Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
DOI : 10.47974/JIOS-2111

Translating clinical data into neuroimaging : A deep latent space framework for predicting Alzheimer's disease progression

Firas Saadullah Raheem [†]
Leili Farzinvash *
Mohammad Asadpour [§]
Department of Computer Engineering
University of Tabriz
29 Bahman Boulevard
Tabriz 5166616471
Iran

Abstract

Early and accurate prediction of Alzheimer's disease progression remains a critical challenge in clinical neuroscience. In this study, we propose a multimodal framework based on a Canonical Correlation Analysis Variational Autoencoder (CCA-VAE) for learning a shared latent representation from heterogeneous data modalities, specifically MRI and clinical features. This latent representation is used as input to classical and neural classifiers including Logistic Regression (LR), Random Forest (RF), and Artificial Neural Network (ANN). Experimental results on the UDS dataset demonstrate that our CCA-VAE-based approach improves classification performance across all metrics (precision, recall, F1-score, and AUC) compared to baseline unimodal and concatenation-based models. Notably, the CCA-VAE-ANN model achieved the highest accuracy and robustness, highlighting the benefit of deep, correlated latent spaces for multimodal fusion. These results underscore the potential of CCA-VAE architectures in developing reliable early diagnostic tools for Alzheimer's disease progression.

Subject Classification: *92C55, 68T07, 62P10.*

Keywords: *Alzheimer's disease, Deep learning, Variational autoencoder, Canonical correlation analysis, Multi-modal data, Neuroimaging, Clinical data integration, Disease progression modeling.*

[†] *ORCID-ID:* **0000-0002-9059-5874**
* *ORCID-ID:* **0000-0002-2369-9833**
[§] *ORCID-ID:* **0000-0002-1672-149X**

[†] *E-mail:* **firas.a.raheem@uotechnology.edu.iq**
* *E-mail:* **l.farzinvash@tabrizu.ac.ir (Corresponding Author)**
[§] *E-mail:* **m_asadpour@tabrizu.ac.ir**

© Taru Publications

1. Introduction

The integration of artificial intelligence into medical imaging has reached a transformative phase, where synthetic image generation now promises to bridge critical gaps in clinical decision-making. The ability to generate magnetic resonance imaging (MRI) scans directly from routine clinical data-including laboratory results, demographic information, and medical history-represents a paradigm shift with three profound implications:

- Compensating for missing imaging in resource-constrained settings or when repeated scans are contraindicated;
- Predicting disease progression through simulated anatomical changes over time;
- Personalizing treatment planning via patient-specific pathology modeling.

Several studies have addressed the early prediction of Alzheimer's disease (AD) by identifying key biomarkers and applying traditional machine learning (ML) techniques. Irfan, Shahrestani, and Elkhodr [1] used feature selection to extract cognitive features and trained an ensemble model with adaptive voting. Franciotti et al. [2] selected cognitive, MRI, and PET biomarkers using Random Forest (RF) and Gradient Boosting. Bucholc et al. [3] proposed a hybrid unsupervised-supervised ML framework employing RF, SVM, and Logistic Regression (LR) to identify MCI cases at high risk of conversion to AD. Likewise, another study used factor analysis to derive latent variables from neuropsychological, MRI, and CSF features and evaluated their predictive value using LR [4]. A common practice is categorizing MCI into progressive (pMCI) and stable (sMCI) forms based on AD conversion [3, 4], which enables better risk modeling. Kung et al. [5] trained a multi-layer perceptron using neuroimaging biomarkers to distinguish pMCI from sMCI. Although stable cases are more frequent and easier to predict [6-8], forecasting MCI-to-AD conversion is critical for timely intervention. Our work focuses on this transition, emphasizing interpretability. While traditional models such as SVM, LR, and RF are commonly used, they often require domain-driven biomarker selection. Deep learning methods have also been explored. Aghili et al. [9] applied LSTM and GRU networks to longitudinal data to predict diagnostic groups. Nguyen et al. [10] proposed a lightweight RNN capable of handling missing values in temporal cognitive and MRI data. El-Sappagh et al. [11] introduced a two-stage LSTM for disease progression and conversion time estimation, noting interpretability challenges due to nonlinear projections. For imaging data, Convolutional Neural Networks (CNNs) and Vision Transformers have been used effectively [12, 13, 29]. Bae et al. [14] utilized Inception-v4 on 2D coronal MRI slices to identify Alzheimer's disease. Other works fused PET and MRI data using Inception-based architectures [15]. While powerful, such models are computationally expensive and often lack interpretability and temporal context. Variational autoencoders have also been applied to model Alzheimer's disease progression from imaging and longitudinal data [16, 17].

Unlike image-based approaches, FATE-Net [18] takes tabular data (e.g., cognitive scores, MRI volumes), making it more practical in clinical settings. It incorporates time through embeddings and uses self-attention to learn feature relevance over time, improving MCI-to-AD recall and interpretability.

The foundational premise of this study is inspired by prior research [17, 19, 20], which demonstrated that clinical variables and genetic markers — specifically those from the Uniform Data Set Neuropsychological (UDS NP) battery and genetic datasets, including frontotemporal lobar degeneration (FTLD) and Lewy body dementia (LBD) — can predict future clinical outcomes with high accuracy (often exceeding 90%) over time horizons of two to three years. These findings underscore that clinical data collected at a given time point t_i encapsulates significant prognostic information relevant to a patient's future state at t_{i+1}. Such temporal relationships can often be effectively captured using relatively simple machine learning models.

In contrast, imaging modalities such as MRI do not exhibit the same readily exploitable predictive characteristics. The complex and high-dimensional nature of imaging data makes it more difficult to establish direct associations with future outcomes using straightforward analytical techniques.

Consequently, the central objective of this work is to develop a methodology for integrating heterogeneous data sources—specifically, to enrich imaging data by fusing it with complementary clinical information. This is accomplished through the construction of a latent common space, a low-dimensional representation that jointly encodes features from both clinical and imaging domains. The latent space is characterized by its capacity to preserve the essential structure of the original datasets, enabling accurate reconstruction of the input data while capturing shared and complementary predictive signals across modalities.

Key innovations include:

- **Predictive Modeling from Clinical Data:** The model aims to predict the transition from mild cognitive impairment (MCI) to Alzheimer's disease (AD) over short clinical follow-up periods (e.g., 2-3 years). It estimates the likelihood of progression by entering a small number of selected clinical variables. These variables are often collected from patients visiting Alzheimer's Disease Centers (ADCs), making this approach practical for trial recruitment and real-world deployment.
- **Latent Space Fusion via CCA for Multimodal Alignment:** A shared latent space is constructed to align clinical variables with tomographic imaging data (e.g., MRI), enabling extraction of features that are predictive of disease progression. Canonical Correlation Analysis (CCA) is incorporated to explicitly maximize the correlation between the clinical and imaging modalities, ensuring that the shared latent space captures relevant cross-modal information. Orthogonality constraints and time-

decaying correlation weights $\lambda(t)$ are used to balance disentanglement and reconstruction quality during training.

- **Architecture, Training Objectives, and Applicability:** The architecture uses a dual-branch encoder (MLP for clinical data, CNN for MRI) and fuses their latent representations through a correlation-aware mean projection. The training objective jointly optimizes reconstruction loss, KL divergence, cross-modal correlation, and latent space orthogonality. The resulting fused vector $\mathbf{z} = \frac{1}{2}\left(\mathbf{W}_x\mathbf{h}_x + \mathbf{W}_y\mathbf{h}_y\right)$ serves as input for both traditional ML and ANN classifiers. This approach is interpretable, stable, and capable of handling multimodal datasets for downstream prediction tasks.

2. Theoretical Foundations

Definition 2.1: (Latent Representation). Let $\mathcal{X} \subseteq \mathbb{R}^d$ be the input space. A latent representation $\mathbf{z} \in \mathbb{R}^k$ of an input $\mathbf{x} \in \mathcal{X}$ is a lower-dimensional embedding produced by an encoder $f_\theta: \mathbb{R}^d \to \mathbb{R}^k$ such that $\mathbf{z} = f_\theta(\mathbf{x})$, where $k \ll d$ [21].

Definition 2.2: (Desirable Properties of Latent Space). An ideal latent space $\mathcal{Z} \subseteq \mathbb{R}^k$ satisfies the following [21, 22]:

1. **Compactness:** $\dim(\mathcal{Z}) \ll \dim(\mathcal{X})$
2. **Re-constructability:** $\forall \epsilon > 0$ there exists a decoder g_ϕ such that
$$\left\| g_\phi(f_\theta(\mathbf{x})) - \mathbf{x} \right\| < \epsilon, \forall \mathbf{x} \in \mathcal{X}$$
3. **Disentanglement:** The latent distribution factorizes: $p(\mathbf{z}) = \prod_{i=1}^{k} p(z_i)$.

Definition 2.3: (Autoencoder). An autoencoder is a tuple $\mathcal{A} = \left(f_\theta, g_\phi, \mathcal{L}\right)$ where f_θ encodes and g_ϕ decodes, and $\mathcal{L}$ measures reconstruction loss [21]:

$$\min_{\theta,\phi} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}} \left[\mathcal{L}\left(\mathbf{x}, g_\phi\left(f_\theta(\mathbf{x})\right)\right)\right] \quad \ldots (1)$$

Definition 2.4: (Variational Autoencoder). A variational autoencoder replaces deterministic encodings with a distribution [21]:

$$q_\theta(\mathbf{z} \mid \mathbf{x}) = \mathcal{N}\left(\mathbf{z} \mid \boldsymbol{\mu}_\theta(\mathbf{x}), \text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{x}))\right).$$

Definition 2.5: (VAE Loss). The VAE objective:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q_\theta}\left[\log p_\phi(\mathbf{x} \mid \mathbf{z})\right] - \beta D_{KL}(q_\theta \| p(\mathbf{z}))$$

with $p(\mathbf{z}) = \mathcal{N}(0, \mathbf{I})$ [21].

Theorem 2.1: (Manifold Reconstruction Bound). *Let λ_{k+1} be the ($k+1$)-th eigenvalue of $Cov(\mathcal{X})$. Then [22]:*

$$\mathbb{E}\left[\left\|\mathbf{x} - g_\phi(f_\theta(\mathbf{x}))\right\|^2\right] \leq C \cdot \lambda_{k+1}.$$

Definition 2.6: (Canonical Correlation Analysis). CCA seeks projections $\mathbf{u}, \mathbf{v}$ that maximize [23]:

$$\rho = \max_{\mathbf{u},\mathbf{v}} \frac{\mathbb{E}[\mathbf{u}^\top \mathbf{x} \cdot \mathbf{v}^\top \mathbf{y}]}{\sqrt{\mathbb{E}[\mathbf{u}^\top \mathbf{x}\mathbf{x}^\top \mathbf{u}] \cdot \mathbb{E}[\mathbf{v}^\top \mathbf{y}\mathbf{y}^\top \mathbf{v}]}}$$

Definition 2.7: (Deep Canonical Correlation Analysis). DCCA replaces linear projections with deep nets and maximizes:

$\mathrm{tr}\left(\mathbf{U}^\top \mathbf{F}_\theta^\top \mathbf{G}_\phi \mathbf{V}\right)$ subject to orthogonality constraints [24].

Definition 2.8: (Modern Variants). Modern variants of Deep Canonical Correlation Analysis (Deep CCA) include several important extensions. DGCCA is a multi-view generalization that extends Deep CCA to more than two modalities [24]. CCA-AE combines the autoencoder loss with a correlation objective to jointly optimize reconstruction and feature alignment [25]. Contrastive $\boldsymbol{CCA}$ introduces contrastive learning techniques to improve the quality and discrimination of the learned representations [26].

Theorem 2.2: (Sample Complexity). *If* ρ_{min} *is the smallest canonical correlation:*

$$n \geq \frac{k \log(1/\epsilon)}{\rho_{\min}^2 \epsilon^2}$$

Proof: Follows from concentration bounds and matrix Bernstein inequality [27].

Theorem 2.3: (Identifiability). *Deep CCA representations are identifiable up to orthogonal transformation if*

$$\mathrm{rank}\left(\mathbb{E}\left[\mathbf{z}_x \mathbf{z}_y^\top\right]\right) = k.$$

Proof: Follows from the rotational invariance of canonical correlation [28].

3. Experimental Study and the Proposed Model

3.1 *UDS Dataset*

The Uniform Data Set (UDS) Version 3.0, released by the National Alzheimer's Coordinating Center (NACC) in 2015, provides a longitudinal dataset for Alzheimer's and related dementias. It includes 850+ validated variables covering demographics, medical history, cognitive assessments, MRI scans, and clinician notes. Documentation is available at: https://files.alz.washington.edu/documentation/uds3rdd.pdf. After filtering missing data, the final cohort includes 10,870 patients and 39,674 visits. Five folds (5-fold cross-validation): 80% (8,696 patients) for training, 20%(2,174 patients) for testing, with 6,956 patients for training and 1,740 for validation in

each split. Visit labels are derived from the NACCUDSD field, with 11,450 visits labeled as MCI and 28,224 as AD. About 17% of patients (1,810) convert from MCI to AD. Most patients have fewer than six visits, with missing data imputed using forward and backward filling.

3.2 *Proposed Model Architecture*

The proposed CCA-VAE framework integrates deep convolutional neural networks with canonical correlation analysis (CCA) in a variational autoencoder (VAE) structure to enable joint representation learning from paired clinical variables $\mathbf{x} \in \mathbb{R}^p$ and MRI images $\mathbf{y} \in \mathbb{R}^{H\times W\times C}$. The architectural design consists of three main components: dual-branch encoders, a correlation-optimized latent space, and symmetric decoders.

Encoders. The model employs a dual-branch encoder to process each modality independently. For clinical variables, a multilayer perceptron (MLP) with two hidden layers (sizes $512 \to 256$), using batch normalization and LeakyReLU activation ($\alpha = 0.2$), generates Gaussian parameters $\boldsymbol{\mu}_x \in \mathbb{R}^d$ and $\boldsymbol{\sigma}_x \in \mathbb{R}^d$. The MRI image branch consists of a four-block convolutional encoder (channels: $1 \to 32 \to 64 \to 128 \to 256$), where each block performs a 4×4 convolution (stride 2, padding 1), followed by batch normalization and LeakyReLU activation. The convolutional encoder downsamples the input by a factor of 16, yielding latent parameters $\boldsymbol{\mu}_y \in \mathbb{R}^d$ and $\boldsymbol{\sigma}_y \in \mathbb{R}^d$.

Latent Space and CCA Projection. As detailed in Algorithm 1, the view-specific latent variables are projected via learnable CCA matrices:

$$\mathbf{z}_x = \mathbf{W}_x\mathbf{h}_x, \mathbf{z}_y = \mathbf{W}_y\mathbf{h}_y, \mathbf{W}_x, \mathbf{W}_y \in \mathbb{R}^{d\times k}\# \tag{2}$$

where $\mathbf{h}_x$ and $\mathbf{h}_y$ are sampled from their respective Gaussian encoders. The shared latent representation is then defined as $\mathbf{z} = \frac{1}{2}(\mathbf{z}_x + \mathbf{z}_y)$. To maintain stable and disentangled correlation learning, orthogonality regularization is imposed on the CCA matrices via the constraint $\|\mathbf{W}_x^\top\mathbf{W}_x - \mathbf{I}\|_F^2 + \|\mathbf{W}_y^\top\mathbf{W}_y - \mathbf{I}\|_F^2$.

Decoders. A symmetric decoder reconstructs each modality from the shared latent space. The MRI decoder consists of four transposed convolutional layers (channels: $256 \to 128 \to 64 \to 32 \to 1$), each performing $2\times$ up-sampling and followed by a final sigmoid activation to constrain output values. For clinical data, a two-layer dense decoder (sizes $256 \to 512$) with a linear output layer reconstructs the variable vector.

Training Objective. The combined loss function is formulated as:

$$\mathcal{L} = \underbrace{\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]}_{\text{Clinical reconstruction}} + \underbrace{\mathbb{E}[\|\mathbf{y} - \hat{\mathbf{y}}\|^2]}_{\text{MRI reconstruction}} \# \tag{3}$$

$$+\beta \underbrace{D_{\mathrm{KL}}(q\|p)}_{\text{KL regularization}} - \lambda(t) \underbrace{\mathrm{tr}\left(\mathbf{W}_x^\top\Sigma_{xy}\mathbf{W}_y\right)}_{\text{CCA correlation}} \#$$

$$+\gamma \underbrace{\mathcal{L}_{\text{ortho}}}_{\text{Orthogonality constraints}} \#$$

where $\lambda(t) = \lambda_0 \exp(-\alpha t)$ adaptively balances the emphasis between reconstruction and correlation during training.

Training proceeds as described in Algorithm 1, utilizing the Adam optimizer ($\eta = 3 \times 10^{-4}$), learning rate decay ($\gamma = 0.95$ every 10 epochs), dropout ($p = 0.2$ in fully connected layers), and gradient clipping (max norm 5.0). Input feature normalization standardizes clinical variables to $\mathcal{N}(0,1)$ and scales MRI intensities to [0, 1].

Downstream Prediction. After model training, the fused latent vector $\mathbf{z}$ is extracted as per Algorithm 2 and utilized as input for downstream predictors, such as support vector machines (SVMs), logistic regression, random forests, or artificial neural networks (ANNs), to forecast subsequent disease state.

Ultimately, the CCA-VAE model combines data from clinical variables and MRI images into a unified latent space that preserves important correlations and maintains interpretability through orthogonality constraints. This approach leads to richer and more meaningful representations than considering each modality separately, enhancing the model's ability to capture complex relationships and improve predictive performance.

See Algorithms 1 and 2 for training and inference procedures, respectively.

3.3 *Performance Results*

The performance comparison of various machine learning models for predicting the diagnosis of Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) in patients using the UDS dataset is summarized in Table 1. The models were evaluated using several metrics, including Precision, Recall, F1-score, AUC, and Conversion Recall, to capture both classification accuracy and the ability to detect disease progression from MCI to AD. Among the traditional models, Logistic Regression (LogReg) and Support Vector Machine (SVM) show strong performance, with Precision (MCI) values of 0.800 and 0.796, and Precision (AD) values of 0.932 and 0.929, respectively. However, their lower Conversion Recall values of 0.519 and 0.514 indicate that they are less effective at identifying patients transitioning from MCI to AD. Similarly, XGBoost performs moderately, with Precision (AD) of 0.934, but it lags in Conversion Recall (0.478), demonstrating that traditional models face challenges in capturing the complexities of early disease progression.

In contrast, deep learning models such as the LSTM and Transformer architectures outperform traditional models in terms of Conversion Recall, with LSTM achieving a Conversion Recall of 0.659. However, the Transformer model exhibits instability, as evidenced by the high standard deviation of 0.255 in the Conversion Recall metric. Hybrid models like LSTM+FATE-Net, 1D-CNN+FATE-Net, and Transformer+FATE-Net combine the strengths of deep learning with feature extraction but still fail to surpass the CCA-VAE model. These models demonstrate promising feature extraction capabilities but are

hindered by their inability to optimize cross-modal correlations between clinical and imaging data effectively.

The proposed CCA-VAE model excels across all evaluation metrics. It achieves a Precision (MCI) of 0.924, Precision (AD) of 0.994, Recall (MCI) of 0.920, Recall (AD) of 0.990, F1-score of 0.957, AUC of 0.943, and a Conversion Recall of 0.798. The CCA-VAE outperforms other models due to its innovative architecture, which integrates clinical variables and MRI data into a shared latent space optimized for correlation and orthogonality. This architecture enables more accurate disease state prediction and superior performance in detecting patients transitioning from MCI to AD. The model's success in maintaining stable correlations between modalities is key to its superior performance.

Table 1

Comparative performance of models in predicting the diagnosis at the next visit (1-year follow-up). Standard deviations are shown in parentheses.

Model	Precision (MCI)	Precision (AD)	Recall (MCI)	Recall (AD)	F1 Score	AUC	Conv. Recall
LogReg [1]	0.800 (0.009)	0.932 (0.008)	0.829 (0.006)	0.917 (0.006)	0.893 (0.007)	0.875 (0.004)	0.519 (0.021)
SVM [1]	0.796 (0.007)	0.929 (0.008)	0.836 (0.009)	0.910 (0.006)	0.891 (0.006)	0.874 (0.005)	0.514 (0.016)
CNN [5]	0.806 (0.007)	0.923 (0.005)	0.818 (0.005)	0.918 (0.005)	0.889 (0.007)	0.868 (0.005)	0.542 (0.027)
XGBoost [2]	0.788 (0.011)	0.934 (0.006)	0.849 (0.012)	0.907 (0.007)	0.891 (0.006)	0.877 (0.007)	0.478 (0.022)
Transformer [18]	0.811 (0.032)	0.905 (0.016)	0.750 (0.045)	0.927 (0.020)	0.877 (0.009)	0.842 (0.012)	0.722 (0.255)
LSTM [10]	0.823 (0.011)	0.910 (0.004)	0.776 (0.013)	0.932 (0.009)	0.887 (0.006)	0.855 (0.003)	0.659 (0.016)
TabNet [18]	0.792 (0.017)	0.936 (0.013)	0.857 (0.021)	0.908 (0.008)	0.894 (0.007)	0.880 (0.009)	0.471 (0.030)
FATE-Net + FC [18]	0.774 (0.007)	0.916 (0.009)	0.797 (0.006)	0.906 (0.008)	0.875 (0.007)	0.851 (0.006)	0.546 (0.015)
LSTM+FATE-Net [18]	0.800 (0.009)	0.894 (0.015)	0.729 (0.023)	0.926 (0.007)	0.867 (0.012)	0.827 (0.015)	0.714 (0.048)
1D-CNN+FATE-Net [18]	0.777 (0.007)	0.911 (0.007)	0.790 (0.009)	0.906 (0.008)	0.874 (0.008)	0.848 (0.008)	0.555 (0.018)
Transformer+FATE-Net [18]	0.774 (0.011)	0.916 (0.007)	0.802 (0.007)	0.903 (0.011)	0.875 (0.010)	0.853 (0.008)	0.621 (0.025)

Contd...

FATE-Net + FC w/o t2 Vec [18]	0.778 (0.005)	0.912 (0.010)	0.793 (0.011)	0.906 (0.005)	0.851 (0.007)	0.850 (0.009)	0.559 (0.017)
CCA-VAE-LR	**0.919 (0.006)**	**0.992 (0.005)**	**0.914 (0.007)**	**0.986 (0.006)**	**0.954 (0.005)**	**0.941 (0.004)**	**0.794 (0.010)**
CCA-VAE-RF	**0.918 (0.006)**	**0.991 (0.005)**	**0.913 (0.007)**	**0.985 (0.006)**	**0.953 (0.005)**	**0.940 (0.004)**	**0.793 (0.010)**
CCA-VAE-ANN	**0.924 (0.005)**	**0.994 (0.004)**	**0.920 (0.006)**	**0.990 (0.005)**	**0.957 (0.004)**	**0.943 (0.003)**	**0.798 (0.009)**

4. Conclusion and Discussion

This study presents a novel CCA-VAE model, which combines deep convolutional networks with canonical correlation analysis in a variational autoencoder structure to jointly model clinical and imaging data. The model has shown significant improvement in both classification accuracy and conversion prediction compared to traditional machine learning models and hybrid models. The integration of clinical variables with MRI data into a shared latent space optimized for correlation and orthogonality is a major strength of this approach, allowing the model to capture complex relationships between different data modalities.

The CCA-VAE model's high Precision (MCI), Precision (AD), Recall (MCI), Recall (AD), F1-score, AUC, and Conversion Recall metrics demonstrate its effectiveness in predicting both the current diagnosis and future disease state transitions. Notably, the Conversion Recall of 0.798 is substantially higher than that of traditional models, highlighting the model's superior ability to identify patients at risk of transitioning from MCI to AD. This is crucial for early intervention and better patient outcomes. While deep learning models such as LSTM and Transformer show promise, they are limited by issues like instability and the inability to fully capture cross-modal correlations. Hybrid models, while effective at feature extraction, still do not outperform the CCA-VAE model, emphasizing the importance of optimizing the interaction between different data types. The results demonstrate that the CCA-VAE model's architecture, which allows for the optimization of correlation between clinical and imaging data, plays a crucial role in its superior performance.

Future work could explore further optimizations of the CCA-VAE model, including incorporating additional biomarkers or other imaging modalities, and applying it to real-world clinical datasets. Furthermore, there is potential to enhance the interpretability of the model, which would be highly valuable in clinical settings, where understanding the underlying factors contributing to predictions is as important as the predictions themselves.

Appendix

Algorithm 1 Training CCA-VAE for Joint Clinical and MRI Data

Require: Paired clinical variables $\mathbf{x} \in R^p$; MRI images $\mathbf{y} \in R^{H\times W\times C}$
Ensure: Trained model parameters and shared latents $\mathbf{z}$
Initialize network weights and CCA matrices $\boldsymbol{W}_x, \boldsymbol{W}_y$
Normalize clinical variables to $\mathcal{N}(0, 1)$; scale MRI values to [0, 1]
for each epoch **do**
    for each mini batch **do**
        // **Encoder Forward Pass**
        Encode $\mathbf{x}$ with MLP $\rightarrow (\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x)$
        Encode $\mathbf{y}$ with CNN $\rightarrow (\boldsymbol{\mu}_y, \boldsymbol{\sigma}_y)$
        // **Latent Sampling**
        Sample $\boldsymbol{h}_x \sim N(\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x)$, $\boldsymbol{h}_y \sim N(\boldsymbol{\mu}_y, \boldsymbol{\sigma}_y)$
        // **CCA Projection**
        $\boldsymbol{z}_x = \boldsymbol{W}_x \boldsymbol{h}_x;\ \boldsymbol{z}_y = \boldsymbol{W}_y \boldsymbol{h}_y$
        $\boldsymbol{z} = \frac{1}{2}(\boldsymbol{z}_x + \boldsymbol{z}_y)$
        // **Decoder Forward Pass**
        Decode $\mathbf{z}$ to $\hat{\mathbf{x}}$ (MLP) and $\hat{\mathbf{y}}$ (CNN)
        // **Loss Computation**
        $L_{rec}^x = \|x - \hat{x}\|^2$
        $L_{rec}^y = \|y - \hat{y}\|^2$
        L_{KL} = sum of KL divergences for both branches
        $L_{\mathrm{CCA}} = tr(\boldsymbol{W}_x^T \sum_{xy} \boldsymbol{W}_y)$
        $L_{ortho} = \|\boldsymbol{W}_x^T \boldsymbol{W}_x - \boldsymbol{I}\|_F^2 + \|\boldsymbol{W}_y^T \boldsymbol{W}_y - \boldsymbol{I}\|_F^2$
        $\mathcal{L} = L_{rec}^x + L_{rec}^y + \beta\, L_{KL} - \lambda(t) L_{CCA} + \gamma\, L_{ortho}$
        // **Optimization**
        Update parameters by Adam optimizer with learning rate decay and gradient clipping
        Apply dropout in fully-connected layers
    end for
end for

Algorithm 2 CCA-VAE Inference and Disease Prediction

Require: Clinical variables $\mathbf{x}$, MRI $\mathbf{y}$
Ensure: Predicted disease state
Normalize and scale inputs as during training
Encode $\mathbf{x}$ and $\mathbf{y}$ to obtain $\boldsymbol{h}_x, \boldsymbol{h}_y$
$\boldsymbol{z}_x = \boldsymbol{W}_x \boldsymbol{h}_x;\qquad \boldsymbol{z}_y = \boldsymbol{W}_y \boldsymbol{h}_y$
$\boldsymbol{z} = \frac{1}{2}(\boldsymbol{z}_x + \boldsymbol{z}_y)$
Feed $\boldsymbol{z}$ to downstream predictor (e.g., SVM, ANN)
return predicted disease state

References

[1] M. Irfan, S. Shahrestani, and M. Elkhodr, "Early detection of the Alzheimer's disease: A novel cognitive feature selection approach using machine learning," in *Advances in Information, Communication and Cybersecurity (ICI2C'21)*. Cham, Switzerland: Springer, pp. 383–392 (2022), doi: 10.1007/978-3-030-91738-8_35.

[2] R. Franciotti, D. Nardini, M. Russo, M. Onofrj, and S. L. Sensi, "Comparison of machine learning-based approaches to predict the conversion to Alzheimer's disease from mild cognitive impairment," *Neuroscience,* vol. 514, pp. 143–152 (2023), doi: 10.1016/j.neuroscience.2023.01.029.

[3] M. Bucholc, S. Titarenko, X. Ding, C. Canavan, and T. Chen, "A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia," *Expert Syst. Appl.,* vol. 217, Art. no. 119541 (2023), doi: 10.1016/j.eswa.2023.119541.

[4] M. Chang and C. J. Brainerd, "Predicting conversion from mild cognitive impairment to Alzheimer's disease with multimodal latent factors," *J. Clin. Exp. Neuropsychol*., vol. 44, no. 4, pp. 316–335 (2022).

[5] T.-H. Kung, T.-C. Chao, Y.-R. Xie, M.-C. Pai, Y.-M. Kuo, and G. G. C. Lee, "Neuroimage biomarker identification of the conversion of mild cognitive impairment to Alzheimer's disease," *Front. Neurosci.,* vol. 15, Art. no. 584641 (2021).

[6] M. Ansart, S. Epelbaum, G. Bassignana, A. Bône, S. Bottani, T. Cattai, R. Couronné, J. Faouzi, I. Koval, M. Louis, E. Thibeau-Sutre, J. Wen, A. Wild, N. Burgos, D. Dormont, O. Colliot, and S. Durrleman, "Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review," *Med. Image Anal.,* vol. 67, Art. no. 101848 (2021).

[7] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: A systematic review," *Alzheimers Res. Ther.*, vol. 13, Art. no. 162 (2021).

[8] Y. Chen, X. Qian, Y. Zhang, W. Su, Y. Huang, X. Wang, X. Chen, E. Zhao, L. Han, and Y. Ma, "Prediction models for conversion from mild cognitive impairment to Alzheimer's disease: A systematic review and meta-analysis," *Front. Aging Neurosci.,* vol. 14, Art. no. 840386 (2022).

[9] M. Aghili, S. Tabarestani, M. Adjouadi, and E. Adeli, "Predictive modeling of longitudinal data for Alzheimer's disease diagnosis using RNNs," in *Predictive Intelligence in Medicine (PRIME 2018), Lecture Notes in Computer Science*, vol. 11121. Cham, Switzerland: Springer, pp. 112–119 (2018).

[10] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, and B. T. T. Yeo, "Predicting Alzheimer's disease progression using deep recurrent neural networks," *NeuroImage*, vol. 222, Art. no. 117203 (2020).

[11] S. El-Sappagh, H. Saleh, F. Ali, E. Amer, and T. Abuhmed, "Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time," *Neural Comput. Appl.,* vol. 34, no. 17, pp. 14487–14509 (2022), doi: 10.1007/s00521-022-07263-9.

[12] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 574–584 (2022).

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv:2010.11929 (2020).

[14] J. B. Bae, S. Lee, W. Jung, S. Park, W. Kim, H. Oh, J. W. Han, G. E. Kim, J. S. Kim, J. H. Kim, and K. W. Kim, "Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging," *Sci. Rep.*, vol. 10, Art. no. 22252 (2020).

[15] Z. Yuan, N. Qi, Z. Zhou, J. Ding, X. Chen, J. Wu, J. Wang, and J. Zhao, "Diagnosis of Alzheimer's disease using transfer learning with multi-

modal 3D Inception-v4," *Quant. Imaging Med. Surg.*, vol. 15, no. 2, pp. 1455–1467 (2025).

[16] S. Basu, K. Wagstyl, A. Zandifar, L. Collins, A. Romero, and D. Precup, "Early prediction of Alzheimer's disease progression using variational autoencoders," in *Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*. Cham, Switzerland: Springer, pp. 205–213 (2019).

[17] B. Sauty and S. Durrleman, "Progression models for imaging data with longitudinal variational autoencoders," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*. Cham, Switzerland: Springer, pp. 3–13 (2022).

[18] X. Fang, Y. Fan, Y. Zhao, Z. Lin, K. Xu, and Z. Wen, "Conversion-aware forecasting of Alzheimer's disease via featurewise attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 18191–18201 (2023).

[19] Y. Pang, W. Kukull, M. Sano, R. L. Albin, C. Shen, J. Zhou, and H. H. Dodge, "Predicting progression from normal to MCI and from MCI to AD using clinical variables in the National Alzheimer's Coordinating Center uniform data set version 3: Application of machine learning models and a probability calculator," *J. Prev. Alzheimers Dis.*, vol. 10, no. 2, pp. 301–313 (2023).

[20] M. Lin, P. Gong, T. Yang, J. Ye, R. L. Albin, and H. H. Dodge, "Big data analytical approaches to the NACC dataset: Aiding preclinical trial enrichment," *Alzheimer Dis. Assoc. Disord.,* vol. 32, no. 1, pp. 18–27 (2018).

[21] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press (2016). [Online]. Available: https://www.deeplearningbook.org/

[22] A. Shrivastava, R. Rameshan, and S. Agnihotri, "Latent space characterization of autoencoder variants," arXiv:2412.04755 (2024).

[23] H. Hotelling, "Relations between two sets of variates," *Biometrika,* vol. 28, nos. 3–4, pp. 321–377 (1936).

[24] A. Benton, H. Khayrallah, B. Gujral, D. A. Reisinger, S. Zhang, and R. Arora, "Deep generalized canonical correlation analysis," arXiv:1702.02519 (2017).

[25] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 37, pp. 1083–1092 (2015).

[26] F. Feng, X. Wang, and R. Li, "Cross-modal retrieval with correspondence autoencoder," in *Proc. ACM Int. Conf. Multimedia (MM)*, pp. 7–16 (2014).

[27] C. Gao, Z. Ma, Z. Ren, and H. H. Zhou, "Minimax estimation in sparse canonical correlation analysis," *Ann. Statist.*, vol. 43, no. 5, pp. 2168–2197 (2015).

[28] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, PMLR, vol. 28, no. 3, pp. 1247–1255 (2013).

[29] G. K. Yadav, A. K. Tiwari, V. K. Singh, A. Kumar, D. Varshney, and A. K. S. Sanger, "Empirical Algorithm to Detect Skin Disease Using Deep Convolution Neural Networks through MobileNet Algorithm," *J. Inf. Optim. Sci.*, vol. 45, no. 4, pp. 1017–1027 (2024), doi: 10.47974/JIOS-1624.

Received July, 2025