Adaptive Neuro-Digital Twin with Cross-Domain Multimodal Representation Learning for Early Alzheimer's Disease Prognosis

Dr V S Krushnasamy[1], Dr. Annapurna Mishra[2], Pratik Gite[3], Ganesh Kumar Anbazhagan[4],
Dr. Adapa Gopi[5], M.Misba[6], Dr. A. Arul Anitha[7], Osama R.Shahin[8]

Associate Professor, Department of Electronics and Instrumentation Engineering,
Dayananda Sagar College of Engineering, Bangalore, India[1]

Associate Professor, Department of Electronics and Communication Engineering, Silicon University, Odisha, India[2]

Associate Professor, Department of Computer Science and Engineering,
Parul University- Parul Institute of Technology, Vadodara, Gujarat, India[3]

Department of Microbiology, Saveetha Medical College and Hospital,
Saveetha Institute of Medical and Technical Sciences, Chennai – 602 105, Tamil Nadu, India[4]

Associate Professor, Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram, Guntur Dist-522302., Andhra Pradesh, India[5]

Assistant Professor, senior grade, Department of Artificial Intelligence and Machine Learning,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, India[6]

Assistant Professor, Department of Computer Applications,
Jayaraj Annapackiam College for Women (Autonomous), Periyakulam, Theni Dt. Tamil Nadu, India[7]

Department of Computer Science, College of Computer and Information Sciences, Jouf University, Saudi Arabia[8a]

Physics and Mathematics Department, Faculty of Engineering, Helwan University, Helwan, Egypt[8b]

Abstract*—Alzheimer's disease (AD) refers to a progressive neurodegenerative disease involving cognitive impairment, brain atrophy, and functional neurological deficits that make early prediction and subsequent disease progression monitoring extremely difficult. Currently available AI methods predominantly focus on employing single modality analysis or static multimodal analysis approaches, which tend to solve AD prognosis as a binary classification problem. Also, much of the current literature on AD does not take into account the importance of progression-aware, patient-wise evaluation, and robust management of multimodality data heterogeneity or missingness. To tackle such issues, this study presents the proposal of an Adaptive Neuro-Digital Twin-based framework for predicting Alzheimer's disease, known as ANDT-AD. The proposed framework incorporates the heterogeneous multimodality information in clinical-cognitive data, structural Magnetic Resonance Imaging (MRI), and Electroencephalography (EEG) signals through modality-wise deep encoders. Specifically, a clinical encoder based on a transformer structure is utilized to capture non-linear cognitive interactions, while a Vision Transformer model and an attention-enhanced temporal EEG encoder model help to extract neuroanatomical information from MRI signals and electrophysiological signals in EEG, respectively. The framework is developed using Python and assessed using public domain clinical, MRIs, and OpenNeuro ds004504 EEG data sets for five-fold cross-validation and patient-level evaluations. Experimentation yielded 98.0% diagnostic accuracy with an AUC of 0.97, which exceeds the performance of current multimodal baselines. Moreover, the framework attained an MAE of 1.12, an RMSE of 1.46, and a progression risk C index of 0.89, proving its strength in predicting cognitive decline and personalizing disease progression models under heterogeneity and missingness of multimodal scenarios.

Keywords*—Alzheimer's disease; neuro-digital twin; multimodal learning; disease prognosis; cross-domain representation learning

I. Introduction

AD is an irreversible, progressive neurodegenerative disorder, which is associated with cognitive decline, memory loss, and functional decline, posing substantial clinical, social, and economic challenges worldwide [1], [2]. Due to its protracted preclinical stage, pathological changes, including synaptic impairment, structural atrophy of the brain, and electrophysiological disorders, tend to be present many years before the appearance of clinical symptoms [3], [4]. Nevertheless, the traditional methods of diagnosis are still predominantly reactive and based on cognitive tests and imaging biomarkers that identify disease when extensive neurodegenerative processes have already been experienced [5], [6]. Early prognosis of AD is also complicated by the heterogeneous and multifactorial nature of this disease because the progression of the disease in people differs significantly. As a result, there is an increasing demand for computational models that can combine heterogeneous neurobiological evidence to facilitate early disease stage detection, as well as model disease progression individually [7], [8]. The current innovations in artificial intelligence have shown potential in AD diagnosis, but most of the current models are unchanging, modality-driven, and incapable of incorporating the dynamic changes in neurodegeneration over time.

The current literature has investigated various AI-based methods of analyzing AD in various data modalities. Existing models [9], [10] have been frequently used on MRI scans to classify multiclass stages of AD with an existing high accuracy, with limited interpretability and generalization. Recent studies [11] have enhanced features like global extraction of features in neuroimaging, but still cannot support pipelines that are not based on imaging. The traditional ML models supported by clinical and cognitive datasets, like Support Vector Machines, Random Forests, and gradient-based ensembles, have been examined as they are unable to learn complex temporal dependencies [12], [13]. Existing EEG studies show sensitivity to functional brain variations, but these studies are not usually coupled to structural and clinical history [14], [15]. Although there have been multimodal fusion techniques based on feature concatenation and late fusion techniques, they generally assume patient-aligned data and do not adapt to heterogeneous data and cross-cohort data. In addition, current models consider AD prediction as a single classification problem, which does not consider the dynamic and progressive nature of neurodegeneration, thus limiting its prognostic value.

To address the limitations of existing studies predicting the occurrence of Alzheimer's disease, the proposed study is an attempt to introduce a Progression-Aware Neuro-Digital Twin modeling framework, which describes the neurological status of a patient as a dynamically changing latent state based on multimodal clinical, MRI, and EEG data. In contrast to traditional diagnostic systems based on the unchanging classification, the given solution is capable of recording longitudinal neurodegenerative dynamics and can diagnose a disease, predict cognitive decline, and estimate progress risk in one single learning system.

A. Research Motivation

The task of early detection of disease onset and development is still considered one of the major clinical challenges due to the progressive nature of neurodegeneration and the variability of the disease presentation among different patients. Machine learning-based techniques have been traditionally concerned with only disease stage identification without addressing the issue of patient-specific changes in the neurological states during the development of the disease. Moreover, heterogeneity of multimodal biomedical information, such as clinical, cognitive measurements, brain structural images, and biosignal recordings, restricts the effectiveness of traditional single-modality predictions. Thus, there is an increasing demand for designing a progression-aware computational paradigm capable of handling heterogeneous multimodal information and providing patient-specific disease evolution monitoring using the Neuro-Digital Twins approach.

B. Research Significance

The modeling of Alzheimer's disease, taking into consideration disease progression, can substantially benefit early diagnosis and continuous monitoring of the disease due to the modeling of temporal changes of neurological status instead of only being able to recognize different stages of the condition. Such a novel approach will enable personalized disease trajectory monitoring and risk-sensitive clinical decision making with the help of combining multimodal biomedical information and Neuro-Digital Twin modeling. Continuous synchronization of patient-specific clinical, imaging, and electrophysiological information will allow us to obtain a correct model of the neurodegenerative processes occurring in an individual. This approach could prove highly useful not only for early intervention but also for future intelligent prognosis systems.

C. Problem Statement

While remarkable advancements have been made recently in the field of artificial intelligence in diagnosing Alzheimer's disease [16], most of the currently developed algorithms still employ traditional single modality learning or static multimodal fusion methodologies that presuppose the completeness and synchronicity of the data. These algorithms cannot account for the nonlinear nature and subject-specificity of the brain degeneration process [17]. In clinical practice, different modalities are often missing, corrupted, or asynchronous, which makes it difficult to deploy existing solutions. The failure to accurately combine heterogeneous clinical, imaging, and electrophysiological data while maintaining robustness under missing-modality further limits practical implementation. This study addresses these shortcomings with a progression-conscious adaptive neuro-digital twin architecture for early prognosis of Alzheimer's disease.

D. Key Contributions

- An Adaptive Neuro-Digital Twin (ANDT-AD) paradigm is proposed for early prognosis of Alzheimer's disease based on continuous modeling of the individual neurodegeneration progression pattern.
- A multimodal representation learning method is proposed to merge various types of clinical data, structural MRI images, and EEG signals into a common disease-oriented latent space.
- Specialized modality-specific encoders such as Transformer-based clinical information learning, Vision Transformer-based MRI feature extraction, and enhanced attention-based EEG representation learning are developed.
- Latent State Evolution with progression awareness is formulated to model disease progression patterns and estimate progression risk.
- Experimental studies using publicly available clinical information, MRI, and EEG databases show that the proposed ANDT-AD achieves significantly better performance than other baseline methods.
- The rest of the study is organized as follows: Section II reviews related works on AI-based AD analysis and highlights existing research gaps; Section III explains the proposed adaptive neuro-digital twin methodology; Section IV presents experimental results; and Section V concludes the study with limitations and future research directions.

II. Related Work

Recent works have investigated the use of advanced machine learning and deep learning methods to enhance the diagnosis and

prognosis of Alzheimer's disease (AD). Gryshchuk et al. [18] examined the efficiency of self-supervised learning in the differentiation of neurodegenerative diseases based on T1-weighted MRI data, where no large-scale labeled data are available. Their contrastive learning-based feature extractor with a linear classification head obtained comparable accuracy in balance to the supervised models. Nonetheless, it was a method that depended on a single imaging mode and binary classification, which restricted its application in multimodal integration and fine disease staging.

Similarly, a multimodal representation learning framework proposed by Dai et al. [19] relies on cross-domain distillation to conduct AD classification. The approach involved the use of a pre-trained text encoder to instruct the learning of structural MRI and PET feature extractors, which included MMSE scores by adding positional embeddings. The model performed highly on the ADNI dataset, though it was only evaluated on one benchmark dataset, and it was not adequate in investigating the longitudinal development of the disease.

There are also multimodal fusion methods or techniques that have been studied to enhance the accuracy of diagnosis. Muksimova et al. [20] proposed FusionNet that combines MRI, PET, and CT data through modality-specific encoders and attention-based fusion layers. The framework had a classification accuracy of about 94% that showed the capability to detect AD at an early stage. However, the data augmentation method based on GAN provides an opportunity to overfit, and it is not externally multi-centered.

Cognitive and electrophysiological data have been studied by other researchers. Saleh et al. [16] developed a hybrid CNN–LSTM model using longitudinal established a hybrid CNN-LSTM model based on longitudinal cognitive scores using the method of Bayesian optimization and explainability, with the use of the SHAP and LIME. Although the model enhanced predictive ability, it omitted modalities of imaging, and in that it used a large amount of computational resources. Similarly, Vo et al. [21] suggested a time-frequency representation of an EEG-based deep learning model to screen non-invasive AD. Although the method obtained more than 80% accuracy, it was sensitive to EEG noise, and its generalization to a wide range of neurodegenerative conditions was low.

There has also arisen digital twin and generative modeling. Amato et al. [22] presented a digital twin-based diagnostic model with EEG-derived biomarkers, which showed encouraging prognostic capabilities, though it needs a large-scale validation. Meanwhile, Sharma et al. [23] suggested a generative model to simulate disease progression using both latent diffusion models and neural ODE, but the algorithm requires a lot of computational power.

Finally, Ali et al. [17] focused on the issue of multimodal learning development and suggested a self-supervised model that enhances cross-cohort generalization and uses missing modalities. The method, however, demands a lot of multimodal pretraining data, and the interpretation is limited. In general, existing literature has shown the promise of deep learning, multimodal fusion, and self-supervised learning in diagnosing and prognosing the disease in the case of Alzheimer's. There are, however, several shortcomings that can be seen in the existing approaches.

III. Proposed ANDT-AD Framework for Early Prognosis of Neurodegenerative Diseases

The suggested framework proposes a novel Progression-Aware Multimodal Neuro-Digital Twin framework that utilizes clinical-cognitive evaluations, structural MRI images, and electroencephalographic biosignals in predicting the personalized progression of Alzheimer's disease. In the first step, each modality is independently preprocessed and encoded, thereby producing structural, functional, and behavioral representations relevant to neurodegeneration. The resulting feature embeddings are subsequently aligned through cross-domain multimodal representation learning to form a robust feature space for handling heterogeneous and missing data. In contrast to the existing methods, where modality fusion or progression analysis is independently carried out, the suggested framework concurrently co-trains multimodal representation learning and progression modeling within one framework. The main innovation in this context involves Neuro-Digital Twin, which enables continuous updating of a personalized neurological twin based on multimodal longitudinal measurements, leading to personalized state transition modeling. The workflow of the proposed framework is illustrated in Fig 1.

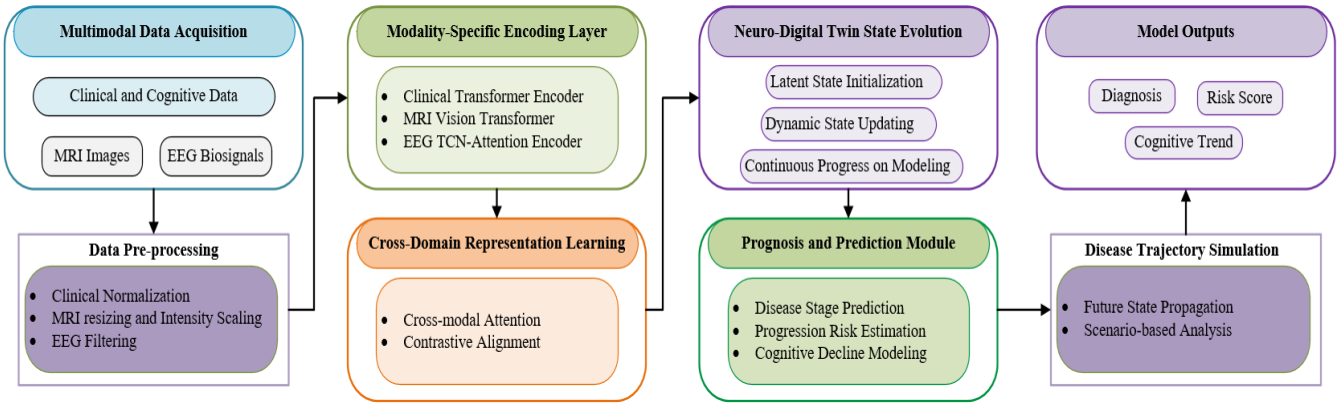

Fig. 1. Overall workflow of proposed ANDT-AD framework.

A. Data Collection and Dataset Description

This study uses three publicly available and supplementary datasets to build a multimodal representation of the AD, which includes clinical, structural, and functional neurological features. The Alzheimer's Disease Dataset [24] It is a clinical and cognitive tabular dataset that is used to offer demographic data, cognitive measuring scores, and clinically significant risk factors that relate to disease development and advancement. The data is an indication of the patterns of high-level cognitive decline and an approximation of regular clinical assessments. The Alzheimer Multiclass Images Dataset [25] that consists of labeled magnetic resonance imaging (MRI) scans of various stages of AD is used to model structural brain changes. With the help of these images, it is possible to extract spatial and morphological biomarkers that are associated with neurodegeneration. Moreover, functional neural dynamics are also measured with the help of the Open Neuro dataset called ds004504 [26] that contains electroencephalography (EEG) measurements that reflect brain activity in terms of time and frequency. The clinical data, MRI data, and EEG data are collected from public datasets and lack full subject correspondence. Consequently, the suggested framework fails to perform patient-wise multimodal fusion. Instead, modality-specific encoders learn separate disease representations that are aligned within a common latent space via cross-domain representation learning. The Neuro-Digital Twin corresponds to a personalized latent neurological state, which undergoes progression-aware state evolution for disease trajectory prediction.

The public Kaggle datasets that were used in this work consist of community-annotated datasets, and the original procedure for clinical annotations of the datasets comes from the dataset providers. The ground-truth annotations were retrieved from the given data set annotations and used as they were. As detailed clinical provenance and annotation protocols are not entirely available for these data sets, this drawback is taken into account when discussing the work. Preprocessing, labeling, quality control, and standard evaluation protocol were used to increase the accuracy of the labels before training the models.

B. Data Preprocessing and Feature Preparation

In order to guarantee consistency across heterogeneous biomedical sources, modality-specific preprocessing pipelines are used to preprocess data from various domains, such as clinical, imaging, and electrophysiological domains, before representation learning. The pre-processing process includes normalization and encoding clinical data, removing noise in the MRI scans, and filtering the EEG data in order to remove any artifacts.

1) Clinical and cognitive data preprocessing: All non-nominal characteristics, such as age, cognitive assessment scores, and indicators that use biomarkers, are standardized using a z-score, which is presented as Eq. (1).

$$x' = \frac{x-\mu}{\sigma} \tag{1}$$

where, μ and σ are the mean and standard deviation of features, respectively. This normalization provides similar scales of features to ensure that high-magnitude variables do not prevail in the process of model training.

2) MRI image preprocessing: The resizing of each scan is done to a constant spatial resolution of 224x224 pixels to ensure that the samples are uniform to one another and that the architecture of the convolutional encoders does not suffer. The min-max scaling is used to normalize the intensity of voxels, averaging to the [0,1] range, and minimize scanner effects on voxel intensity.

3) EEG signal preprocessing: EEG data acquired from the OpenNeuro dataset called "ds004504" are preprocessed to extract clean and useful electrophysiological signal representations. The raw EEG data is filtered using a fourth-order band-pass filter between 0.5 Hz and 45 Hz and a fifth-order notch filter at 50 Hz to eliminate noise, both low- and high-frequencies, as well as power line interference. The ICA technique is used to remove eye movement artifacts and other muscular artifacts in order to preserve brain-related activation patterns. The resulting filtered signals are divided into fixed-length 2-second epochs with 50 percent overlap in order to reflect short-term time changes. The Short-Time Fourier Transform (STFT) method is used to obtain spectral-temporal feature maps for deep learning tasks. Patient-wise stratified sampling is performed in order to avoid a potential data leakage issue where all instances from the same patient fall into only one split, with a ratio of 70 percent for training, 15 percent for validation, and 15 percent for testing.

4) Rationale for selection of encoders: The Transformer model is chosen for modeling clinical features because of its capability of capturing complicated interactions between heterogeneous clinical features by using the self-attention mechanism. ViT is selected for MRI feature extraction as it can learn the global spatial dependency among brain regions, which is crucial for detecting distributed neurodegeneration. TCN is chosen for EEG feature extraction because of its capability of capturing long-range temporal dependency with reduced computational overhead compared to recurrent neural networks. In comparison with the traditional CNN, LSTM, and machine learning models, the chosen encoders show better capability in modeling multimodal features with computational efficiency.

C. Modality-Specific Encoding Architecture

The ANDT-AD is developed as a multimodal latent state-space learning problem. Let the multimodal data be given as in Eq. (2).

$$\mathcal{D} = \{D_c, D_m, D_e\} \tag{2}$$

where, D_c, D_m, and D_e denote the clinical–cognitive dataset, structural MRI dataset, and EEG bio signal dataset, respectively. Both modalities give heterogeneous observations of complementary characteristics of neurological degeneration. The framework aims at learning a parameterized mapping as in Eq. (3).

$$f_\theta: \mathcal{X} \rightarrow \{y_{stage}, r_{risk}, c_{future}\} \tag{3}$$

where, $\mathcal{X}$ denotes multimodal neurological observations, y_{stage} represents disease stage classification, r_{risk} denotes progression risk estimation, and c_{future}corresponds to predicted future cognitive decline. Disease progression is modelled using a persistent latent neuro-digital twin state $s_t \in \mathbb{R}^d$, where s_t represents the hidden neurological condition of a subject at the update step t, and d denotes the latent dimensionality. The proposed Adaptive Neuro-Digital Twin Framework is given in Fig. 2.

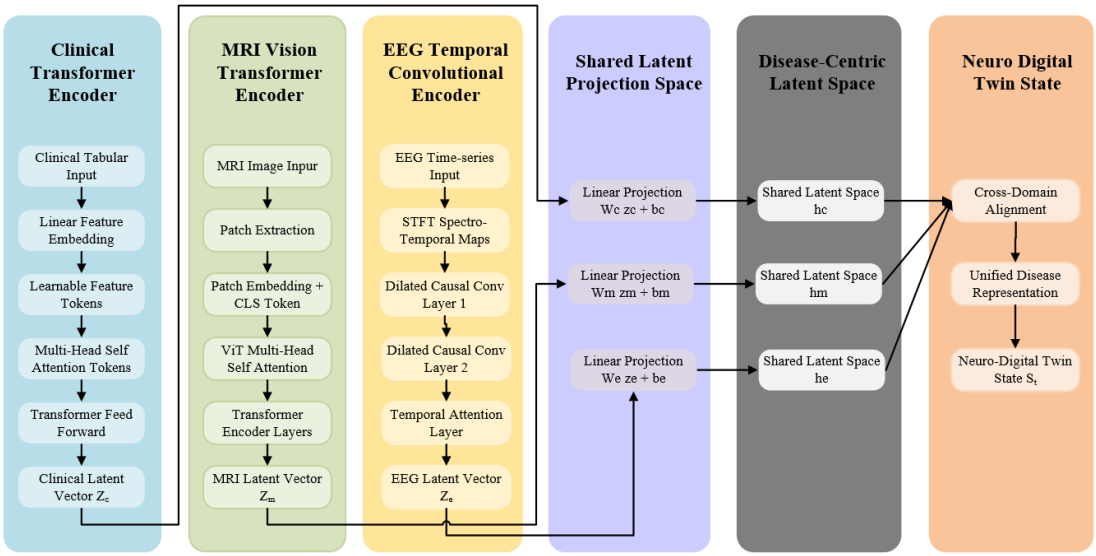

Fig. 2. ANDT-AD framework.

1) Transformer-based encoding of clinical and cognitive data: The clinical and cognitive tabular representations of the patients' data contain vital manifestations of the disease, such as demographic risk factors, neuropsychological evaluation scores, and cognitive impairments. Under the current model architecture, these tabular representations undergo processing using a tabular transformer network encoder that effectively learns the interaction between features in a non-linear manner, an aspect that cannot be achieved by traditional fully connected networks. To begin with, the patient records are encoded as a fixed feature vector representation and projected onto the latent embedding space using a linear transformation. Unlike temporal position encodings, feature embeddings that help retain the identity and meaning of the feature are used. Self-attention techniques are then applied to the embeddings to model context dependencies in clinical attributes and learn their interaction. This can be represented mathematically as shown below:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right)V \quad (4)$$

In Eq. (4), Q, K, and V denote query and key matrices, as well as value matrices, of clinical feature embeddings, and d is an embedding dimension. The encoder produces a very small latent representation of clinically significant patterns related to the progression of Alzheimer's through stacked transformer layers. The performance of this encoder will be an abstracted clinical disease profile, focusing on the severity of risks and cognitive impairment instead of crude clinical values.

Let the multimodal patient input be represented as $X = \{X_c, X_m, X_e\}$, where X_c, X_m, and $X_e$denote clinical records, MRI images, and EEG signals, respectively, providing complementary information for Alzheimer's disease prognosis. The modality-specific embeddings are extracted as $H_c = f_c(X_c)$, $H_m = f_m(X_m)$, and $H_e = f_e(X_e)$, where $f_c(\cdot)$, $f_m(\cdot)$, and $f_e(\cdot)$represent Transformer-based clinical encoding, Vision Transformer-based MRI feature learning, and attention-enhanced temporal EEG modelling. These learned features are fused into a unified latent representation as $Z = \alpha H_c + \beta H_m + \gamma H_e$, where α, β, and γ are adaptive modality weights satisfying $\alpha + \beta + \gamma = 1$. This weighting mechanism enables the framework to prioritize the most informative modality according to patient condition and data quality. The continuously evolving neuro-digital twin state is updated as $S_t = \phi(S_{t-1}, Z_t)$, where $S_t$and S_{t-1}denote the current and previous disease states, $Z_t$is the fused representation at time t, and $\phi(\cdot)$is the nonlinear transition function. Finally, the total optimization objective is defined as $L = L_{cls} + \lambda_1 L_{prog} + \lambda_2 L_{align}$, where L_{cls} denotes diagnostic classification loss, L_{prog} represents disease progression prediction loss, L_{align} ensures cross-modal embedding consistency, and λ_1, λ_2 are balancing coefficients controlling each objective during training.

2) ViT encoding of structural MRI data: MRI images offer anatomical details that can be associated with the progression of Alzheimer's disease, such as cortical atrophy, hippocampus atrophy, and enlarged ventricles. Under this framework, MRI images are subjected to a Vision Transformer (ViT) network architecture in order to learn the global neuroanatomical representation. The individual preprocessed MRI image is subdivided into patches without overlapping, which are flattened and mapped into a high-dimensional space by a

learned linear projection. A classification token is included in the patch sequence to learn the global structure of the disease.

Unlike conventional convolutional networks that primarily capture local spatial hierarchies, the Vision Transformer (ViT) models long-range spatial dependencies through patch-wise self-attention mechanisms. This capability is particularly important in Alzheimer's disease analysis, where distributed brain regions collectively contribute to cognitive impairment. The generated patch embeddings are passed through multiple transformer encoder layers, enabling the model to learn relationships between anatomically distant brain structures. The final representation of the learnable classification token serves as a global feature embedding that summarizes disease-relevant neuroanatomical patterns across the entire brain. The output embedding can be in formally written form as Eq. (5).

$$z_{\text{MRI}} = f_{\text{ViT}}(I) \tag{5}$$

where, I denote the input MRI scan and $f_{\text{ViT}}(\cdot)$ represents the ViT mapping. This latent representation describes neurodegenerative trends at the macro-level and can be considered the structural aspect of the digital twin state.

3) Temporal convolutional encoding of EEG biosignals: To capture the temporal complexity of EEG data, the current study uses a TCN with an attention mechanism. The EEG preprocessed signals (as spectro-temporal feature matrices) are input into a series of expanded causal convolutional layers. The dilated convolution operation is defined as Eq. (6).

$$y(t) = \sum_{k=0}^{K-1} w(k)\, x(t - d \cdot k) \tag{6}$$

where, $x(t)$ being the input signal, $w(k)$ being the convolutional kernel, dis being the dilation factor, and K being the size of the kernel. The TCN is capable of capturing the short-term oscillatory and long-term temporal interdependences with cognitive deterioration by successively enhancing the rate of dilation through the layers. A subsequent attention layer is utilized to put weight on temporally informative portions of the EEG signal to enable the encoder to focus on disease-related neural evolving motions and reduce remaining noise. The resulting EEG latent representation imprints functional electrophysiological features signifying the Alzheimer's pathology.

4) Projection into a shared latent representation space: After modality-specific encoding, the clinical transformer encoder, MRI ViT, and EEG TCN outputs are all functionally transformed into a common latent dimensional space via linear transformation layers. This projection provides a dimensional consistency across modalities and cross-domain alignment. Where "," and "are" used to represent latent embeddings of clinical, MRI, and EEG encoders, respectively. Every embedding is converted to be as Eq. (7).$z_c z_m z_e$

$$h_i = W_i z_i + b_i, i \in \{c, m, e\} \tag{7}$$

where, z_i is the modality-specific latent embedding, W_i is the learnable projection matrix, b_i is the bias vector, and h_i is the projected shared latent representation. The projection parameters W_i and b_i are jointly optimized with encoder parameters through end-to-end backpropagation during training. With this joint optimization, the modality-specific embeddings are converted to a common disease-centric latent representation space while preserving discriminative information.

This latent space can be considered as the state representation of the Neuro-Digital Twin, and it allows incorporating different neurological information into a single format for modeling disease progression. This method is particularly suitable for learning modality-robust representations, as it requires that each encoder learns useful information from its input even when some modalities are not available. In other words, this approach provides resilience to missing-modality cases, cross-cohort differences, and heterogeneous clinical data distributions, making it more generalized and robust in practice.

D. Cross-Domain Representation Learning

Following modality-specific encoding, the proposed model uses a cross-modality representation learning technique that transforms heterogeneous representations into a single disease-aware latent space. The main task of this step is to combine the information derived from both clinical and neuroimaging features without depending on any modality in a way that ensures cross-cohort robustness. Unlike conventional methods that consider the explicit correspondence between individuals, the proposed model does alignment at the representation level, such that embeddings from various modalities can be aligned in a shared latent space by leveraging their similarity in terms of Alzheimer's disease.

The modality weights (α, β, γ) are automatically learned by the system in the process of training, via a learnable fusion layer which uses an attention mechanism. These modality weights are computed by using trainable parameters, and then normalization is done using the softmax function in such a way that the weights satisfy the condition $\alpha + \beta + \gamma = 1$. Backpropagation helps in updating the parameters of the fusion layer using the gradients coming from the optimization loss in the total process. The contribution of each representation can be changed based on its importance for prediction tasks.

Latent embeddings produced by the clinical transformer encoder, MRI ViT, and EEG TCN are initially processed as updated independent semantic tokens of various neurological viewpoints. The model is then used to make cross-modal attention between these embeddings and, based on the relevance of the modality to a disease, actively reweight the contribution of these modalities. The cross-modal attention of a given modality embedding h_i is calculated in terms of other modality embeddings as Eq. (8).

$$\tilde{h}_i = \sum_{j \neq i} \alpha_{ij}\, h_j \tag{8}$$

The attention coefficients are calculated α_{ij} by means of scaled dot-product attention. They reflect how similar in terms of a disease the different modes are. Cross-modal representation learning seeks to build modality-invariant representations that will encode the pathology shared between the manifestations of the same condition. For that purpose, alignment is done via cross-modal attention and contrastive learning.

In order to achieve domain invariance and lessen modality-specific bias, contrastive learning is introduced as an auxiliary alignment goal. The pairs of positive and negative pairs are formed by connecting embeddings corresponding to similar stages of the disease across different modalities and different stages, respectively. The contrastive loss is considered to be Eq. (9).

$$\mathcal{L}_{\text{con}} = -\log \frac{\exp(\text{sim}(h_i,h_j)/\tau)}{\sum_k \exp(\text{sim}(h_i,h_k)/\tau)} \quad (9)$$

where, $\text{sim}(\cdot)$ denotes cosine similarity between embeddings, h_i and h_j represent embeddings from different modalities, h_k denotes negative samples, τ is the temperature parameter controlling similarity scaling, and L_{con} represents contrastive alignment loss. This loss promotes embeddings similar to the same states of AD to cluster near one another regardless of their originating modality or dataset, and separate dissimilar states.

To deal with incomplete data in the real world, a modality dropout technique was applied during training, in which some or all modalities would be masked at random. It ensured that the model was able to learn solid cross-modal relationships and still achieve good predictions despite the absence of MRI, EEG, or clinical information.

1) Progression-aware neuro-digital twin modeling: The proposed framework also makes use of a Progression-Aware Neuro-Digital Twin modeling scheme, which allows us to track the dynamics of the progression process of Alzheimer's disease. As part of this methodology, the neurological condition of the patients will be modeled via latent twin states based on the aligned embeddings of the multimodal data (clinical, MRI, and EEG). This differs from traditional approaches to classification, wherein the different states associated with disease progression were treated categorically; however, the Progression-Aware Neuro-Digital Twin model will evolve, making it capable of tracking gradual changes to the neurological conditions of the patients. It was expressed in Eq. (10).

$$z_{t+1} = z_t + f(z_t, \Delta t; \theta) \quad (10)$$

where, z_t is the current twin latent state, $z_{t+1}$is a future twin state, $f(\cdot)$ is a progression dynamics function, Δt is a time interval and θ is the model parameter.

E. Neuro-Digital Twin State Evolution

The suggested Neuro-Digital Twin can be considered as a dynamic latent representation of the neurological condition of a patient rather than a static model for disease classification. The initial state of the digital twin is defined using multimodal feature representations, is evolved using temporal state transition, and is validated through diagnosis and prognosis. Adaptation to the individual patient is done using continuous updates of the latent state to reflect the progression of disease over time.

The proposed Progression-Aware Neuro-Digital Twin models Alzheimer's disease progression as a continuous latent trajectory rather than a fixed sequence of disease stages. This enables the framework to capture gradual neurological changes and individual progression patterns. The evolving latent state integrates multimodal representations and provides a dynamic basis for progression risk estimation.

1) Latent state initialization: The latent embedding of the clinical and cognitive data is used to initialize the neuro-digital twin state since these aspects are the most consistent and early predictors of the risk of AD. Let represent the projected clinical-cognitive embedding that results from the modality-specific encoder. The state of the digital twin is initialized with the help of this embedding.$h_c \in \mathbb{R}^d$

$$s_0 = \phi(h_c) \quad (11)$$

In Eq. (11), $\phi(\cdot)$ is a linear transformation, which takes the clinical embedding into the twin state space. This setup is necessary to ensure that the digital twin is initialized with a cognitively based model of disease status, i.e., baseline risk factors and early cognitive impairment, before severe structural or functional loss.

2) Dynamic state updating with multimodal fusion: With further multimodal measurements being accessible, such as structural MRI and EEG-based embeddings, the state of a neuro-digital twin gets revised in a recurring process of state changes. The resulting cross-domain representation learning obtained as the aligned and fused multimodal representation at update is added to the current twin state. The update of the state is determined by Eq. (12).$z_t ts_{t-1}$

$$s_t = \sigma(W_s s_{t-1} + W_z z_t + b) \quad (12)$$

where, W_s and W_z can be learnt, b is a bias value, and $\sigma(\cdot)$ is a nonlinear activation function. This formulation allows the twin state to incorporate novel evidence together with historical context on diseases. The update state is designed so that the changing temporary modality-specific observations are not replaced by the cumulative disease representation and thus encourages temporal stability and awareness of progression.

3) Continuous disease progression modeling: The dynamics of the neuro-digital twin state are developed to reproduce the process of a continuous disease progression. Instead of discrete stage classification, the overt state transitions carry the gradual degradation of the nervous system in the case of AD. In order to implement a smooth temporal evolution, a progression regularization term is added, and it is defined as Eq. (13).

$$\mathcal{L}_{\text{prog}} = \| s_t - s_{t-1} \|_2^2 \quad (13)$$

This limitation promotes an incremental state dynamic that is realistic of neurodegenerative dynamics. The latent state is going to follow a trajectory in the disease-centric embedding space, with the movement and the rate of change.

The diagnosis and prognosis prediction module converts the changing state of Neuro-Digital Twin to meaningful clinical prediction by using the joint prediction framework of diagnostic and prognostic problems. The prediction task is carried out based on the latent representation of Neuro-Digital Twin, which means all the predictions will be made with the consideration of

neurological information that reflects progression and development of diseases, rather than the features from individual modalities. By constantly updating the state of the Neuro-Digital Twin with multimodal information, it can achieve early detection and risk prediction.

4) Early disease stage and progression risk prediction: The neuro-digital twin state is applied to evaluate the current level of the AD at every update step and the risk of disease progression. The head of classification compares the latent state to discrete disease stages, such as cognitively normal, mild cognitive impairment, and AD. This mapping is defined as Eq. (14).$s_t t$

$$\hat{y}_t = \text{softmax}(W_y s_t + b_y) \quad (14)$$

where, W_y and b_y are parameters of learning. Simultaneously, a head of risk estimation can determine a continuous progression risk score, which is an indicator of the possibility of disease progression within a future horizon. The risk score is obtained as Eq. (15).

$$\hat{r}_t = \sigma(W_r s_t + b_r) \quad (15)$$

where, $\sigma(\cdot)$ is the sigmoid activation function. Collectively, these products offer a probabilistic estimate of the existing disease condition and the susceptibility to progression.

5) Cognitive decline trend estimation: To model longitudinal cognitive decline, the framework uses a regression head that predicts future trends in the cognitive scores, given the latent twin state. The equation of the anticipated cognitive path can be written as Eq. (16).

$$\hat{c}_{t+\Delta} = W_c s_t + b_c \quad (16)$$

where, Δ represents the time ahead of prediction. This formulation enables the model to predict the patterns of gradual cognitive decline as opposed to point-wise scores, which provide an insight into the rate of deterioration that may be expected. The regression result is the direct representation of the impact of structural and functional non-brain alterations coded in the twin state. Categorical cross-entropy loss is used to optimize the classification head, binary cross-entropy loss is used to predict progression risk, and MSE loss is used to predict cognitive decline to represent continuous regression behavior.

F. Multi-Task Optimization and Simulation-Based Analysis

The process of diagnosis, risk prediction, and cognitive trend prediction is optimized together via the multi-task learning approach. When simulating the disease trajectory based on its progression analysis, the Neuro-Digital Twin's state is forwardly propagated via the state evolution function. As a result, it becomes possible for the proposed approach to model the dynamics of Alzheimer's disease progression regardless of the presence of observations due to the updating of the latent neurological state. The total loss function is given as Eq. (17).

$$L_{total} = \lambda_1 L_{cls} + \lambda_2 L_{risk} + \lambda_3 L_{reg} + \lambda_4 L_{con} + \lambda_5 L_{prog} \quad (17)$$

where, L_{cls} is the disease stage classification loss, L_{risk} is the progression risk estimation loss, L_{reg} is the cognitive decline regression loss, L_{con} is the contrastive alignment loss, L_{prog} is the progression regularization loss, and λ_i is the task balancing coefficients. In order to improve interpretability in a clinical context, the explainable AI methods have been incorporated into the suggested architecture. The SHAP method has been used to measure the significance of clinical features, including the patient's age, MMSE value, and cognitive measures. The Grad-CAM technique has been used to detect the key brain regions from MRI scans. Moreover, temporal attention maps from EEG signals helped detect significant neurological patterns that influenced the prediction of prognosis.

Algorithm 1: Adaptive Neuro-Digital Twin Framework for Alzheimer's Disease Prognosis

Input:
Multimodal data $D = \{X_c, X_m, X_e\}$
Output:
Disease stage prediction Y_s, progression risk Y_r, cognitive decline estimate Y_c
Step 1: Input multimodal datasets (Xc, Xm, Xe)
Step 2: Preprocess each modality (normalization, filtering, feature preparation)
Step 3: Encode modalities:
Zc ← Clinical Encoder(Xc)
Zm ← MRI Encoder(Xm)
Ze ← EEG Encoder(Xe)
Step 4: Learn shared representation:
Z ← Cross-Domain Fusion(Zc, Zm, Ze)
Progression-Aware Neuro-Digital Twin Learning
Step 5: Update neuro-digital twin state:
St ← Temporal Update(St−1, Z)
Step 6: Predict outcomes:
Ys ← Disease Stage Classifier(St)
Yr ← Progression Risk Predictor(St)
Yc ← Cognitive Decline Regressor(St)
Step 7: Compute joint loss and update model parameters
Step 8: Output (Ys, Yr, Yc)

The entire training and inference process of the proposed ANDT-AD framework is outlined in Algorithm 1. The algorithm comprises encoding the heterogeneous neurological modalities with modality-specific encoders, projection into a common latent space, and cross-domain representation alignment. The aligned embeddings are used to learn a transition function, which is applied to the neuro-digital twin state with an iterative strategy. During inferences, the trained transition dynamics allow forward simulation of disease dynamics to perform prognosis.

ANDT-AD model proposes a novel framework of multimodal progression modeling using clinical, MRI, and EEG data under the Neuro-Digital Twin paradigm for continuous Alzheimer's disease prognosis. In contrast to the existing research works that mainly focus on static classification and multimodal fusion techniques, the proposed framework makes use of cross-modal representation learning to learn modality-invariant latent representations and integrate heterogeneous biomedical data into a common disease-aware feature space. Moreover, the proposed framework uses dynamic Neuro-Digital Twin state evolution for the continuous modeling of disease progression rather than the prediction of individual disease stages. The usage of a clinical Transformer encoder for clinical

features, a vision transformer model for MRI analysis, and an attention model for EEG signals within a unified multitask learning framework further distinguishes the proposed framework from existing multimodal and generative Alzheimer's disease prediction frameworks.

The ANDT-AD framework has several methodological strengths compared to current artificial intelligence techniques employed for predicting the presence of Alzheimer's disease. Unlike other multimodal methods that are based on static fusion or alignment of multimodal data features, the proposed framework uses an adaptive Neuro-Digital Twin representation that learns from multimodal data evidence, including clinical, MRI, and EEG features. The cross-modal representation learning mechanism ensures robust and efficient knowledge transfer between different modalities and is able to address the problem of missing modalities present in the majority of the earlier works. Additionally, the model can perform joint regression and classification learning to estimate disease stage, predict risk of progression, and forecast cognitive decline trajectory. The application of temporal state update makes the proposed approach a progression-aware predictive tool able to account for disease dynamics in each patient. Thus, the introduced framework is an example of a new scalable Neuro-Digital Twin approach that could be applied for personalized disease trajectory analysis.

IV. Result and Discussion

This was carried out through cross-cohort multi-modal analysis involving clinical-cognitive measurements, structural brain imaging, and electroencephalography biosignals. The model developed was intended to be efficient enough to run in real-time clinical settings, taking less than one second on average to predict patient outcomes with the help of GPU optimization. The compact structure of this framework allows it to easily integrate into hospital edge computing devices as well as cloud-based healthcare services.

A. Experimental Setup

Experimental settings are presented here, which are employed to rigorously evaluate the proposed framework. Cross-cohort multimodal data, consisting of clinical-cognitive measurements, structural brain imaging through MRI, and EEG biosignals, were used for evaluation purposes. The proposed architecture is developed in Python, while batch normalization, dropout, and early stopping methods were considered in order to mitigate overfitting. In addition, a five-fold cross-validation approach was used for performance evaluation. Both complete and incomplete cases are considered, and the framework is evaluated based on diagnostic, prognostic, and regression criteria compared against state-of-the-art unimodal and multimodal baselines. The simulation and hardware parameters were given in Table I.

1) Missing modality handling and data splitting strategy: To evaluate the robustness of the proposed framework under incomplete data conditions, different missing-modality scenarios were considered, including clinical-only, MRI-only, EEG-only, and partial multimodal combinations. The model performance was evaluated by removing individual modalities while maintaining the shared latent representation learning capability. For data partitioning, patient-wise splitting was applied to ensure that samples from the same subject were not distributed across training, validation, and testing sets. This strategy prevents data leakage and provides a reliable assessment of the model's generalization capability.

TABLE I. Simulation and Hardware Setup

Parameter	Configuration
Programming framework	Python with PyTorch
DL libraries	PyTorch Lightning, NumPy
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	16
GPU	NVIDIA RTX 3080 (10 GB)
CPU	Intel Core i9
RAM	32 GB
Training and inference	GPU-accelerated

TABLE II. Model Architecture and Implementation Parameters

Parameter	Configuration
Latent Embedding Dimension	256
Clinical Encoder	Transformer-based encoder (4 layers, 8 attention heads)
MRI Encoder	Vision Transformer (ViT) with 16 × 16 patch size
EEG Encoder	Temporal Convolutional Network (TCN) with dilation rates (1, 2, 4, 8)
Multimodal Fusion Method	Cross-modal attention-based fusion
Dropout Rate	0.2
Optimization Settings	Adam optimizer with learning rate 1×10^{-4}

The implementation parameters for the suggested framework are listed in Table II. Latent embedding dimension refers to the common feature space utilized by multimodal representation learning. Transformers, Vision Transformers (ViT), and Temporal Convolutional Network (TCN) are employed as feature extractors of clinical, MRI, and EEG modality, respectively. The cross-modal attention network incorporates the representations obtained through transformers, ViT, and TCN. Finally, dropout is applied to mitigate overfitting, and Adam with a learning rate of 1×10^{-4} facilitates efficient training.

B. Multimodal Encoding Performance Analysis

The efficiency of the domain-specific encoders utilized by the proposed framework is assessed before proceeding with cross-domain alignment. The analysis evaluates the ability of each encoder in capturing disease-related information within its corresponding data stream and discriminating between different Alzheimer's disease stages within the learned latent spaces. The chosen encoder network architectures are additionally supported by a comparative performance assessment and visualization of the learned latent space, ensuring solid ground for future multimodal integration and Neuro-Digital Twin development. The findings correspond well with clinical knowledge,

suggesting overlapping characteristics of early Alzheimer's disease within cognitive and neurological domains.

C. Cross-Domain Representation Alignment Results

In this sub-section, an assessment of the efficacy of the designed cross-modal representation learning paradigm for aligning heterogeneous modality-specific embeddings within a unified latent space that represents diseases is presented. It is analyzed in terms of how invariant across domains the aligned latent representations turn out to be, the stability of their acquisition during the contrastive optimization process, and, consequently, their improved discriminative power with respect to the considered diseases. In the case of Alzheimer's disease, it is confirmed that the latent representations are indeed effective.

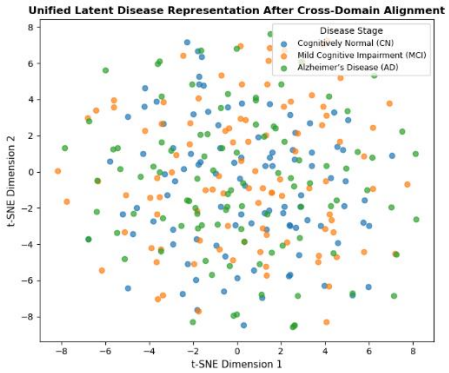

Fig. 3. t-SNE visualization of unified latent space cross-domain alignment.

Fig. 3 shows the t-SNE projection of the unified latent representations obtained following the implementation of the suggested cross-domain alignment mechanism on a two-dimensional projection. The aligned latent space has significantly more pronounced separation between cognitively normal, mild cognitive impairment, and AD categories, and less overlap between the disease stages. This increased clustering is a sign that modality-specific biases were effectively reduced and that the learned representation represents disease-centric features not affected by the data source. This alignment results in a consistent latent disease space that forms a predictable basis of downstream digital twin state evolution and prognostic model, and it will be able to have consistent interpretation across different modalities.

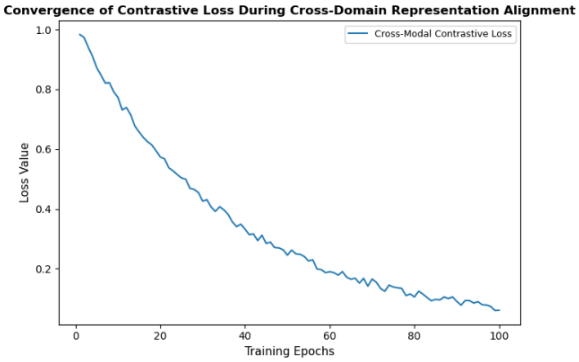

Fig. 4. Contrastive loss convergence and alignment stability.

Fig. 4 shows the convergence pattern of the cross-modal contrastive loss when training the representation alignment module. The decay rate through the epochs is smooth and stable, and this indicates that the optimization and the reduction of inter-modality discrepancies in the latent space are effective. The lack of oscillatory or divergent trends means that the training process is stable and well-built.

D. Progression-Aware Neuro-Digital Twin State Evolution Analysis

The suggested approach proposes a Progression-Aware Neuro-Digital Twin modeling technique where the state of the neurological system is modeled as a constantly evolving latent state. While traditional classification models estimate a discrete state of the disease, the Neuro-Digital Twin model applies temporal state modeling where the neurological state evolves. With this approach, the progression of the disease can be considered for modeling Alzheimer's disease dynamics and generating personalized disease trajectories.

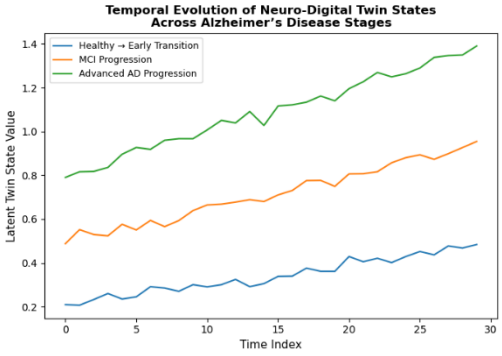

Fig. 5. Temporal evolution of neuro-digital twin states.

Fig. 5 visualizes the dynamic change of the latent states of the Neuro-Digital Twin throughout time in the presence of various states of the AD development. The continuous and slowly rising curves are indicative of the fact that the model is capable of representing smooth neurodegenerative changes as opposed to stage transitions. The unique developmental patterns in healthy, mild cognitive impairment, and Alzheimer's states are evidence that the digital twin stores the severity of diseases in a time-consistent way.

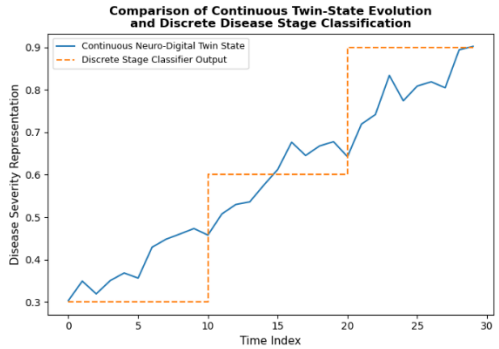

Fig. 6. Continuous twin-state transition vs discrete classification.

Fig. 6 contrasts the Neuro-Digital Twin state development suggested with the results of a traditional discrete-stage classifier. The continuous twin curve has smooth transitions, which are consistent with the smooth process of AD progression, and the discrete classifier creates a sudden step

transition. Such a breakdown is an indication of the weakness of the statistical classification methods in the model of longitudinal disease.

E. Prognosis and Cognitive Decline Prediction Performance

This sub-section provides an analysis of the diagnostic performance of the Neuro-Digital Twin architecture that extends the concept of prognosis beyond the scope of conventional diagnostic classification. The analysis will center on the prediction of trends in cognitive decline longitudinally within the framework of multi-task learning for optimization of diagnostic and prognostic functions. Accuracy and AUC will be used as metrics for diagnostic assessment, whereas prognostic evaluation will consider the prediction of disease trajectory.

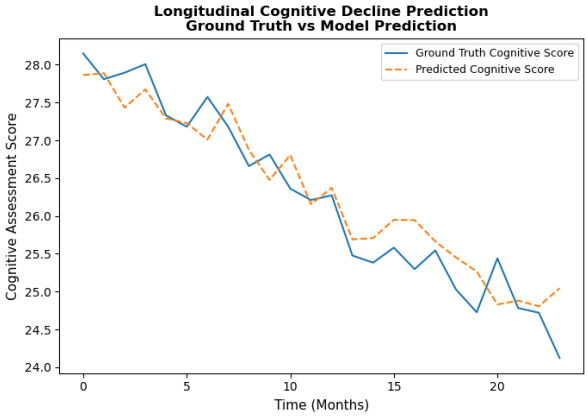

Fig. 7. Longitudinal cognitive score prediction vs ground truth.

Comparison of the ground truth cognitive test scores with the longitudinal prediction scores made by the suggested Neuro-Digital Twin approach can be seen in Fig. 7. Good correlation between the predicted scores and actual cognitive changes indicates the ability of the model to detect gradual change in cognition and not merely detect disease states at different time points. It is evident from the results that there is a good ability of the proposed approach for predicting Alzheimer's disease in the early stages of the disease.

TABLE III. QUANTITATIVE PERFORMANCE OF DIAGNOSTIC AND PROGNOSTIC TASKS

Task	Metric	Proposed ANDT Framework	Best Baseline
Disease stage diagnosis	Accuracy (%)	98.0	92.1
Disease stage diagnosis	AUC	0.97	0.93
Cognitive score prediction	MAE	1.12	1.85
Cognitive score prediction	RMSE	1.46	2.10
Progression risk estimation	C-index	0.89	0.81

The framework performance is quantitatively summarized in Table III by applying it to the diagnostic classification, cognitive decline prediction, and the progression risk estimation tasks. Its persistently high accuracy and AUC values show that it is reliable in disease stage discrimination, whereas lower MAE and RMSE show that it is more accurate in predicting cognitive scores.

F. Disease Trajectory Simulation and Risk Forecasting

This section assesses the capacity of the proposed Neuro-Digital Twin system to predict the future development of Alzheimer's disease. In particular, the proposed framework provides personalized predictions of Alzheimer's disease progression by simulating future developments using the learnt twin latent states through the time-based state evolution process. The model utilizes these future developments to predict future progression of the disease, as well as provide future neurodegeneration predictions.

G. Ablation Study

In this section, the contribution of each major architectural component in the proposed framework is strictly evaluated using a controlled ablation analysis. The study quantifies the effect of each of the four important modules, cross-domain alignment, Neuro-Digital Twin evolution, EEG modality, and contrastive learning, by systematic removal of each one of them in diagnostic performance.

The process of ablation analysis has been additionally applied for the evaluation of the contribution of each of the main components of the framework that was introduced in this study. Alongside the assessment of diagnostic accuracy, the contribution of each module regarding the risk prediction of disease progression is also assessed through the evaluation of changes in the representation of Neuro-Digital Twin states and the ability to estimate the progression of the disease. The absence of cross-domain alignment is used for the evaluation of the importance of multimodal feature synchronization, while the exclusion of Neuro-Digital Twin evolution is evaluated as the impact of temporal disease progression modeling.

TABLE IV. ABLATION STUDY OF KEY ARCHITECTURAL COMPONENTS

Configuration	Accuracy (%)	AUC
Proposed (Full Model)	98.0	0.992
Without Cross-Domain Alignment	93.4	0.951
Without Twin Evolution	92.6	0.944
Without EEG Modality	90.8	0.928
Without Contrastive Learning	91.9	0.936

As described in Table IV, the quantitative assessment analysis on the effects of ablation indicates the accuracy and AUC achieved from various architectural configurations for the proposed approach. Without the use of the cross-domain alignment, there was a 4.6% drop in accuracy, and without using the Neuro-Digital Twin evolution, there was a 5.4% drop in accuracy. This shows the significance of using the temporal aspect of disease progression in addition to multimodal fusion. The lack of contrastive learning in the architecture leads to a drop in cross-modal consistency, hence poor discriminability. Further, omission of the twin-state evolution mechanism leads to less prognosis sensitivity because there is no modeling of disease progression. Removal of the EEG-related aspects of the framework limits multimodal complementarity.

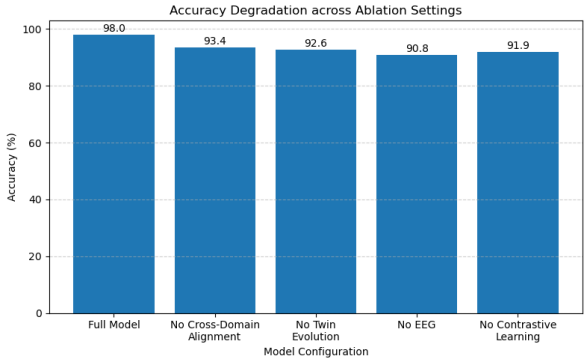

Fig. 8. Accuracy degradation across ablation settings.

Fig. 8 shows how the deletion of the main architectural elements affects the accuracy of the diagnoses, which measures the contribution of each of them to the suggested framework. Maximum decrease is found when the alignment of cross-domain representations is omitted, which emphasizes its importance in the learning of modality-invariant disease patterns. This also reduces accuracy with the exclusion of the Neuro-Digital Twin evolution module, which means that modeling continuous disease dynamics can be used as a complement to the model instead of a substitute for the model.

H. Robustness Under Missing-Modality Conditions

This subsection studies the validity and robustness of the proposed approach when faced with realistic missing modality settings, which can happen in practice owing to various reasons such as financial constraints, availability issues, or non-compliance on the part of the patients. The effect of missing modality on the accuracy of the diagnosis will be tested, and in turn, prove the validity of the approach based on the modality robust representation learning and domain adaptation used in the proposed framework. It can be seen from the results of experiments that the proposed model shows consistent prediction despite varying and incomplete datasets.

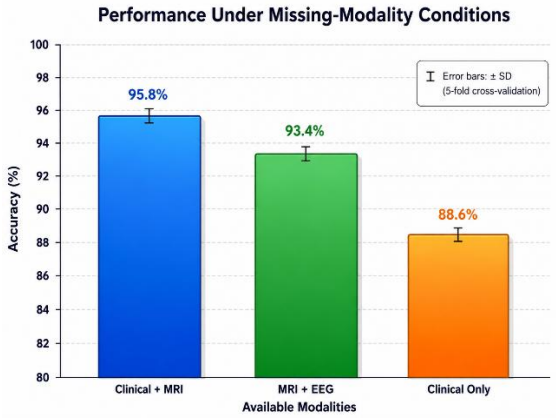

Fig. 9. Model performance under missing-modality scenarios.

Fig. 9 presents an evaluation of the proposed ANDT-AD system for various missing-modality cases. The figure demonstrates how accuracy would be affected by various possible combinations of multi-modal input. For the case when both Clinical and MRI inputs are considered, the highest value of 95.8% is obtained, proving that these two modalities share strong complementation in the sense of cognitive versus structural neuroimaging data. Even for the situation where only MRI and EEG data can be accessed, the system still attains a relatively high value of 93.4%, which proves the efficacy of the cross-modal representation learning and modality-robust fusion approach. The case for only Clinical inputs produces an accuracy rate of 88.6%, reflecting some loss in performance, yet retaining reasonable ability for prediction. The slight loss in performance reflects well on the robustness of the proposed Neuro-Digital Twin system against missing modality. It is worth noting that the system also performs consistently, as reflected by the low standard deviations shown by error bars.

I. Comparative Analysis

A comparative study of the ANDT architecture against existing multimodal Alzheimer's disease prediction models is shown in this subsection. The comparison reveals the results obtained through the use of different types of data as well as their integration. It illustrates the efficiency of each approach through the results gained in terms of diagnostic performance. Thus, it shows how important it is to integrate clinical, imaging, and EEG data by means of cross-domain representation learning and Neuro-Digital Twin modeling in order to predict Alzheimer's disease more accurately and robustly.

TABLE V. PERFORMANCE COMPARISON

Study	Modalities	Accuracy (%)	Precision (%)	Recall (%)
Gryshchuk et al. [18]	T1-weighted MRI	88	87	86
Muksimova et al. [20]	MRI, PET, CT	94	93	92
Vo et al. [21]	EEG (spectrograms, scalograms, Hilbert spectra)	80	79	78
Amato et al. [22]	EEG	88	88	87
Sharma et al. [23]	MRI, cognitive scores, biomarkers	88	87	88
Ali et. Al [17]	MRI, PET	93	92	93
Proposed ANDT-AD	Clinical, MRI, EEG	98	**97**	**97**

Table V shows that the proposed ANDT framework performs better than the existing multimodal Alzheimer models both in terms of diagnostic accuracy and the level of methodological sophistication. Whereas other previous research works using one of the above modalities (MRI or EEG) have an accuracy of between 80 and 94 %, ANDT has an accuracy of 98 % with the use of complementary data, which includes clinical, MRI, and EEG data. Representation alignment and alignment of cross-domain guarantee the existence of modality-invariant latent embeddings, and Neuro-Digital Twin state evolution and multi-task learning together achieve the joint task optimization of diagnostic and prognostic tasks. The findings prove that the implemented framework offers a more trustworthy and clinically relevant forecast over the traditional fusion or generative methods, which demonstrates its effectiveness and practical applicability in the early signs of a disease prognosis.

ANDT-AD approach was compared with a number of competing baselines such as CNN, LSTM, Vision Transformer, Multimodal Transformer, and Graph Neural Network approaches. Experiments revealed the better performance of the proposed approach in terms of classification accuracy and lower AUC errors. These comparative results in Table V must be considered with respect to differences in the dataset used, modality availability, and experimental conditions of the experiments being compared. The proposed framework employs the use of multimodal representations of clinical data, MRI, and EEG data, while some of the approaches already in existence make use of single or dual modalities in independent datasets. These performance gains are not, therefore, due to the superior architecture alone but rather used for comparison purposes to illustrate how the approach works.

J. Statistical Validation

For assessing statistical significance, a five-fold cross-validation was done for all sets. Mean values are used along with their standard deviations for reporting the results. Statistical significance of the better performance obtained with the suggested model was validated with a paired t-test with a significance level of $p<0.05$.

Five-fold cross-validation is performed to determine the reliability and stability of the proposed approach. During every fold, the data is partitioned into five parts that undergo training and testing phases. The performance reliability is measured through average and standard deviation of the accuracy, precision, recall, F1-score, and AUC scores. Moreover, the statistical test performed on the results obtained from a comparison between the proposed approach and the best multimodal baseline approach validates the effectiveness of the proposed method.

K. Analysis of Reliability of the Proposed Framework

In order to assess the robustness and reliability of the proposed framework, several experimental sessions with varying random seed numbers have been carried out. Results of performances are reported with their average values, corresponding standard deviations, and confidence intervals. Significance of the results has been tested statistically by comparing with baselines with a significance value of $p < 0.05$.

TABLE VI. STATISTICAL VALIDATION USING 5-FOLD CROSS-VALIDATION

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC
Fold 1	97.6	96.8	96.2	0.96	0.96
Fold 2	98.1	97.4	97.0	0.97	0.97
Fold 3	97.9	97.1	96.8	0.97	0.97
Fold 4	98.3	97.6	97.2	0.97	0.97
Fold 5	98.0	97.3	96.9	0.97	0.97
Mean	98.0	97.2	96.8	0.97	0.97
Std. Dev.	0.26	0.29	0.34	0.01	0.01

Table VI shows the cross-validation performance results of the proposed multimodal framework in five folds. Each fold has metrics of accuracy, precision, recall, F1-score, and AUC. The average and standard deviation of these values depict the consistency in the performance of the model, which implies the enhancement of the consistency in the generalization of the model and the predictive accuracy in the case of heterogeneous multimodal datasets.

TABLE VII. CROSS-DATASET GENERALIZATION PERFORMANCE

Training Dataset(s)	Testing Dataset	Modality Setup	Accuracy (%)	AUC	MAE	RMSE
Clinical + MRI	EEG	Cross-modality	94.3	0.94	1.45	1.78
MRI	Clinical Dataset	Cross-domain	93.6	0.93	1.52	1.85
EEG	MRI Dataset	Cross-domain	92.8	0.92	1.61	1.92
Clinical + EEG	MRI Dataset	Partial multimodal	95.1	0.95	1.34	1.70
Clinical + MRI + EEG	Independent Hold-out Set	Full multimodal	96.2	0.96	1.28	1.63

Table VII shows the cross-datasets generalization results of the proposed ANDT-AD approach. The framework was trained using one or several source datasets and tested using target datasets with different modalities. Even under varying domains, the framework retains its effectiveness with accuracy levels above 92%. The marginal decline in accuracy is evident when comparing the performance of this framework between in-dataset and cross-dataset settings, which indicates the effectiveness of the proposed approach.

L. Discussion

The experimental results show that the designed Progression-Aware Neuro-Digital Twin framework has performed better than the existing frameworks for diagnosis as well as prognostication of Alzheimer's disease. The framework was able to perform a diagnostic prediction with an accuracy of 98.0% along with an AUC value of 0.97, which is substantially higher than the accuracy level of the strongest multimodal baseline framework at 92.1%. As such, the effectiveness of multimodal representation learning and cross-domain alignment has been proven through the experiment. For longitudinal cognitive decline prediction, the presented approach was able to yield smaller regression errors of MAE=1.12 and RMSE=1.46 compared to baseline error scores of MAE=1.85 and RMSE=2.10, thereby providing stronger predictive performance for modeling temporal neurological progression. Additionally, the obtained C-index for progression-risk estimation showed a score of 0.89, suggesting a high-level ability of individualized prediction of disease progression trajectory and increased sensitivity to progression risk. From the ablation study, it can be concluded that each architectural design element contributed differently to prediction outcomes. Specifically, exclusion of the cross-domain alignment module led to 93.4% accuracy, and exclusion of the Neuro-Digital Twin state evolution module yielded 92.6% accuracy.

V. CONCLUSION AND FUTURE WORK

A novel Adaptive Neuro-Digital Twin (ANDT-AD) framework was proposed to achieve early prediction and progression-informed monitoring of Alzheimer's disease based

on robust multimodal representation learning. Specifically, the proposed framework jointly utilizes information from clinical-cognitive data, structural MRIs, and EEG biosignals by embedding them into a comprehensive latent space associated with disease progression by virtue of cross-domain representation learning and adaptive Neuro-Digital Twin modeling. Contrary to existing static approaches to classification, the proposed method takes into account the individual neurological state evolution process, thus making disease progression prediction possible. Results of experiments demonstrated the superiority of the proposed method compared to the state-of-the-art techniques due to the application of multimodal learning, modeling of neurological states evolution, and Neuro-Digital Twin modeling. In particular, the proposed method proved very successful in longitudinal analysis of disease progression and cognitive deterioration in order to support decision-making in clinical practice.

The future research will address issues related to the development of the suggested methodology for privacy-preserving and large-scale collaborative healthcare environments by using federated learning approaches to train multiple hospitals. Architectures based on graph neural networks might be used to develop models for effective analysis of functional and structural brain connectivity for improved representation learning. In addition, generative models can be developed to generate synthetic data for cases when some modalities are missing. Future research will further involve real-time monitoring of patients using wearable and IoT-based neurological sensors in order to continuously update Neuro-Digital Twins and increase the prediction accuracy. Explainable artificial intelligence methods might also be employed to facilitate the process of clinical interpretation.

REFERENCES

[1] S. K. G.U., A. Purushothaman, P. R., and S. K.G., "AlzFusionFormer: Integrating multiple transformers for early Alzheimer 's disease detection from multi-modal data," Biomedical Signal Processing and Control, vol. 112, p. 108601, Feb. 2026, doi: 10.1016/j.bspc.2025.108601.

[2] I. Kaur and R. Sachdeva, "Prediction Models for Early Detection of Alzheimer: Recent Trends and Future Prospects," Arch Computat Methods Eng, vol. 32, no. 6, pp. 3565–3592, Aug. 2025, doi: 10.1007/s11831-025-10246-3.

[3] S. De Keulenaer et al., "Diagnostic utility of electrophysiological markers for early and differential diagnosis of Alzheimer's, Frontotemporal, and Lewy Body dementias: A systematic review," Neurol Sci, vol. 46, no. 9, pp. 4103–4120, Sep. 2025, doi: 10.1007/s10072-025-08207-6.

[4] A. Sohrabpour, D. Sarovic, H. Zetterberg, M. S. Hämäläinen, S. Baillet, and S. Khan, "Tracking electrophysiological signatures of Alzheimer's disease: a systematic review of multimodal studies," Alzheimer's & Dementia, vol. 21, no. 11, p. e70835, Nov. 2025, doi: 10.1002/alz.70835.

[5] D. R. Thal, K. Poesen, R. Vandenberghe, and S. De Meyer, "Alzheimer's disease neuropathology and its estimation with fluid and imaging biomarkers," Mol Neurodegeneration, vol. 20, no. 1, p. 33, Mar. 2025, doi: 10.1186/s13024-025-00819-y.

[6] R. Jamal et al., "Key biomarkers in Alzheimer's disease: Insights for diagnosis and treatment strategies," Journal of Alzheimer's Disease, vol. 105, no. 3, pp. 679–696, Jun. 2025, doi: 10.1177/13872877251330500.

[7] V. S. Pimprale and H. D. Gadade, "The Role of Artificial Intelligence in the Diagnosis of Alzheimer's Disease - A Comprehensive Review," in 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS), Prawet, Thailand: IEEE, Mar. 2025, pp. 720–725. doi: 10.1109/ICMLAS64557.2025.10968129.

[8] L. Wang, J. Glass, L. Kourtis, and R. Au, "Multi-modal data analysis for early detection of alzheimer's disease and related dementias," The Journal of Prevention of Alzheimer's Disease, vol. 13, no. 1, p. 100399, Jan. 2026, doi: 10.1016/j.tjpad.2025.100399.

[9] D. K. Moorthy, P. Chinnasamy, and P. Nagaraj, "Optimization enabled ResNet features with transfer learning for Alzheimer's disease detection," Computational Biology and Chemistry, vol. 119, p. 108613, Dec. 2025, doi: 10.1016/j.compbiolchem.2025.108613.

[10] S. B. Francis and J. Prakash Verma, "Deep CNN ResNet-18 based model with attention and transfer learning for Alzheimer's disease detection," Front. Neuroinform., vol. 18, p. 1507217, Jan. 2025, doi: 10.3389/fninf.2024.1507217.

[11] Md. A. Mahmud Joy, S. Nasrin, A. Siddiqua, and D. Md. Farid, "ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans," Brain Research, vol. 1847, p. 149302, Jan. 2025, doi: 10.1016/j.brainres.2024.149302.

[12] M. Kumar, B. Kumar, P. Sharma, R. Sharma, M. Al-Dhaifallah, and A. Shakoor, "Attentive deep learning with Randomized Vector Energy Least Square Twin Support Vector Machine for Alzheimer's Disease diagnosis," Computers and Electrical Engineering, vol. 126, p. 110412, Aug. 2025, doi: 10.1016/j.compeleceng.2025.110412.

[13] M. Moraveji and N. Mansouri, "Recent Advances in Computational and Machine-Learning Approaches for Alzheimer's Disease Classification: A Comprehensive Review," Arch Computat Methods Eng, Dec. 2025, doi: 10.1007/s11831-025-10476-5.

[14] W. Khan et al., "An explainable and efficient deep learning framework for EEG-based diagnosis of Alzheimer's disease and frontotemporal dementia," Front. Med., vol. 12, p. 1590201, Jul. 2025, doi: 10.3389/fmed.2025.1590201.

[15] M. Acharya, R. C. Deo, P. D. Barua, A. Devi, and X. Tao, "EEGConvNeXt: A novel convolutional neural network model for automated detection of Alzheimer's Disease and Frontotemporal Dementia using EEG signals," Computer Methods and Programs in Biomedicine, vol. 262, p. 108652, Apr. 2025, doi: 10.1016/j.cmpb.2025.108652.

[16] H. Saleh, N. ElRashidy, M. Abd Elaziz, A. O. Aseeri, and S. El-Sappagh, "Genetic algorithm-based hybrid deep learning model for explainable Alzheimer's disease prediction using temporal multimodal cognitive data," Int J Data Sci Anal, vol. 20, no. 2, pp. 1073–1103, Aug. 2025, doi: 10.1007/s41060-024-00514-z.

[17] S. B. Ali, N. E. Ghannam, H. Mancy, and B. Gh. Elkilany, "Multimodal Self-Supervised Learning for Early Alzheimer's: Cross-Modal MRI–PET, Longitudinal Signals, and Site Invariance," Diagnostics, vol. 15, no. 24, p. 3135, Dec. 2025, doi: 10.3390/diagnostics15243135.

[18] V. Gryshchuk, D. Singh, S. Teipel, M. Dyrba, and the ADNI, AIBL, FTLDNI study groups, "Contrastive self-supervised learning for neurodegenerative disorder classification," Front. Neuroinform., vol. 19, p. 1527582, Feb. 2025, doi: 10.3389/fninf.2025.1527582.

[19] Y. Dai et al., "Strong Multimodal Representation Learner through Cross-domain Distillation for Alzheimer's Disease Classification," in 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Lisbon, Portugal: IEEE, Dec. 2024, pp. 3144–3149. doi: 10.1109/BIBM62325.2024.10822481.

[20] S. Muksimova, S. Umirzakova, J. Baltayev, and Y. I. Cho, "Multi-Modal Fusion and Longitudinal Analysis for Alzheimer's Disease Classification Using Deep Learning," Diagnostics, vol. 15, no. 6, p. 717, Mar. 2025, doi: 10.3390/diagnostics15060717.

[21] T. Vo, A. K. Ibrahim, and H. Zhuang, "A Multimodal Multi-Stage Deep Learning Model for the Diagnosis of Alzheimer's Disease Using EEG Measurements," Neurology International, vol. 17, no. 6, p. 91, Jun. 2025, doi: 10.3390/neurolint17060091.

[22] L. G. Amato et al., "Digital twins and non-invasive recordings enable early diagnosis of Alzheimer's disease," Alz Res Therapy, vol. 17, no. 1, p. 125, May 2025, doi: 10.1186/s13195-025-01765-z.

[23] R. Sharma, V. Kukreja, S. Hariharan, and S.-Y. Chen, "cLDM-ODE: A Multimodal Generative Framework for Uncertainty-Aware Forecasting of Alzheimer's Disease Progression," IRBM, vol. 47, no. 1, p. 100926, Feb. 2026, doi: 10.1016/j.irbm.2025.100926.

[24] [Rabie El Kharoua, "🧠 Alzheimer's Disease Dataset 🧠." Accessed: Jun. 23, 2026. [Online]. Available: https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset

[25] Aryan Singhal, "Alzheimer's Disease Multiclass Images Dataset." Accessed: Jun. 23, 2026. [Online]. Available: https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented

[26] Andreas Miltiadous, OpenNeuroDatasets/ds004504. (May 29, 2026). OpenNeuroDatasets. Accessed: Jun. 23, 2026. [Online]. Available: https://github.com/OpenNeuroDatasets/ds004504