

A Combination of Functional Principal Component Analysis and Long Short-Term Memory Networks for Alzheimer's Disease Prediction

Yong-Shiuan Lee

Department of Applied Mathematics

Feng Chia University

Taichung City, Taiwan (R.O.C.)

yongslee@fcu.edu.tw

<https://orcid.org/0000-0003-3760-4513>

Huimei Liu

Department of Statistics

National Chengchi University

Taipei City, Taiwan (R.O.C.)

hliu@mail2.nccu.tw

for the Alzheimer's Disease
Neuroimaging Initiative*

Abstract—Alzheimer's disease (AD) research has focused on understanding its causes, progression, and treatments, with an emphasis on early detection, particularly diagnosing mild cognitive impairment. Research AD data are typically longitudinal, sparse, and irregularly spaced, with high-dimensional features due to advances in neuroimaging techniques. Traditional classification methods have limitations on handling such data. This study proposes to combine multivariate functional principal component analysis (MFPCA) with long short-term memory (LSTM) networks to reconstruct variable functions over time and classify data. The empirical results show that this approach outperformed other methods, including the flexible discriminant analysis and the regularized discriminant analysis. These findings suggest that the proposed method holds great potential for the early detection of AD.

Index Terms—Alzheimer's disease, functional principal component analysis, long short-term memory networks, longitudinal data, discriminant analysis

I. INTRODUCTION

Dementia is a long-term effect of impairment in cognitive functions that develops gradually and could be beyond what we expect from biologically normal aging. Alzheimer's Disease International (ADI) declares that 'among all types of dementia, Alzheimer's disease (AD) is the most prevalent and well-known form over the age of 65' [1]. AD is a growing concern due to its increasing prevalence and the significant burden it places on families and the healthcare system. Despite extensive research, the causes of AD remain unclear, and there

is no cure. Research focusing on the progression of the disease is crucial for early detection and risk reduction.

The diagnosis of AD can be achieved through a combination of clinical assessments, including medical or family history, physical and neurological examinations, cognitive assessments, laboratory tests, and structural imaging. Typically, medical data of AD are collected longitudinally, with variables recorded at multiple time points. However, the observation time points vary among subjects. The data can be sparse due to the limited number of measurements per subject and irregularly spaced. Additionally, neuroimaging data generate high-dimensional features such as regions of interest (ROI). Traditional statistical models have limitations in analyzing such high-dimensional, sparse, and irregular longitudinal data effectively. Therefore, dealing with longitudinal data with missing values is one of the main purposes of this study.

This study introduces a novel approach by combining functional principal component analysis (FPCA) with long short-term memory (LSTM) networks to model AD data. The FPCA reconstructs sparse and irregular data that have time evolving variables, and hence is a novel imputation method to predict missing information. LSTM, a recurrent neural network designed for sequence and time series data, is applied to model the reconstructed AD data. We also apply some extended discriminant methods, the flexible discriminant analysis (FDA) and the regularized discriminant analysis (RDA), by incorporating the features reconstructed by the FPCA as benchmark models. Unlike LSTM, these methods can only perform classification at one time point without considering the disease progress over a period of time. Our empirical results indicate that this method shows great potential for early diagnosis of AD.

The contributions of this study are as follows.

- We handle longitudinal and multimodal data with missing values by applying FPCA. The statistical methods used in this study cannot be applied without the functional data analysis technique. Furthermore, the method provides

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [ADNI Acknowledgement List.pdf](#) ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

both lower computational cost and higher explainability in comparison to other state-of-the-art approaches.

- Conventional statistical and machine learning models often rely on cross-sectional features, limiting their ability to capture temporal dynamics in the data. LSTM incorporates temporal patterns in disease progression for patient classification.
- The combination of FPCA and LSTM is a novel approach and is promising in detecting the early stage of AD.

II. RELATED WORK

Research on AD focuses mainly on early prediction of disease onset and identification of biomarkers. In the field of medical researches, most data analyses of the disease progression are made through the approaches of generalized linear models [2]–[4], especially the mixed-effects model. Furthermore, the search for better classification models for disease diagnosis continues by investigating a variety of methods such as logistic regression, linear discriminant analysis (LDA), K-nearest neighbors (KNN), naïve Bayes (NB), decision trees (DT) or random forest (RF), and support vector machine (SVM) [5]–[7].

However, there are challenges including complicated correlation structures, irregularly spaced observation points, and missing data for the longitudinal study [8]. Lately, more researchers are turning to deep learning models to investigate the classification problem (diagnosis) of AD, the convolutional neural network (CNN), and the recurrent neural network (RNN), for example [9]–[17].

Recent advances in machine learning (ML) and deep learning (DL) have significantly improved performance in medical image analysis, enabling efficient extraction of features from single or multimodal data to identify biomarkers for the detection, classification and diagnosis of brain diseases [18]. However, there are still some challenges in recent studies from the data and model point of view. Data-based challenges include data availability, missing data, longitudinal and multimodal data, and imbalanced data. Model-based challenges include interpretability, explainability, reproducibility, applicability, comparability, and reliability [18].

Most research deals with missing data by simple linear interpolation [19], filling based on the RNN model [20], sophisticated statistical methods such as MFPCA [20], [21], and deep learning models such as deep generative models [22]. Imbalanced data can be handled using several sampling techniques. The fusion of deep learning models is proposed to predict AD from multimodal data [23]. However, a unified method that deals with all complex longitudinal data is not available.

CNN and RNN models are most promising in AD prediction for processing neuroimaging data [24], [25] and multimodal data [26]. However, the complexity of the model, training time, resources, and data requirements lead to model-based challenges. The interpretability of machine learning models helps researchers and clinicians gain insights and confidence in their decision-making processes. On the other hand, deep

learning models are more data-intensive and require more resources on computational power and training time. Recent research shows that AI-supported brain volumetry improves dementia diagnosis and integrates into radiological workflows [27]. Systematic reviews can be found in [10], [18], [28]–[30].

III. METHODOLOGY

A. Functional Principal Component Analysis

Recently, large p small n problems ($p \gg n$) have emerged increasingly in various fields of applications due to advances in data collection. Under high-dimensional circumstances, one needs to overcome the curse of dimensionality. The functional data are realizations $x_i(s)$, $i = 1, 2, \dots, N$ collected discretely from an underlying stochastic process, $X : T \rightarrow \mathbb{R}$, over the common domain T with the assumption that X lies in the Hilbert space such as $L^2(T)$ [31].

1) *Univariate FPCA*: The Karhunen-Loève theorem [32], [33] expands the random process X as

$$X(s) = \mu(s) + \sum_{k=1}^{\infty} A_k \phi_k(s), \quad (1)$$

where $\phi_k(s)$ is the k th eigenfunction of the covariance function with respect to the k th eigenvalue, and $A_k = \int_T [X(s) - \mu(s)] \phi_k(s) ds$ are the functional principal component scores (or FPCs). It can be shown that

$$E(A_k) = 0, \quad E(A_k^2) = \lambda_k, \quad \text{Cov}(A_k, A_j) = 0, \text{ if } k \neq j, \quad (2)$$

$$\text{and } E \int [X(s) - \mu(s)]^2 ds = \sum_{k=1}^{\infty} \lambda_k. \quad (3)$$

The eigenvalue λ_k is the variance of X in the direction of ϕ_k , all of which can be summed to the total variance of X . Thus, the Karhunen-Loève decomposition is an analysis or a decomposition of variance [31], [34]. Usually, the eigenvalues λ_k decrease to zero very fast and the first few FPCs contribute a major proportion of variability of X . In this sense, it is optimal that X can be well approximated by only few ϕ_k . That is, the observations $x_1(s), \dots, x_N(s)$ can be approximated by the first m FPCs as

$$\begin{aligned} x_i(s) &= \mu(s) + \sum_{k=1}^{\infty} A_{ik} \phi_k(s) \\ &\approx \mu(s) + \sum_{k=1}^m A_{ik} \phi_k(s), \quad s \in T, \quad i = 1, \dots, N. \end{aligned} \quad (4)$$

In practice, $\Sigma(s, t)$ and hence the functional principal components (with corresponding eigenvalues and scores) are unknown, and need to be estimated from the observations. Several approaches have been proposed for the estimation. One is the principal components analysis through conditional expectation (PACE), which aims at irregularly-spaced sparse longitudinal data with measurement error [35]. The PACE estimator is a best linear unbiased predictor (BLUP) even the assumption of Gaussian does not hold. Refer more details to [35].

2) *Multivariate FPCA*: The FPCA is extended to multivariate functional data (MFPCA) with some methods being proposed during recent years. Motivated by the neuroimaging data of AD, a new method [36]–[38] is developed for multivariate functional data on different domains, whose dimensions are allowed to be different. This approach can be extended to non orthonormal basis for expanding the univariate component of the random process, and can also be useful for sparse longitudinal data with measurement error. A two-stage approach by implementing this MFPCA method [36] and using the resulting scores as features in the Cox regression model is proposed to predict the progression of AD [39]. Other applications of this MFPCA method [36] include analysis of waveform features for hip osteoarthritis [40] and missing data imputation for prediction of AD progress by deep learning models [20].

B. Discriminant Analysis

The linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) are two commonly applied classification methods based on the normal distribution

$$f_k(\mathbf{X}) = (2\pi)^{-\frac{p}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{X} - \boldsymbol{\mu}_k) \right], \quad (5)$$

where $\boldsymbol{\mu}_k$ and Σ_k are the vector of population mean and the covariance matrix for the k th class ($1 \leq k \leq K$), respectively.

The classification rule for QDA is

$$d_k(\mathbf{X}) = \min_{1 \leq k \leq K} d_k(\mathbf{X}), \quad (6)$$

with

$$d_k(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{X} - \boldsymbol{\mu}_k) + \ln |\Sigma_k| - 2 \ln \pi_k, \quad (7)$$

where π_k is the unconditional prior probability of observing a member in the k th class. The LDA is a special case of the QDA, where all of the within-group covariance matrices are assumed to be identical, i.e.

$$\Sigma_k = \Sigma, \quad 1 \leq k \leq K. \quad (8)$$

For multi-group classification or discrimination, especially for high-dimensional features, the flexible discriminant analysis (FDA) [41], [42] is proposed by using optimal scorings to represent the groups in an adaptive classification procedure. Two strategies for adaptive nonparametric regression are proposed to use in the procedure, namely MARS (multivariate adaptive regression splines) and BRUTO [43].

Similar to the FDA, the regularized discriminant analysis (RDA) [44] is proposed, which combines the covariance of QDA ($\hat{\Sigma}_k$) with the covariance of LDA ($\hat{\Sigma}$) by introducing two ridge-type tuning parameters λ ($\in [0, 1]$) and an γ ($\in [0, 1]$),

$$\hat{\Sigma}_k(\lambda, \gamma) = (1 - \gamma) \hat{\Sigma}_k(\lambda) + \frac{\gamma}{p} \text{tr}[\hat{\Sigma}_k(\lambda)] I. \quad (9)$$

The tuning parameter λ controls the level of shrinkage of each group covariance matrix estimates toward the pooled estimate

while γ controls shrinkage toward a multiple of the identity matrix.

C. Long Short-Term Memory Networks

Recurrent neural networks (RNNs) are models designed for sequence data prediction, where network depth corresponds to input sequence length [45]. A standard RNN processes input sequence $\mathbf{x} = (x_1, \dots, x_T)$ to produce hidden states $\mathbf{h} = (h_1, \dots, h_T)$ and outputs $\mathbf{y} = (y_1, \dots, y_T)$ through:

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (10)$$

$$y_t = W_{hy}h_t + b_y, \quad (11)$$

where W represents weights and b biases. Long short-term memory (LSTM) [46] extends RNNs by implementing the hidden layer function $\mathcal{H}$ with memory cells:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (12)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (13)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (14)$$

$$h_t = o_t \tanh(c_t), \quad (15)$$

where i_t , f_t , and o_t represent input, forget, and output gates respectively.

IV. DATA

The data, sampled from the ADNI study, consist of 364 subjects. All subjects are divided into three groups, cognitively normal (CN), MCI, and dementia (Alzheimer's disease; AD). There are 276, 58, and 30 subjects in each group, respectively. The observation time points for each subject range from 0 to 132 months. At each observation time point, the subject is assessed by two Alzheimer's disease assessment scales, i.e., MMSE and ADAS-Cog-13 [47], [48]. Low MMSE scores indicate dementia, while high ADAS-Cog-13 scores reflect severe cognitive impairment. Additionally, 315 regions of interest (ROI) variables converted from fMRI data are also recorded for all subjects at their observation time points. As a result, a total of 317 features are included in the data. Figure 1 shows the sample paths for one of the 317 features, the right hippocampus, of the AD group, illustrating the sparse and irregularly spaced longitudinal data.

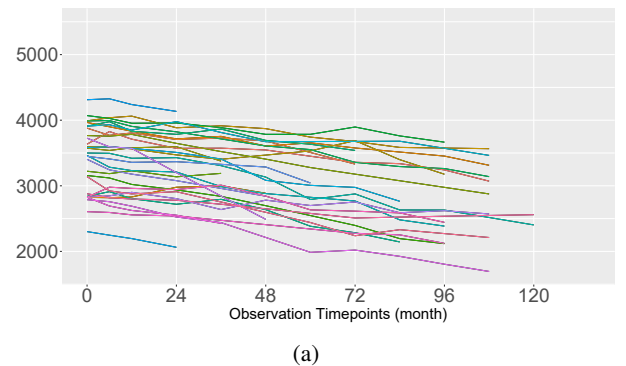


Fig. 1: Right Hippocampus vs.Time from AD Group.

V. EMPIRICAL RESULTS

A. Multivariate FPCA

We consider MFPCA to reconstruct the functions of 317 features simultaneously. Reconstruction allows us to perform other tasks, such as classification of subjects. We implement the MFPCA method by the R packages **funData** and **MFPCA** [37], [38].

The proportion of variance explained by the multivariate FPCs decreases rapidly. The first 31 PCs explain more than 90% of the cumulative explained variance, while they represent less than 10% of the total 317 extracted features.

B. Classification by Discriminant Analysis

Since there is only one observed subject at the time point of month 132, we focus on the data from the baseline (0 month) to the month of 120. We split the data into training set and testing set by a 80/20 train-test split ratio. We implemented discriminant analysis to perform classification at the last observation time point (120 months) using the reconstructed values of the 317 variables by MFPCA. Three discriminant methods are applied to the data, i.e., the FDA with MARS, the FDA with BRUTO, and the RDA.

The confusion matrix for each method in the testing set is listed in Tables I, II, and III, respectively. The prediction accuracy is 66.67% for the FDA using MARS, 77.78% for the FDA using BRUTO, and 61.11% for the RDA, respectively.

		Actual		
		CN	MCI	AD
Predicted	CN	47	6	5
	MCI	2	1	1
	AD	6	4	0

TABLE I: Confusion Matrix for FDA-MARS at 120 Months

		Actual		
		CN	MCI	AD
Predicted	CN	53	9	3
	MCI	0	0	0
	AD	2	2	3

TABLE II: Confusion Matrix for FDA-BRUTO at 120 Months

		Actual		
		CN	MCI	AD
Predicted	CN	41	9	3
	MCI	5	1	1
	AD	9	1	2

TABLE III: Confusion Matrix for RDA at 120 Months

		Actual		
		CN	MCI	AD
Predicted	CN	46	6	4
	MCI	5	5	1
	AD	4	0	1

TABLE IV: Confusion Matrix for LSTM at 120 Months

C. Classification by LSTM

The LSTM model is constructed and trained by Keras, the Python deep learning API [49]–[51]. The settings of the LSTM structure are detailed as follows. The model we use contains four layers, namely *hidden layer (LSTM)* $\rightarrow$ *hidden layer (LSTM)* $\rightarrow$ *dense layer* $\rightarrow$ *dense layer*, where $\rightarrow$ represents the connection between layers. The activation functions between LSTM layers we use include the rectified linear activation function (ReLU) for the first dense layer, the

linear function for the second, and the softmax function for the output.

The number of LSTM neurons in each hidden layer is set as 32 and 64 while the neurons in the dense layer are 128 and 256. The batch size is set as 128 and the number of epochs is 1000. Batch normalization is applied since it has been shown to have great benefits in terms of training time and prediction accuracy [52]. We choose the categorical cross-entropy as the loss function, and Adam optimizer [53] for gradient-based optimization of stochastic objective functions.

The confusion matrix for classification of the testing set is in Table IV. Table V shows that the overall prediction accuracy of the LSTM model is 72.22%, the second best among these methods. Moreover, the LSTM correctly predicts the most MCI cases (5 correct predictions; 0 or 1 by FDA or RDA).

Apart from the prediction accuracy and confusion matrix, we can examine the prediction performance by other metrics, the precision and the recall, for example. The precision of one of the three groups (CN, MCI, and AD) is the number of correctly predicted cases of all subjects predicted as a group. The recall for one of the three groups is the number of correctly predicted cases out of the number of subjects actually in that group. That is, if we denote M_{ij} as the number of counts in the (i, j) element of the confusion matrix. The precision and recall are defined as

$$\text{Precision}_i = \frac{M_{ii}}{\sum_j M_{ij}}, \quad \text{Recall}_i = \frac{M_{ii}}{\sum_j M_{ji}}.$$

Table V compares the precision and recall of the three groups in the test set by the four methods. Although the best overall prediction performance belongs to FDA-MARS or FDA-BRUTO, the LSTM has advantages regarding precision and recall measurements. The LSTM can create the highest precision and recall for MCI cases.

		FDA MARS	FDA BRUTO	RDA	LSTM
CN	precision	0.810	0.815	0.774	0.821
	recall	0.855	0.964	0.746	0.836
MCI	precision	0.250	0	0.143	0.455
	recall	0.091	0	0.091	0.455
Dementia	precision	0	0.429	0.167	0.200
	recall	0	0.500	0.333	0.167
accuracy		66.67%	77.78%	61.11%	72.22%

TABLE V: Prediction Evaluation of The four Methods In The Testing Dataset

^a Bold numbers indicate the largest value among the methods.

^b We define the value $0/(0+0+0)$ as 0.

The LSTM model demonstrates notable strengths in Table V, showing its effectiveness across different cognitive groups, particularly in early stage detection, which is of significant clinical importance.

For the CN (cognitively normal) group, the LSTM achieves a precision of 0.821, the highest among all compared methods. This indicates that the model excels at accurately identifying individuals with normal cognition, minimizing the risk of misclassifying patients with MCI or dementia as cognitively

normal. Although its recall of 0.836 is not the highest, it remains robust and contributes to reliable predictions.

In the MCI (mild cognitive impairment) group, the LSTM shows a recall of 0.455, significantly outperforming other methods. This is a crucial advantage, as MCI represents a transition stage between normal cognition and dementia, and accurate identification at this stage is vital for early intervention and treatment. Although its precision for this group is relatively low at 0.143, the ability of the LSTM to recognize this challenging transitional state surpasses that of certain other methods, making it a valuable tool in clinical contexts.

Regarding overall accuracy, the LSTM achieves 72.22%, which, while not the highest, surpasses the RDA method. Given the ability of the LSTM to handle temporal data, this performance highlights its potential to capture disease progression patterns effectively.

The primary advantage of the LSTM lies in its balanced performance across three groups, particularly its superior capability to detect MCI, a critical stage in clinical practice. This balanced and reliable performance could prove to be more practical in real-world applications than focusing on a single metric, highlighting its potential to enhance diagnostic accuracy in cognitive assessments.

VI. CONCLUSION

This study aims to classify subjects into multi-classes, focusing on early detection. Longitudinal, sparse, and irregularly spaced data due to varying subject conditions and observation times present challenges for analysis. To address this, the study proposes reconstructing time-evolving features using FPCA. The discriminant methods perform classification only based on features at a fixed time point, while LSTM considers their progress.

Although the LSTM model does not achieve the highest overall accuracy, it excels in predicting MCI cases. LSTM uses historical data for predictions, unlike FDA and RDA, which classify on the basis of fixed time points. The advantage of LSTM is its ability to capture long-term dependencies. However, the data set used in this study is relatively small and imbalanced, with 75% of subjects being cognitively normal, which potentially limits the ability of the proposed procedure to detect AD. Future improvements include resampling methods for better early detection. Replacing LSTM by other state-of-the-art models is also our future work.

ACKNOWLEDGMENT

We would like to thank Dr. Wei-Wen Hsu at the University of Cincinnati for the assistance with data preprocessing. Yong-Shiuan Lee is grateful for funding support from the National Science and Technology Council, R.O.C. (NSTC 112-2115-M-035-007-MY2). We also extend our sincere gratitude to the anonymous reviewers for their constructive feedback and thoughtful comments.

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National

Institute on Aging (National Institutes of Health Grant U19 AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

REFERENCES

- [1] S. Gauthier, P. Rosa-Neto, J. A. Morais, C. Webster, *et al.*, "World Alzheimer report 2021 - Journey through the diagnosis of dementia." <https://www.alzint.org/resource/world-alzheimer-report-2021/>. Accessed: 2021-09-28.
- [2] S. A. Mofrad, A. J. Lundervold, A. Vik, and A. S. Lundervold, "Cognitive and MRI trajectories for prediction of Alzheimer's disease," *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.
- [3] C. Ledig, A. Schuh, R. Guerrero, R. A. Heckemann, and D. Rueckert, "Structural brain imaging in Alzheimer's disease and mild cognitive impairment: Biomarker analysis and shared morphometry database," *Scientific reports*, vol. 8, no. 1, pp. 1–16, 2018.
- [4] M. Torso, M. Bozzali, G. Zamboni, M. Jenkinson, S. A. Chance, and A. D. N. Initiative, "Detection of Alzheimer's disease using cortical diffusion tensor imaging," *Human Brain Mapping*, vol. 42, no. 4, pp. 967–977, 2021.
- [5] Y. Gupta, R. K. Lama, G.-R. Kwon, M. W. Weiner, P. Aisen, M. Weiner, R. Petersen, C. R. Jack Jr, W. Jagust, J. Q. Trojanowski, *et al.*, "Prediction and classification of Alzheimer's disease based on combined features from apolipoprotein-E genotype, cerebrospinal fluid, MR, and FDG-PET imaging biomarkers," *Frontiers in Computational Neuroscience*, vol. 13, p. 72, 2019.
- [6] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, G. R. Kwon, N. R. C. for Dementia, and A. D. N. Initiative, "Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images," *PLoS One*, vol. 14, no. 10, p. e0222446, 2019.
- [7] I. Saied, T. Arslan, and S. Chandran, "Classification of Alzheimer's disease using RF signals and machine learning," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 6, no. 1, 2022.
- [8] T. P. Garcia and K. Marder, "Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model," *Current Neurology and Neuroscience Reports*, vol. 17, no. 2, pp. 1–9, 2017.
- [9] M. M. Ghazi, M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, L. Sørensen, A. D. N. Initiative, *et al.*, "Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling," *Medical Image Analysis*, vol. 53, pp. 39–46, 2019.
- [10] L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis," *Frontiers in Neuroscience*, p. 779, 2020.

- [11] R. Cui, M. Liu, A. D. N. Initiative, *et al.*, "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease," *Computerized Medical Imaging and Graphics*, vol. 73, pp. 1–10, 2019.
- [12] S. El-Sappagh, T. Abuhmed, S. R. Islam, and K. S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, 2020.
- [13] M. Liu, D. Cheng, W. Yan, A. D. N. Initiative, *et al.*, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers in Neuroinformatics*, vol. 12, p. 35, 2018.
- [14] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, pp. 1–21, 2020.
- [15] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.
- [16] W. Li, X. Lin, and X. Chen, "Detecting Alzheimer's disease based on 4d fMRI: An exploration under deep learning framework," *Neurocomputing*, vol. 388, pp. 280–287, 2020.
- [17] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.
- [18] A. Elazab, C. Wang, M. Abdelaziz, J. Zhang, J. Gu, J. M. Gorriz, Y. Zhang, and C. Chang, "Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: Achievements and future directions," *Expert Systems with Applications*, p. 124780, 2024.
- [19] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.
- [20] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, B. T. Yeo, A. D. N. Initiative, *et al.*, "Predicting Alzheimer's disease progression using deep recurrent neural networks," *NeuroImage*, vol. 222, p. 117203, 2020.
- [21] Y. Li, W.-W. Hsu, and A. D. N. Initiative, "A classification for complex imbalanced data in disease screening and early diagnosis," *Statistics in medicine*, vol. 41, no. 19, pp. 3679–3695, 2022.
- [22] U. Hwang, S.-W. Kim, D. Jung, S. Kim, H. Lee, S. W. Seo, J.-K. Seong, S. Yoon, A. D. N. Initiative, *et al.*, "Real-world prediction of preclinical Alzheimer's disease with a deep generative model," *Artificial Intelligence in Medicine*, vol. 144, p. 102654, 2023.
- [23] S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, "Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 737–749, 2018.
- [24] Y. Li, A. Haber, C. Preuss, C. John, A. Uyar, H. S. Yang, B. A. Logsdon, V. Philip, R. K. M. Karuturi, G. W. Carter, *et al.*, "Transfer learning-trained convolutional neural networks identify novel MRI biomarkers of Alzheimer's disease progression," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 13, no. 1, p. e12140, 2021.
- [25] J. Zou, D. Park, A. Johnson, X. Feng, M. Pardo, J. France, Z. Tomljanovic, A. M. Brickman, D. P. Devanand, J. A. Luchsinger, *et al.*, "Deep learning improves utility of tau PET in the study of Alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 13, no. 1, p. e12264, 2021.
- [26] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Scientific Reports*, vol. 9, no. 1, p. 1952, 2019.
- [27] J. Rudolph, J. Rueckel, J. Döpfert, W. X. Ling, J. Opalka, C. Brem, N. Hesse, M. Ingenerf, V. Koliogiannis, O. Solyanik, *et al.*, "Artificial intelligence-based rapid brain volumetry substantially improves differential diagnosis in dementia," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 16, no. 4, p. e70037, 2024.
- [28] P. Forouzaneshad, A. Abbaspour, C. Fang, M. Cabrerizo, D. Loewenstein, R. Duara, and M. Adjouadi, "A survey on applications and analysis methods of functional magnetic resonance imaging for Alzheimer's disease," *Journal of neuroscience methods*, vol. 317, pp. 121–140, 2019.
- [29] M. Tanveer, B. Richhariya, R. Khan, A. Rashid, P. Khanna, M. Prasad, and C. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 1s, pp. 1–35, 2020.
- [30] A. Kaur, M. Mittal, J. S. Bhatti, S. Thareja, and S. Singh, "A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease," *Artificial Intelligence in Medicine*, p. 102928, 2024.
- [31] J.-L. Wang, J.-M. Chiou, and H.-G. Müller, "Functional data analysis," *Annual Review of Statistics and Its Application*, vol. 3, pp. 257–295, 2016.
- [32] K. Karhunen, "Über lineare methoden in der wahrscheinlichkeit-srechnung," *Annales Academiae Scientiarum Fennicae. Series A. 1: Mathematica-Physica*, vol. 37, pp. 1–79, 1947.
- [33] M. Loève, "Fonctions aléatoires à décomposition orthogonale exponentielle," *La Revue Scientifique*, vol. 84, pp. 159–162, 1946.
- [34] P. Kokoszka and M. Reimherr, *Introduction to Functional Data Analysis*. Boca Raton: Chapman and Hall/CRC, 2017.
- [35] F. Yao, H.-G. Müller, and J.-L. Wang, "Functional data analysis for sparse longitudinal data," *Journal of the American statistical association*, vol. 100, no. 470, pp. 577–590, 2005.
- [36] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *Journal of the American Statistical Association*, vol. 113, no. 522, pp. 649–659, 2018.
- [37] C. Happ-Kurz, *MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains*, 2021. R package version 1.3-9.
- [38] C. Happ-Kurz, "Object-oriented software for functional data," *Journal of Statistical Software*, vol. 93, no. 5, pp. 1–38, 2020.
- [39] K. Li and S. Luo, "Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data," *Statistics in Medicine*, vol. 38, no. 24, pp. 4804–4818, 2019.
- [40] K. E. Roach, V. Pedoia, J. J. Lee, T. Popovic, T. M. Link, S. Majumdar, and R. B. Souza, "Multivariate functional principal component analysis identifies waveform features of gait biomechanics related to early-to-moderate hip osteoarthritis," *Journal of Orthopaedic Research*, vol. 39, no. 8, pp. 1722–1731, 2021.
- [41] T. Hastie, R. Tibshirani, and A. Buja, "Flexible discriminant analysis by optimal scoring," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1255–1270, 1994.
- [42] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics*, vol. 23, no. 1, pp. 73–102, 1995.
- [43] T. Hastie, "[Flexible Parsimonious Smoothing and Additive Modeling]: Discussion," *Technometrics*, vol. 31, no. 1, pp. 23–29, 1989.
- [44] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.
- [45] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [47] J. K. Kueper, M. Speechley, and M. Montero-Odasso, "The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review," *Journal of Alzheimer's Disease*, vol. 63, no. 2, pp. 423–444, 2018.
- [48] R. C. Mohs, D. Knopman, R. C. Petersen, S. H. Ferris, C. Ernesto, M. Grundman, M. Sano, L. Bieliauskas, D. Geldmacher, C. Clark, *et al.*, "Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's disease assessment scale that broaden its scope," *Alzheimer Disease and Associated Disorders*, 1997.
- [49] F. Chollet *et al.*, "Keras." [urlhttps://github.com/fchollet/keras](https://github.com/fchollet/keras), 2015.
- [50] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [51] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, pp. 265–283, 2016.
- [52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, pp. 448–456, PMLR, 2015.
- [53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, 2015.