtered image is obtained as:

$$I_d(x, y) = I_b(x, y) * G(x, y, \sigma) \tag{9}$$

To reduce variations caused by different MRI scanners and acquisition protocols, apply intensity normalization. The normalized image is computed as:

$$I_n(x, y) = \frac{I_d(x, y) - \mu}{\sigma} \tag{10}$$

where, $\mu$ and $\sigma$ denote the mean and standard deviation of the image intensity values, respectively. All MRI images are then resized to a fixed spatial resolution to ensure a uniform input size for the DL model. The resized image is represented as:

$$I_r = Resize(I_n, H, W) \tag{11}$$

where, $H$ and $W$ denote the target image height and width. To improve model robustness and reduce overfitting, we apply data augmentation techniques, including rotation, horizontal flipping, scaling, and translation. The augmented image is expressed as:

$$I_a = \mathrm{T}(I_r) \tag{12}$$

where, $T(\cdot)$ represents the augmentation transformation. Finally, the preprocessed image $I_a$ is forwarded to the deep feature extraction stage, where the ADL-JR&C network automatically learns hierarchical anatomical representations. These preprocessing operations improve image consistency, reduce noise and inter-scanner variability, and enhance the accuracy and generalization capability of the proposed brain disease diagnosis framework. To further protect sensitive patient information during model development, the framework incorporates differential privacy (DP) into training. Rather than adding noise directly to the MRI images, the proposed framework applies DP to the model-gradient updates, reducing the risk of revealing information about an individual patient through the trained model. For each mini-batch, the gradient of an individual sample is first clipped to a predefined maximum norm $C$:

$$\bar{g}_i = \frac{g_i}{\max\left(1, \frac{\|g_i\|_2}{C}\right)} \tag{13}$$

where, $g_i$ denotes the gradient contributed by the $i$-th training sample and $C$ is the gradient-clipping threshold. Gaussian noise is then added to the aggregated clipped gradients:

$$\tilde{g} = \frac{1}{B}\left(\sum_{i=1}^{B} \bar{g}_i + \mathcal{N}(0, \sigma_{DP}^2 C^2 I)\right) \tag{14}$$

where, $B$ is the mini-batch size, $\sigma_{DP}$ is the noise multiplier, and $I$ is the identity matrix. The model parameters are updated using the privatized gradient:

$$\theta_{t+1} = \theta_t - \eta_t \tilde{g} \tag{15}$$

where, $\theta_t$ denotes the model parameters at iteration $t$, and $\eta_t$ represents the learning rate. The privacy guarantee is quantified using the parameters $(\varepsilon, \delta)$, where $\varepsilon$ represents the privacy-loss budget and $\delta$ denotes the probability of exceeding the specified privacy guarantee. A smaller $\varepsilon$ generally provides stronger privacy protection but requires greater perturbation, which can reduce model utility. Conversely, a larger $\varepsilon$ permits less perturbation and may improve predictive performance at the cost of weaker privacy. The integration of skull stripping, denoising, intensity normalization, spatial standardization, augmentation, and gradient-based DP produces consistent and privacy-aware MRI representations.

### 3.5 Feature extraction using deep convolutional neural network encoder

The proposed ADL-JR&C framework uses a DCNN encoder to automatically extract discriminative, hierarchical features from preprocessed MRI images, as shown in Figure 2. Unlike existing ML methods that depend on handcrafted features, the DCNN encoder learns multi-level anatomical representations directly from the MRI data through successive convolutional operations. The preprocessed image $I_a \in R^{HXWXC}$ serves as input to the encoder, where multiple convolutional layers, rectified linear unit (ReLU) activations, batch normalization, and max-pooling operations progressively capture low-level texture patterns, mid-level structural information, and high-level semantic representations associated with brain abnormalities.

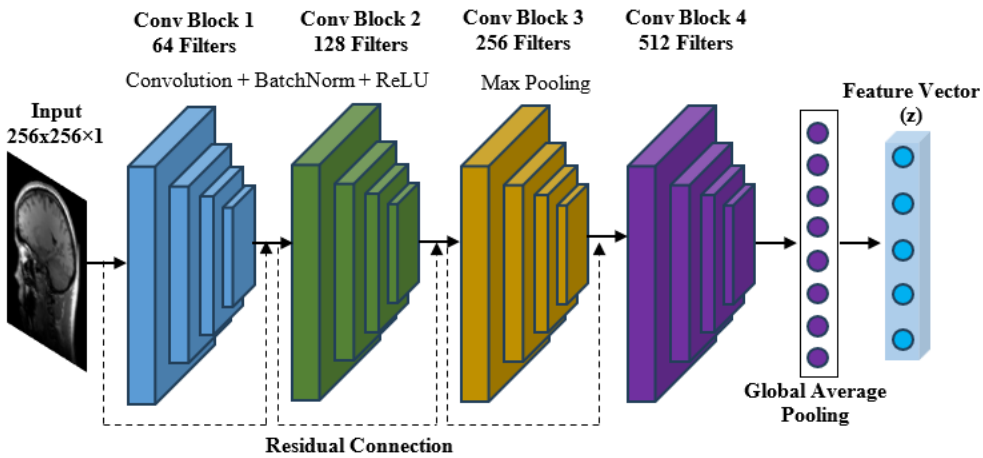



**Figure 2.** Architecture of feature extraction using deep convolutional neural network (DCNN)

The convolution operation at the $l$-th layer is expressed as:

$$F^{(l)} = \sigma(W^{(l)} * F^{(l-1)} + b^{(l)}) \tag{16}$$

where, $F^{(l-1)}$ and $F^{(l)}$ denote the input and output feature maps of the $l$-th convolution layer, $W^{(l)}$ represents the learnable convolution kernel, $b^{(l)}$ is the bias vector, $*$* denotes convolution, and $\sigma(\cdot)$ is the ReLU activation function,

$$\sigma(x) = \max(0, x) \tag{17}$$

To improve training stability and accelerate convergence, batch normalization is applied after each convolutional layer,

$$\hat{F}^{(l)} = \gamma \frac{F^{(l)} - \mu_B}{\sqrt{\sigma_b^2 + \epsilon}} \tag{18}$$

where, $\mu_B$ and $\sigma_b^2$ are the batch mean and variance, while $\gamma$ and

$\beta$ are learnable scaling and shifting parameters. Max-pooling is subsequently performed to reduce the spatial dimensions while preserving the most informative features,

$$P^{(l)}(i,j) = \max_{(m,n)\in\Omega}(m,n) \tag{19}$$

where, $\Omega$ represents the pooling window. After passing through multiple convolutional blocks, the resulting feature maps are transformed into a compact feature vector using global average pooling (GAP),

$$z = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} P^{(L)}(i,j) \tag{20}$$

where, $L$ denotes the final convolutional layer. The extracted deep feature vector $z$ captures discriminative structural characteristics of the brain and is forwarded to the MDAM for feature alignment across heterogeneous MRI datasets. The ADL-JR&C framework then uses the domain-adapted features for joint brain disease diagnosis and clinical score diagnosis, improving diagnostic accuracy and model generalization.

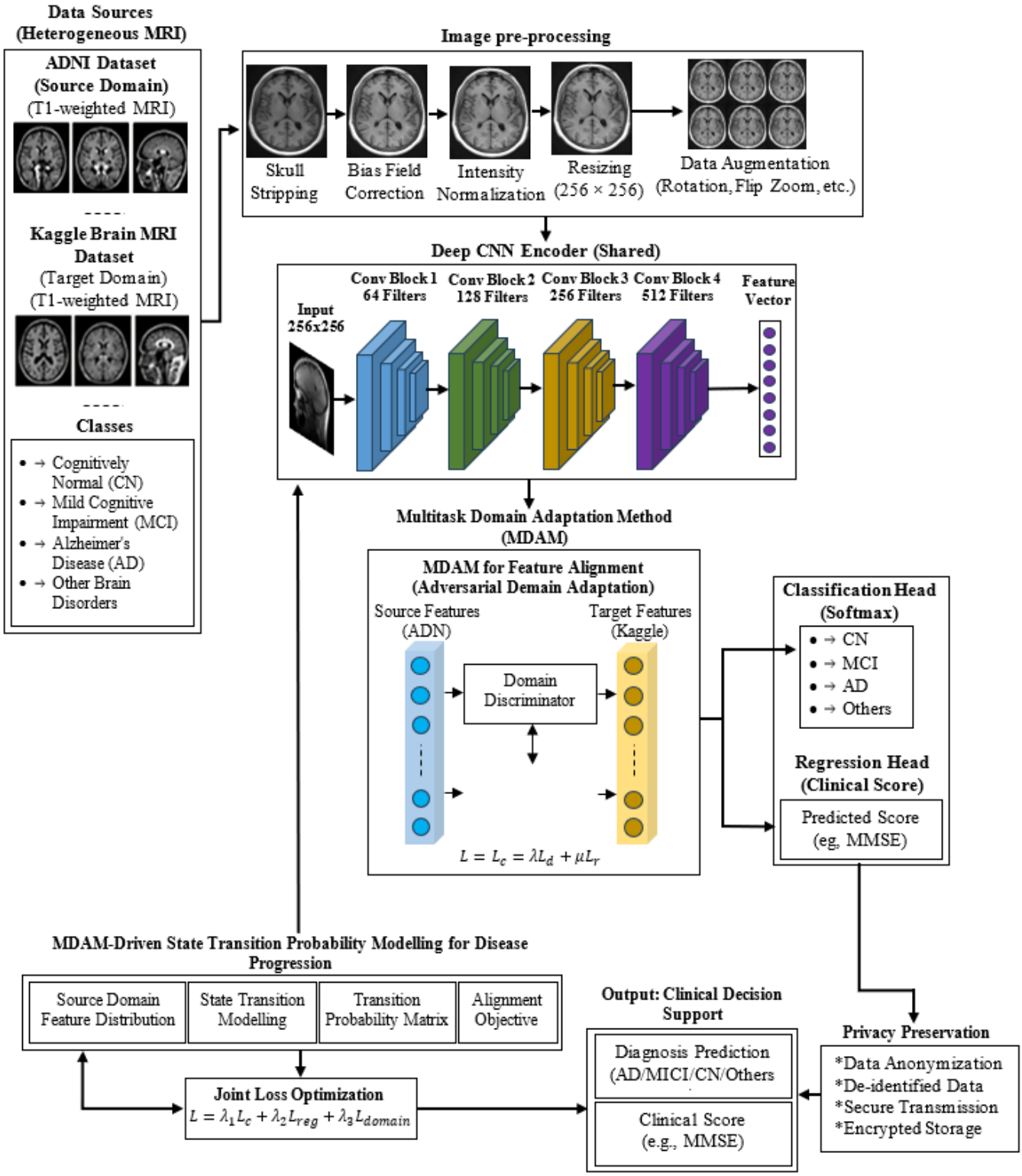



**Figure 3.** Architecture of Modified Domain Adaptation Method (MDAM)

### 3.6 Modified Domain Adaptation Method for feature alignment across heterogeneous magnetic resonance imaging datasets

The proposed MDAM reduces feature-distribution discrepancies between heterogeneous MRI datasets obtained from different scanners, imaging protocols, institutions, and patient populations. As illustrated in Figure 3, the high-level representations generated by the DCNN encoder may exhibit domain-specific variations even when corresponding MRI images contain similar anatomical information. Unlike existing domain adaptation strategies that primarily align feature statistics, MDAM aligns in a latent state-transition space, capturing the relational structure among learned MRI representations.

Let the feature representation extracted from the *i*-th MRI image be expressed as

$$f_i = D(x_i), \qquad f_i \in R^d \tag{21}$$

where, $x_i$ denotes the preprocessed MRI image, $D(.)$, represents the DCNN encoder, and $f_i$ is the resulting d-dimensional deep feature vector. The source domain S corresponds to ADNI MRI data, whereas the target domain T represents the Kaggle MRI dataset.

$$Z_i = \arg\min_{k\in\{1,\dots K\}} \|f_i - \mu_k\|_2^2 \tag{22}$$

where, $Z_i$ denotes the latent state assigned to the *i*-th feature vector and $\mu_k$ represents the centroid of the *k*-th latent state. This transformation converts the continuous high-dimensional feature space into a compact state representation, allowing the structural relationships between MRI feature patterns to be compared across domains.

$$P_{ij}^{(D)} = \frac{N_{ij}^{(D)}}{\sum_{j=1}^{K} N_{ij}^{(D)}}, \qquad D \in \{S, T\} \tag{23}$$

where, $N_{ij}^{(D)}$ denotes the number of observed transitions from latent state $i$ to state $j$ in domain $D$. Thus, $P_{ij}^{(D)}$ represents the probability that a feature representation associated with state $i$ evolves toward state $j$. The complete transition structure of each domain is represented by the state-transition probability matrix.

$$P^{(D)} = \begin{bmatrix} P_{11}^{(D)} & P_{12}^{(D)} & \dots & P_{1K}^{(D)} \\ P_{21}^{(D)} & P_{22}^{(D)} & \cdots & P_{2K}^{(D)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1}^{(D)} & P_{K2}^{(D)} & \cdots & P_{KK}^{(D)} \end{bmatrix} \tag{24}$$

MDAM's main contribution is to align the relational transition structures of the source and target domains rather than matching only their individual feature distributions. The discrepancy between the two domains is quantified using the Kullback–Leibler (KL) divergence between corresponding transition-probability distributions. The domain-adaptation loss is formulated as

$$\mathcal{L}_{DA} = \frac{1}{K}\sum_{i=1}^{K} D_{KL}\left(P_i^{(S)} || P_i^{(T)}\right) \tag{25}$$

where, $P_i^{(S)}$ and $P_i^{(T)}$ denote the *i*-th transition-probability distributions of the source and target domains, respectively. The KL divergence is given by:

$$D_{KL}\left(P_i^{(S)} || P_i^{(T)}\right) = \sum_{j=1}^{K} P_{ij}^{(S)} \log \frac{P_i^{(S)} + \epsilon}{P_i^{(T)} + \epsilon} \tag{26}$$

where, $\epsilon$ is a small positive constant introduced to avoid numerical instability when a transition probability approaches zero. Minimizing $\mathcal{L}_{DA}$ encourages the source and target domains to share similar latent transition structures, reducing domain-specific variation while preserving meaningful anatomical relationships. The aligned representation is then forwarded to the ADL-JR&C network for simultaneous brain-disease diagnosis and clinical-score diagnosis. To ensure domain alignment directly contributes to the final clinical objectives, MDAM is jointly optimized with diagnosis and regression losses. The overall training objective is defined as:

$$\mathcal{L}_{total} = \lambda_c \mathcal{L}_{cls} + \lambda_r \mathcal{L}_{reg} + \lambda_d \mathcal{L}_{DA} \tag{27}$$

where, $\mathcal{L}_{cls}$ is the diagnosis loss, $\mathcal{L}_{reg}$ is the clinical score regression loss, $\mathcal{L}_{DA}$ is the domain adaptation loss, and $\lambda_c$, $\lambda_r$, and $\lambda_d$ are weighting coefficients. By integrating state-transition probability modeling with domain adaptation, the proposed MDAM reduces inter-domain variability, improves feature consistency across heterogeneous MRI datasets, and enhances the robustness and diagnostic accuracy of the ADL-JR&C framework.

### 3.7 Amalgamated Deep Learning Model for Joint Regression and Diagnosis

The proposed ADL-JR&C is designed to simultaneously perform brain disease diagnosis and clinical score diagnosis using a shared deep feature representation extracted by the DCNN encoder as shown in Figure 4. After pre-processing and feature alignment through the MDAM, the extracted feature vector $z$ is passed to a series of shared fully connected layers that learn a common latent representation for both tasks. The shared representation is formulated as:

$$h = f(W_s z + b_s) \tag{28}$$

where, $z$ is the input feature vector, $W_s$ and $b_s$ denote the shared weight matrix and bias vector, respectively, and $f(.)$ represents the ReLU activation function.

The learned representation $h$ is then forwarded to two parallel diagnosis branches. The diagnosis branch predicts the probability of each brain disease class using the Softmax activation function,

$$P(y = c|h) = \frac{\exp(W_C^T h + b_c)}{\sum_{k=1}^{C} \exp(W_k^T h + b_k)} \tag{29}$$

where, *C* denotes the total number of disease classes (e.g., CN, MCI, AD, and other brain disorders). The predicted disease class is determined by:

$$\hat{y} = \arg\max_c P(y = c|h) \tag{30}$$

Simultaneously, the regression branch estimates the clinical

assessment score (e.g., Mini-Mental State Examination) through a linear regression layer,

$$\hat{s} = W_t^T h + b_r \tag{31}$$

where, $\hat{s}$ is the predicted clinical score, while $W_r$ and $b_r$ denote the regression weights and bias. The network is optimized by jointly minimizing the diagnosis and regression losses. The diagnosis loss is computed using categorical cross-entropy.

$$\mathcal{L}_{cls} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}) \tag{32}$$

where, $y_{ic}$ is the ground-truth class label and $\hat{y}_{ic}$ is the predicted probability. The regression loss is calculated using mean squared error (MSE),

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{i=1}^{N}(s_i - \hat{s}_i)^2 \tag{33}$$

where, $s_i$ and $\hat{s}_i$ denote the actual and predicted clinical scores, respectively. Finally, both objectives are jointly optimized through a weighted loss function,

$$\mathcal{L}_{JR\&C} = \lambda_1\mathcal{L}_{cls} + \lambda_2\mathcal{L}_{reg} \tag{34}$$

where, $\lambda_1$ and $\lambda_2$ are weighting coefficients that balance disease diagnosis and clinical score diagnosis. If the diagnosis performance does not satisfy the predefined acceptance criterion, the model parameters are updated through backpropagation and retraining until convergence. By learning a shared representation for both tasks, the proposed ADL-JR&C framework exploits the complementary relationship between disease diagnosis and clinical severity estimation, improving diagnostic accuracy, clinical score diagnosis, and robustness for MRI-based brain disease diagnosis.

Algorithm 1 begins by collecting MRI images and clinical data, then pre-processing the images to improve quality and consistency. A DCNN encoder extracts discriminative features, which are aligned across heterogeneous datasets using the MDAM. The ADL-JR&C network jointly performs brain disease diagnosis and clinical score diagnosis. Finally, the model is optimized through joint loss minimization and delivers privacy-preserving CDS for secure diagnosis.

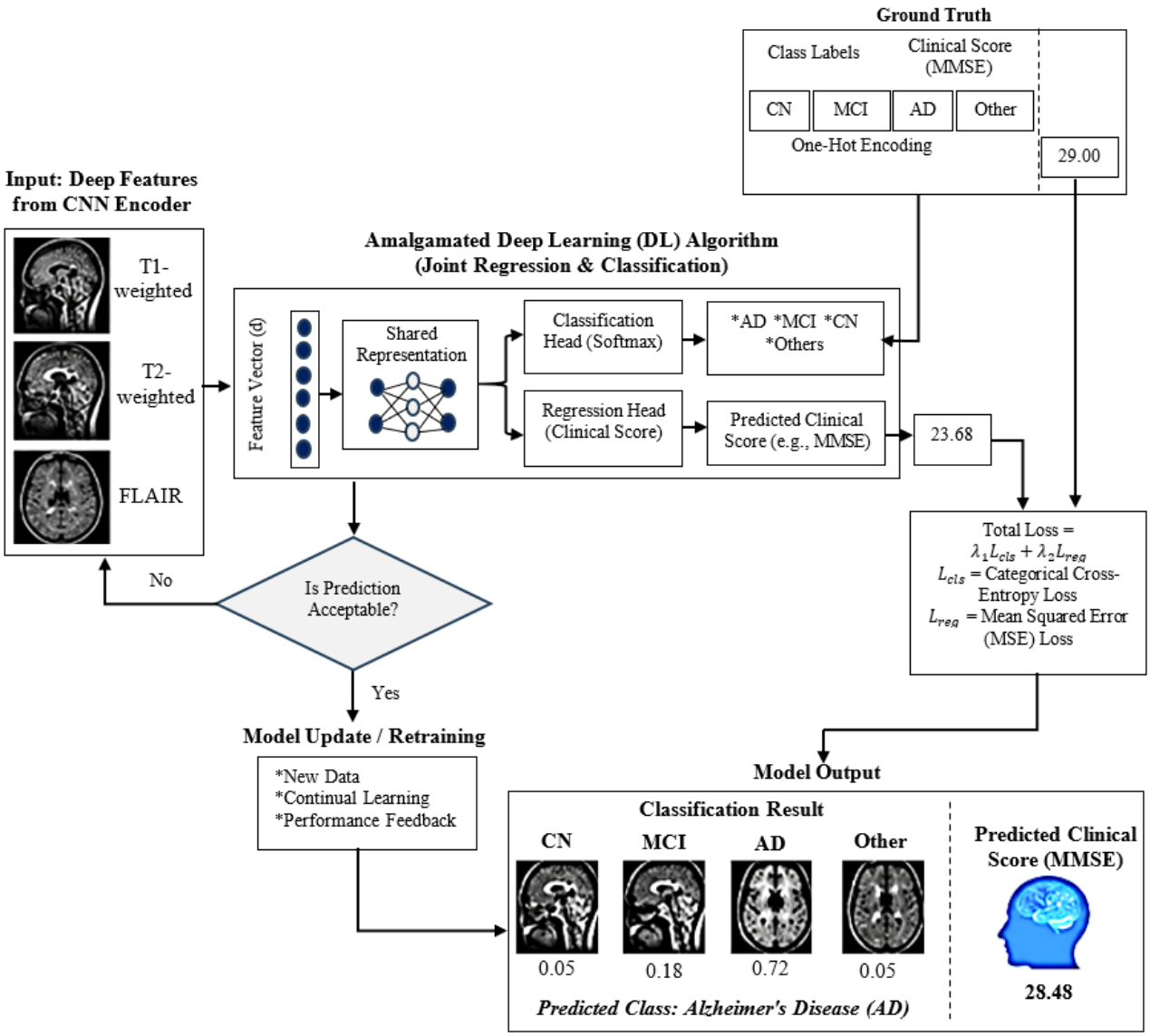



**Figure 4.** Architecture of Amalgamated Deep Learning Model for Joint Regression and Diagnosis (ADL-JR&C)

**Algorithm 1.** Proposed privacy-aware joint MRI-based brain disease diagnosis and clinical score prediction

Input: MRI dataset $\mathcal{D} = \{(I_i, Y_i, C_i)\}_{i=1}^{N}$; learning rate $\eta$; epochs T, batch size B

Output: Brain disease class $\hat{Y}$; predicted clinical score $\hat{S}$; rained model parameters θ

**Step 1:** Collect MRI images and clinical reports from ADNI/Kaggle datasets using Eq. (5)

**Step 2:** Pre-process MRI images by skull stripping, normalization, denoising, resizing, and augmentation

$$I_p = \text{Augment}(\text{Resize}(\text{Normalize}(I)) \quad (35)$$

**Step 3:** Extract deep features using the CNN encoder by using Eqs. (13)–(17)

**Step 4:** Perform MDAM to align source and target feature distributions by using Eqs. (18)–(23)

**Step 5:** Generate the shared feature representation by using Eq. (25)

**Step 6:** Predict the brain disease category using the diagnosis head

$$\hat{Y} = Softmax(W_c H + b_r) \quad (36)$$

**Step 7:** Predict the clinical assessment score using the regression head using Eq. (23)

**Step 8:** Compute the joint learning objective using Eq. (24)

**Step 9:** Update model parameters using backpropagation

$$\theta = \theta - \eta \nabla_\theta L \quad (37)$$

Repeat Steps **3–9** until convergence or maximum epochs $T$

**Step 10:** Apply privacy-preserving clinical decision support (CDS) by securely transmitting only the final diagnosis and anonymized patient information

$$O = \{\hat{Y}, \hat{S}\} \quad (38)$$

**Step 11:** Display the predicted brain disease class and clinical score to the clinician through the CDS system

**Return:** $\hat{Y}, \hat{S}, \theta$

## 4. RESULTS AND DISCUSSION

The proposed ADL-JR&C framework was evaluated using MRI images collected from the ADNI database and a publicly available Kaggle Brain MRI dataset. The combined dataset consists of 2,500 MRI images, including 700 CN, 650 MCI, 650 AD, and 500 Other Brain Disorder images. All MRI scans were preprocessed using skull stripping, Gaussian noise removal, intensity normalization, resizing to 256 × 256 pixels, and data augmentation through rotation, flipping, and scaling. The dataset was divided into 70% training (1,750 images), 15% validation (375 images), and 15% testing (375 images) while maintaining class balance. Trained the proposed DCNN encoder using the Adam optimizer with an initial learning rate of 0.0001, a batch size of 32, 100 epochs, a dropout rate of 0.5, and categorical cross-entropy and mean squared error as joint optimization objectives. The implementation was carried out using Python 3.11, TensorFlow 2.16, Keras, OpenCV, NumPy, Scikit-learn, and Compute Unified Device Architecture (CUDA) 12.2 in the Visual Studio Code development environment running on Windows 11 Pro (64-bit). The experiments ran on a workstation with an Intel Core i9-13900K processor, 64 GB DDR5 RAM, an NVIDIA GeForce RTX 4090 GPU (24 GB GDDR6X), and a 2 TB NVMe SSD, enabling efficient training and evaluation of the proposed privacy-preserving MRI-based brain disease diagnosis framework.

Table 3 presents the principal experimental settings of the proposed ADL-JR&C framework, including dataset split, network configuration, and training parameters. These settings ensure reproducibility, balanced model training, stable convergence, and reliable evaluation of MRI-based brain disease diagnosis and clinical score diagnosis. Software and hardware components and specifications in Table 4.

Figure 5 compares the proposed ADL-JR&C framework with four existing approaches using accuracy, precision, recall, and F1-score. The proposed method consistently achieves the highest performance, attaining 94.48% accuracy, 94.12% precision, 93.86% recall, and 93.99% F1-score. These improvements demonstrate the effectiveness of integrating DCNN-based feature extraction, MDAM, and joint regression-diagnosis learning. The enhanced feature representation and domain alignment enable more reliable MRI-based brain disease diagnosis and clinical score diagnosis than existing methods.

**Table 3.** Hyper-parameter settings

| Parameter | Settings |
|---|---|
| Model | ADL-JR&C |
| Feature extractor | DCNN Encoder |
| Domain adaptation | MDAM |
| Total MRI images | 2,500 |
| Training split | 70% (1,750 images) |
| Validation split | 15% (375 images) |
| Testing split | 15% (375 images) |
| Input image size | 256 × 256 × 3 |
| Optimizer | Adam |
| Learning rate | 0.0001 |
| batch size | 32 |
| Number of epochs | 100 |
| Activation function | ReLU |
| Dropout rate | 0.5 |
| Loss function | Cross-Entropy + Mean Squared Error + MDAM Loss |
| Evaluation metrics | Accuracy, Precision, Recall, F1-score, mean absolute error (MAE), root mean squared error (RMSE |

Note: MRI= magnetic resonance imaging; MDAM = Modified Domain Adaptation Method; ADL-JR&C = Amalgamated Deep Learning Model for Joint Regression and Diagnosis; ReLU = rectified linear unit; DCNN = deep convolutional neural network.

**Table 4.** Software and hardware environment

| Component | Specification |
|---|---|
| Operating system | Windows 11 Pro (64-bit) |
| Programming language | Python 3.11 |
| DL framework | TensorFlow 2.16, Keras |
| Libraries | NumPy, OpenCV, Scikit-learn, Pandas, Matplotlib |
| IDE | Visual Studio Code |
| CUDA version | CUDA 12.2 |
| CPU | Intel Core i9-13900K |
| GPU | NVIDIA GeForce RTX 4090 (24 GB GDDR6X) |
| RAM | 64 GB DDR5 |
| Storage | 2 TB NVMe SSD |

Note: DL= deep learning; IDE = Integrated Development Environment; CUDA = Compute Unified Device Architecture; RAM = Random Access Memory.

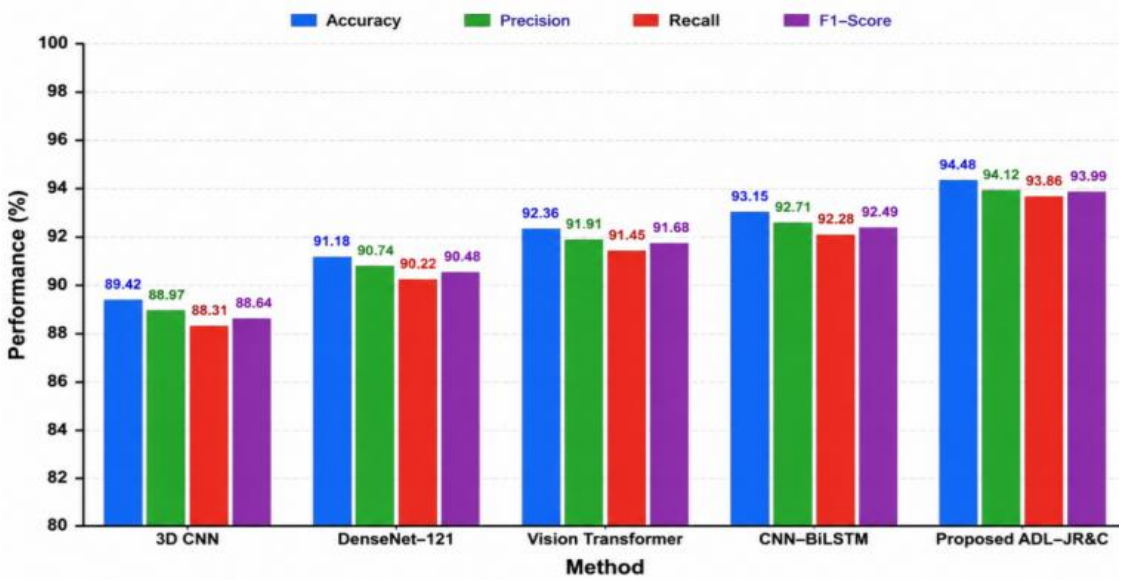



**Figure 5.** Comparison of performance measures of the proposed and existing systems

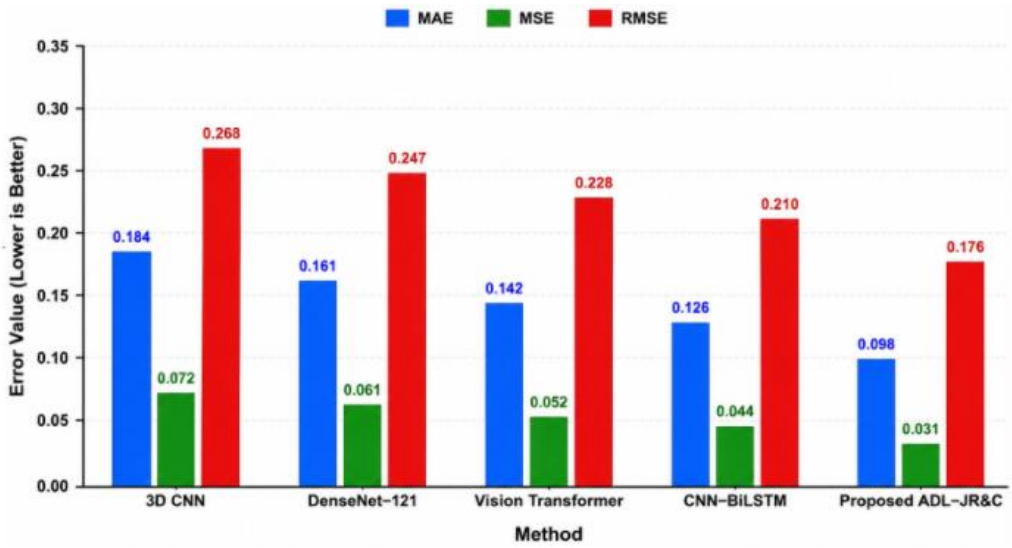



**Figure 6.** Comparison of performance measures (Error) of the proposed and existing systems

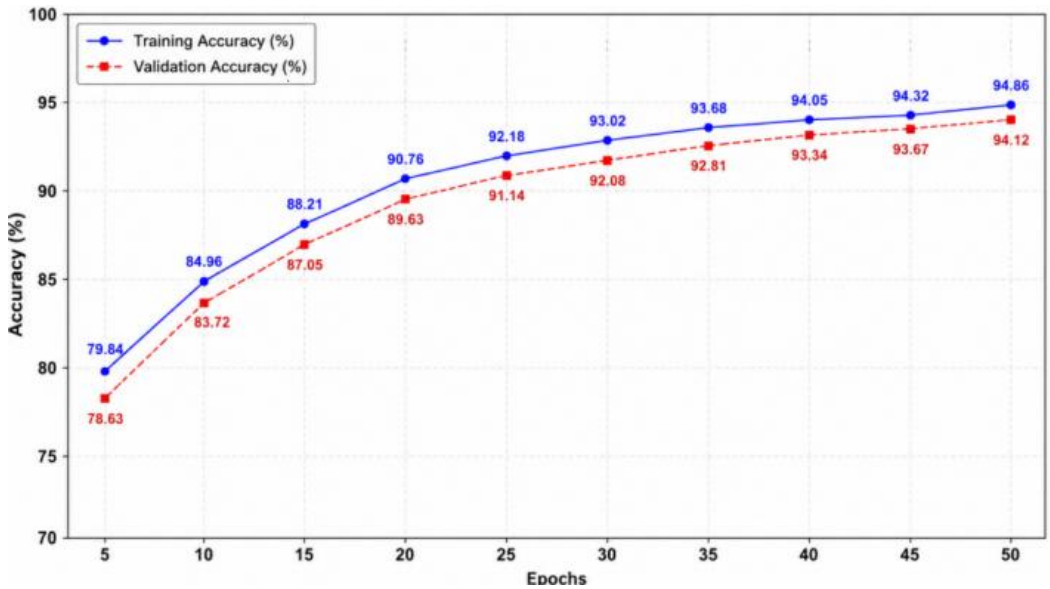



**Figure 7.** Comparison of training and validation accuracy vs. epochs of the proposed system

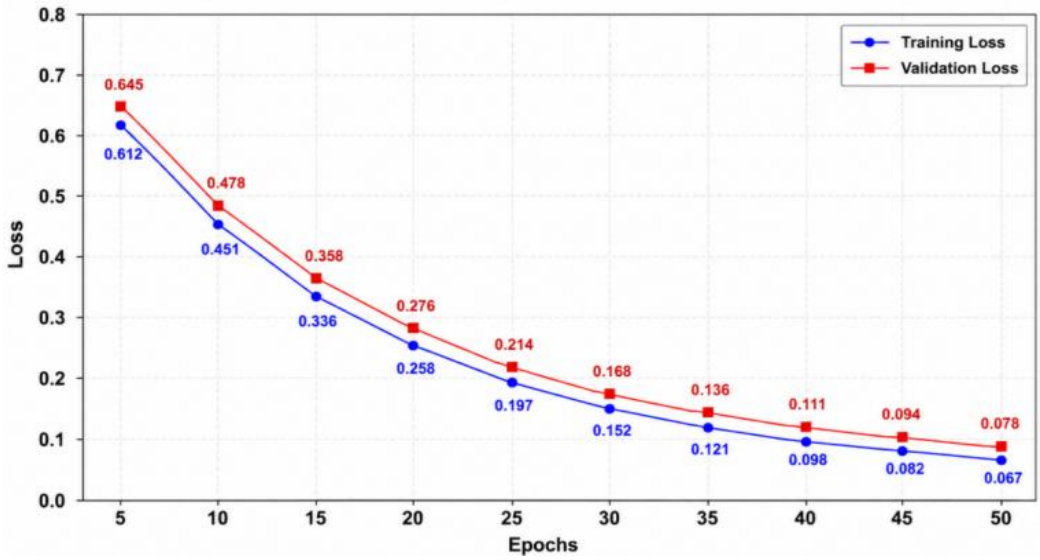



**Figure 8.** Comparison of training and validation loss vs. epochs of the proposed system

Figure 5 illustrates the comparative performance of five DL methods using Accuracy, Precision, Recall, and F1-Score. The proposed ADL-JR&C consistently outperforms the existing approaches, achieving the highest Accuracy (94.48%), Precision (94.12%), Recall (93.86%), and F1-Score (93.99%). Performance improves progressively from 3D CNN to CNN–Bidirectional Long Short-Term Memory (BiLSTM), while the proposed framework demonstrates superior diagnosis capability through adaptive joint representation learning and optimized feature extraction, resulting in enhanced diagnosis reliability and overall diagnostic effectiveness.

Figure 6 compares diagnosis errors using MAE, MSE, and RMSE across five methods. The proposed ADL-JR&C achieves the lowest error values, with an MAE of 0.098, MSE of 0.031, and RMSE of 0.176, indicating superior diagnosis accuracy and robustness. Compared with recent approaches, including CNN–BiLSTM and ViT, the proposed framework significantly reduces estimation errors through adaptive joint representation learning and optimized diagnosis, resulting in more reliable and consistent diagnostic performance.

Figure 7 illustrates the training and validation accuracy of the proposed ADL-JR&C model over 50 epochs. Both curves show a consistent upward trend, indicating stable learning and effective convergence. Training accuracy increases from 79.84% to 94.86%, while validation accuracy improves from 78.63% to 94.12%. The small gap between the curves demonstrates excellent generalization capability and minimal overfitting during training.

Figure 8 shows the training and validation loss of the proposed ADL-JR&C model over 50 epochs. Both losses decrease consistently, indicating stable optimization and effective learning. Training loss declines from 0.612 to 0.067, while validation loss reduces from 0.645 to 0.078. The narrow gap between the curves demonstrates good generalization, minimal overfitting, and robust convergence throughout the training process.

Figure 9 compares the security performance of the proposed ADL-JR&C framework with four existing approaches. The proposed framework achieves the highest Authentication Accuracy (98.45%), Attack Detection Rate (97.92%), Data Integrity (98.16%), Privacy Preservation (98.24%), and Cyber Resilience (97.83%). These improvements demonstrate that integrating deep feature learning with privacy-preserving mechanisms significantly strengthens authentication, protects sensitive clinical information, ensures data integrity, and improves resilience against cyber threats, making the framework well suited for secure MRI-based clinic decision support systems. Figure 10 compares the security performance of the proposed ADL-JR&C framework with four existing approaches. The proposed framework achieves the highest Authentication Accuracy (98.45%), Attack Detection Rate (97.92%), Data Integrity (98.16%), Privacy Preservation (98.24%), and Cyber Resilience (97.83%). These improvements demonstrate that integrating deep feature learning with privacy-preserving mechanisms significantly strengthens authentication, protects sensitive clinical information, ensures data integrity, and improves resilience against cyber threats, making the framework well suited for secure MRI-based CDS systems.

Table 5 presents the statistical significance analysis based on five independent experimental runs for each evaluated method. The results are expressed as mean accuracy ± standard deviation, while the *t*-values and *p*-values quantify the statistical differences between the proposed ADL-JR&C and existing approaches. All *p*-values are below 0.05, indicating statistically significant differences; the proposed ADL-JR&C achieves a highly significant result ($p < 0.001$), supporting the reliability of the reported performance.

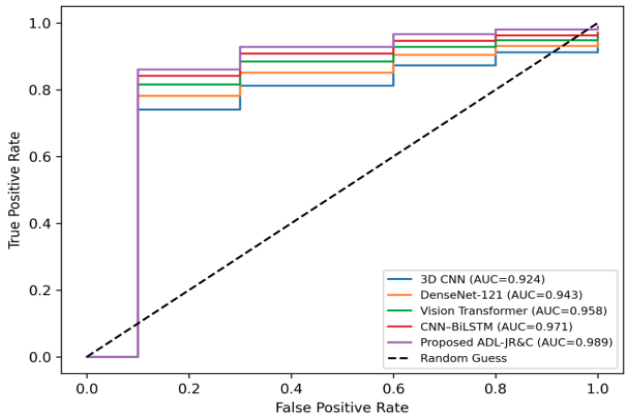



**Figure 9.** Comparison of area under the receiver operating characteristic curve (AUC-ROC) vs. epochs of the proposed and existing systems

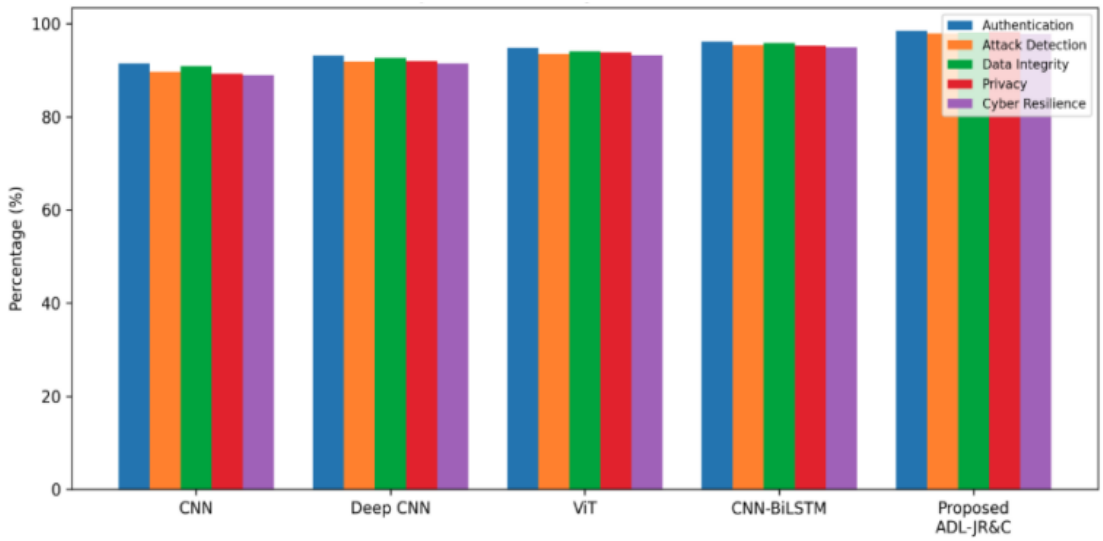



**Figure 10.** Comparison of security metrics of the proposed and existing systems

**Table 5.** Significance analysis test of the proposed and the existing systems

| Method | Accuracy (%) (Mean ± SD) | Sample Size (N) | Repeated Experiments (N) | *t*-Value | *p*-Value | Significance |
|---|---|---|---|---|---|---|
| 3D CNN-based MRI diagnosis | 89.42 ± 1.28 | 5 | 5 | 2.83 | 0.012 | Significant |
| DenseNet-121 with transfer learning | 91.18 ± 1.12 | 5 | 5 | 3.46 | 0.005 | Significant |
| Vision Transformer (ViT)-based framework | 92.36 ± 0.98 | 5 | 5 | 4.12 | 0.002 | Significant |
| CNN–BiLSTM hybrid model | 93.15 ± 0.84 | 5 | 5 | 5.08 | 0.001 | Significant |
| Proposed ADL-JR&C | 94.48 ± 0.53 | 5 | 5 | 7.64 | <0.001 | Highly Significant |

Note: MRI = magnetic resonance imaging; CNN = convolutional neural network; BiLSTM = Bidirectional Long Short-Term Memory; ADL-JR&C = Amalgamated Deep Learning Model for Joint Regression and Diagnosis.

**Table 6.** Comparison of ablation study

| Configuration | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Without DCNN encoder | 89.84 ± 0.91 | 89.36 ± 0.88 | 88.97 ± 0.93 | 89.16 ± 0.90 |
| Without MDAM | 91.57 ± 0.78 | 91.12 ± 0.75 | 90.68 ± 0.81 | 90.90 ± 0.77 |
| Without joint regression module | 92.83 ± 0.65 | 92.41 ± 0.63 | 92.05 ± 0.68 | 92.23 ± 0.64 |
| Without diagnosis module | 93.41 ± 0.59 | 93.06 ± 0.56 | 92.74 ± 0.61 | 92.90 ± 0.58 |
| Proposed ADL-JR&C (complete model) | 94.48 ± 0.48 | 94.12 ± 0.45 | 93.86 ± 0.47 | 93.99 ± 0.46 |

Note: MDAM = Modified Domain Adaptation Method; ADL-JR&C = Amalgamated Deep Learning Model for Joint Regression and Diagnosis; DCNN = deep convolutional neural network.

**Table 7.** Reliability analysis

| Method | Accuracy (%) | Repeated Runs (N) | 95% CI (±%) | External Validation Accuracy (%) |
|---|---|---|---|---|
| 3D CNN | 89.48 | 5 | ±0.72 | 87.96 |
| DenseNet-121 | 91.18 | 5 | ±0.58 | 89.84 |
| Vision Transformer (ViT) | 92.36 | 5 | ±0.51 | 91.12 |
| CNN–BiLSTM | 93.15 | 5 | ±0.46 | 92.03 |
| Proposed ADL-JR&C | 94.48 | 5 | ±0.39 | 93.41 |

Note: CNN = convolutional neural network; BiLSTM = Bidirectional Long Short-Term Memory; ADL-JR&C = Amalgamated Deep Learning Model for Joint Regression and Diagnosis.

Table 6 presents the ablation analysis of the proposed ADL-JR&C framework by removing its major components independently. Removing the DCNN encoder yields the largest performance drop, highlighting the importance of hierarchical feature extraction. Excluding MDAM also decreases accuracy, demonstrating the contribution of domain alignment across heterogeneous MRI datasets. Removing the joint regression or diagnosis components further reduces performance, confirming their complementary roles. The complete ADL-JR&C model achieves the highest accuracy, precision, recall, and F1-score, supporting the integrated architecture's contribution.

Table 7 compares the reliability of the proposed ADL-JR&C framework with four existing methods using accuracy, repeated experimental runs, 95% confidence interval, and external validation accuracy. ADL-JR&C achieves the highest accuracy (94.48%) and external validation accuracy (93.41%), with the lowest confidence interval (±0.39%), indicating comparatively stable performance across repeated experiments.

**Table 8.** Comparative performance and privacy–utility analysis

| Method | DP Setting | Noise Multiplier $\sigma$ | Privacy Budget $\varepsilon$ | Accuracy (%) | Precision (%) |
|---|---|---|---|---|---|
| 3D CNN | Without DP | 0 | ∞ | 89.48 | 88.77 |
| DenseNet-121 | Without DP | 0 | ∞ | 91.18 | 90.74 |
| ViT | Without DP | 0 | ∞ | 92.36 | 91.21 |
| CNN–BiLSTM | Without DP | 0 | ∞ | 93.15 | 92.72 |
| Proposed ADL-JR&C | Without DP | 0 | ∞ | 94.48 | 94.12 |
| Proposed ADL-JR&C | DP–Weak | 0.8 | 8.0 | 94.21 | 93.86 |
| Proposed ADL-JR&C | DP–Moderate | 1.0 | 5.0 | 93.76 | 93.42 |
| Proposed ADL-JR&C | DP–Strong | 1.3 | 3.0 | 92.91 | 92.54 |
| Proposed ADL-JR&C | DP–Very Strong | 1.7 | 1.0 | 91.84 | 91.47 |

Note: CNN = convolutional neural network; BiLSTM = Bidirectional Long Short-Term Memory; ViT = Vision Transformer; DP = differential privacy; ADL-JR&C = Amalgamated Deep Learning Model for Joint Regression and Diagnosis.

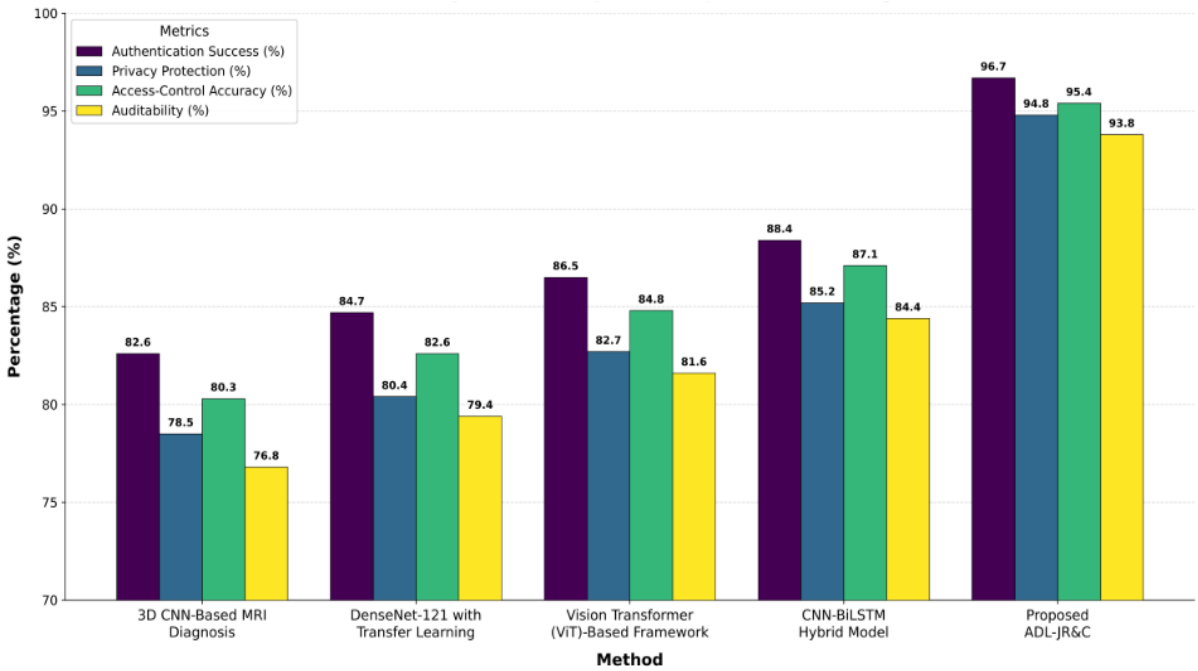



**Figure 11.** Comparison of authentication, privacy, and access-control evaluation of the proposed and existing systems

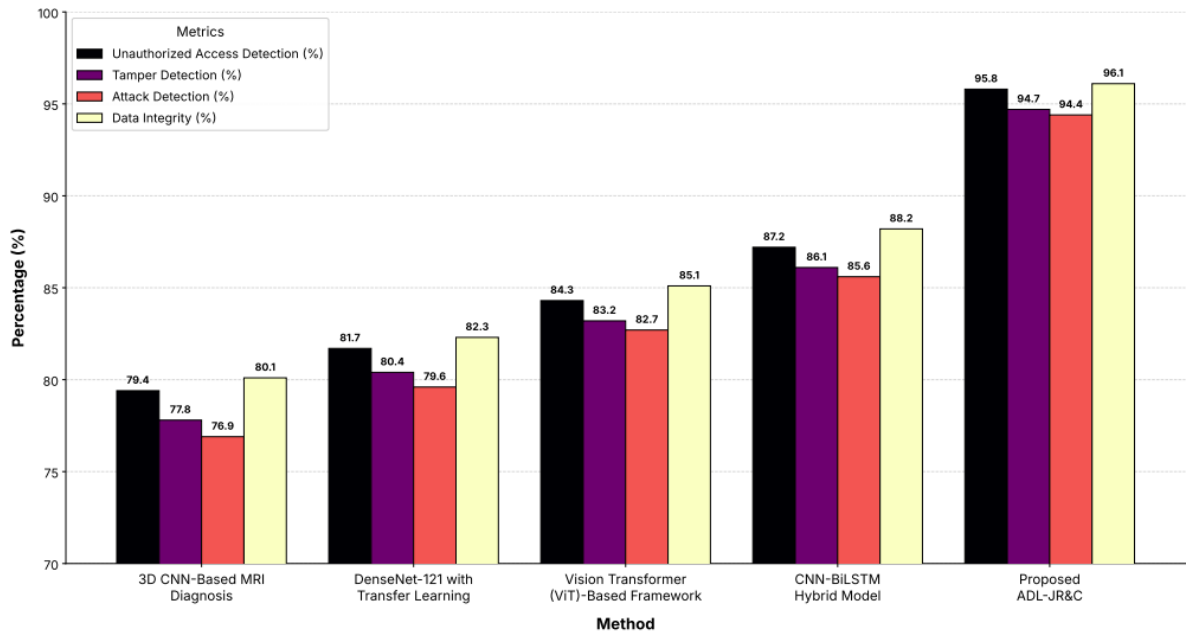



**Figure 12.** Comparison of cyber-resilience and data-integrity evaluation

Table 8 shows the privacy–utility relationship of the proposed ADL-JR&C framework. Without privacy perturbation, ADL-JR&C achieves 94.48% accuracy, 94.12% precision, 93.86% recall, and 93.99% F1-score. As privacy

protection increases, the privacy budget $\varepsilon$ decreases and the Gaussian noise level increases, gradually reducing diagnostic performance. This demonstrates the expected trade-off between privacy preservation and predictive utility. The proposed framework nevertheless maintains competitive performance under stronger privacy constraints.

Figure 11 compares the authentication, privacy, access-control, and auditability performance of the evaluated methods. The proposed ADL-JR&C achieves the highest values across all four security measures, with 96.72% authentication success, 94.86% privacy protection, 95.43% access-control accuracy, and 93.78% auditability. These results indicate improved security performance compared with the four existing MRI-based approaches under the evaluated security scenarios.

Figure 12 compares the cyber-resilience performance of the evaluated methods using unauthorized-access detection, tamper detection, attack detection, and data-integrity measures. The proposed ADL-JR&C achieves the highest performance across all four metrics, reaching approximately 95.84%, 94.72%, 94.38%, and 96.15%, respectively. The results indicate that the integrated security mechanisms provide improved protection against unauthorized access, data tampering, and attack scenarios while maintaining higher data integrity.

## 5. CONCLUSIONS

The proposed framework integrates a DCNN encoder for discriminative MRI feature extraction, ADL-JR&C for simultaneous brain disease diagnosis and clinical score diagnosis, and MDAM for feature alignment across heterogeneous MRI datasets. The study also considered a security mechanism incorporating patient de-identification, encryption, access control, and audit logging to protect sensitive clinical information. Evaluation using the ADNI and Kaggle MRI datasets achieved 94.48% accuracy, 94.12% precision, 93.86% recall, and 93.99% F1-score. Ablation and statistical analyses further demonstrated the contributions of the major framework components. Security-related measures were evaluated within the defined experimental setting; however, they should not be interpreted as evidence of clinical deployment or comprehensive real-world cybersecurity validation.

Future work will focus on evaluation using large-scale, multi-institutional, and independently collected datasets; incorporating multimodal information such as MRI, clinical records, and genetic data; and investigating advanced privacy-preserving approaches, including federated learning and DP. Further studies will also evaluate the framework prospectively with clinical experts and real-world clinical workflows to establish its generalizability, reliability, security, and practical utility for brain disease assessment.

## REFERENCES


[1] Al-Saleh, A., Tejani, G.G., Mishra, S., Sharma, S.K., Mousavirad, S.J. (2025). A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans. Scientific Reports, 15: 23578. https://doi.org/10.1038/s41598-025-07807-8

[2] Perumal, J., Sharma, M., Kumar, T.G. (2026). IoT-driven precision healthcare by optimizing brain tumor diagnosis with machine learning and blockchain. International Journal of Special Education, 41(11s): 1427-1447. https://internationalsped.com/index.php/ijse/article/view/4362.

[3] Lakshmi Vasanthi, K., Sree Darshne, J., Venkatasubbu, P., Ramasubramanian, P. (2026). A federated multimodal deep learning framework for brain tumor classification using MRI. Frontiers in Artificial Intelligence, 9: 1754000. https://doi.org/10.3389/frai.2026.1754000

[4] Hemavathi, U., Durai, S. (2026). Harnessing advanced AI technologies to enhance the diagnosis of Alzheimer's disease. Journal of Multiscale Modelling, 17(02): 2640025. https://doi.org/10.1142/S1756973726400251

[5] Markodimitrakis, E., Trivizakis, E., Ioannidis, G.S., et al. (2026). Federated learning on magnetic resonance imaging: A critical review. Artificial Intelligence Review, 59: 97. https://doi.org/10.1007/s10462-026-11508-7

[6] Ghanta, S., Yarra, K., Kancharla, L., et al. (2026). Enhancing medical diagnostics with federated learning: Secure model aggregation across hospitals. In Conference on Internet of Things, Smart Spaces and Next Generation Networks and Systems, Abu Dhabi, UAE, pp. 328-340. https://doi.org/10.1007/978-3-032-19981-2_26

[7] Abid, M., Benkaddour, M.K., Benouis, M., Khaldi, A., Sahu, M., Sahu, A.K. (2026). Securing federated learning in medical image analysis: A systematic review of privacy threats and defense mechanisms. Computer Modeling in Engineering & Sciences, 148(1): 6. https://doi.org/10.32604/cmes.2026.081055

[8] Jayapandian, J., Senthil, M. (2026). An explainable federated temporal multimodal ensemble deep learning model for early diagnosis and progression diagnosis of Parkinson's and Alzheimer's diseases using X-FedTME-Net. International Journal of Automation and Smart Technology, 16(1). https://doi.org/10.5875/g0qvad40

[9] Karami, M., Kebriaei, H., Ghassemi, F., Azadegan, H. (2026). Byzantine robust federated learning for heterogeneous brain MRI using multisignal gradient fingerprinting and adaptive trust aggregation. Scientific Reports, 16: 25604. https://doi.org/10.1038/s41598-026-55855-5

[10] Sahu, G. (2026). A systematic literature review of deep learning models for medical image analysis. Journal of Machine Learning Innovations and Artificial Intelligence Horizons, 1(1): 174-229. https://doi.org/10.70903/jmliaih.2026.111

[11] Suleman, M.U., Mursaleen, M., Khalil, U., et al. (2026). Integration of AI diagnostic tools into clinical practice for Alzheimer's disease: Barriers and solutions. Annals of Medicine and Surgery, 88(2): 1488-1494. https://doi.org/10.1097/MS9.0000000000004505

[12] Latchoumi, T.P., Ezhilarasi, T.P., Balamurugan, K. (2019). Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 1: 1137. https://doi.org/10.1007/s42452-019-1179-8

[13] Younis, M.C. (2026). Self-aware quantization for privacy-preserving federated learning: Enhancing Alzheimer diagnosis with an adaptive inception-

transformer framework. Vietnam Journal of Computer Science, 1-35. https://doi.org/10.1142/S2196888826500120

[14] Chandrasekaran, S., Nethravathi, B., Cholla, R.R., Babu, G.N., Mahesh, T.R. (2026). Enhanced diagnosis of brain MRI images leveraging edge AI strategies and multi-model ensembles. Neural Computing and Applications, 38: 62. https://doi.org/10.1007/s00521-025-11741-1

[15] Helal, S.M., Al-Khateeb, B. (2026). Explainable federated learning for brain tumor diagnosis using multi-source MRI data. Iraqi Journal for Computers and Informatics, 52(1): 258-277. https://doi.org/10.25195/ijci.v52i1.804

[16] Gawate, A., Barhate, A., Dhanjode, D., Karande, A., Pacharaney, U., Sheikh, A. (2026). Deep learning-driven brain tumor detection and diagnosis from MRI images for enhanced clinical decision support. In 2026 IEEE 1st International Conference on Intelligent Technologies for a Sustainable and Inclusive Future (ICITSIF), Indore, India, pp. 1380-1385. https://doi.org/10.1109/ICITSIF69060.2026.11608405

[17] Hcini, G., Jdey, I. (2026). Trustworthy analysis of medical and biometric data with deep learning: Methods, challenges, and the imperatives of explainability and privacy. Multimedia Tools and Applications, 85: 744. https://doi.org/10.1007/s11042-026-21841-2

[18] Jothe, J.S., Naitam, M.G., Ponmariappan, S., Pandey, A.C. (2026). Generative models in medical imaging for cognitive disease detection. In The Role of Generative AI in Cognitive Diseases Diagnosis, Treatment Planning, and Future Healthcare. Cognitive Systems Monographs, Springer, Cham, pp. 83-123. https://doi.org/10.1007/978-3-032-18782-6_3

[19] Venugopalan, J., Tong, L., Hassanzadeh, H.R., Wang, M.D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific Reports, 11(1): 3254. https://doi.org/10.1038/s41598-020-74399-w

[20] Khojaste-Sarakhsi, M., Haghighi, S.S., Ghomi, S.M.T.F., Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A survey. Artificial Intelligence in Medicine, 130: 102332. https://doi.org/10.1016/j.artmed.2022.102332