A Hybrid 3D-CNN and BiLSTM Framework for Early Alzheimer's Disease Prediction Using Neuroimaging and Clinical Biomarkers

Dr. Jegan.J[1,2], Shashi Kant Gupta[3]

[1]Lincoln University College, 47301, Petaling Jaya, Selangor, Malaysia;

[2]School of Computing, SRM Institute of Science and Technology Tiruchirappalli Campus, Tiruchirappalli, 621105, Tamil Nadu, India;

[3]Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor, Malaysia

Email ID: [1]pdf.jegan@lincoln.edu.my, [2]jeganj@srmist.edu.in, [3]shashigupta@lincoln.edu.my

Abstract: Alzheimer's disease (AD) is the most common cause of dementia, accounting for 60-70% of cases and affecting more than 55 million people worldwide. Early and accurate differentiation of AD, Mild Cognitive Impairment (MCI) and Cognitively Normal (CN) subjects is essential for timely intervention. Here, we propose a hybrid deep learning framework that integrates a 3D Convolutional Neural Network (3D-CNN) and a Bidirectional Long Short-Term Memory (BiLSTM) network to jointly learn spatial atrophy patterns from structural MRI and temporal trends from longitudinal cognitive and cerebrospinal fluid (CSF) biomarkers. Feature-level late fusion combines both representations before classification. The framework achieved an accuracy of 94.8%, sensitivity of 93.6%, specificity of 95.9%, F1-score of 94.3% and AUC of 0.971 with 600 subjects from ADNI database, which outperformed the SVM, Random Forest and standalone 3D-CNN/BiLSTM baselines. The ablation studies and Wilcoxon signed-rank tests ($p < 0.01$) validate the statistically significant gain by combining spatial and temporal information, supporting the potential of the framework for automated three-class AD screening.

Keywords: Alzheimer's Disease; 3D-CNN; BiLSTM; Multimodal Fusion; Neuroimaging; ADNI

Introduction

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder that is responsible for 60-70% of the 55 million dementia cases reported worldwide, a number the WHO predicts will reach over 139 million in three decades. AD is a personal tragedy, and a huge burden on healthcare systems and caregivers as it progresses from mild cognitive decline to complete incapacity.

Pathologically, AD is characterized by the presence of extracellular amyloid-beta plaques and intracellular tau tangles, which lead to neuronal loss in the hippocampus, entorhinal cortex and temporal lobes, and measurable atrophy on structural MRI [1]. Standard diagnosis relies on cognitive tests (e.g., MMSE, ADAS-Cog, CDR-SB) and CSF biomarkers such as Aβ42 and phosphorylated tau [2], but these approaches are labour-intensive, subjective and difficult to scale for population-level screening.

Deep learning is starting to bridge this gap. 3D-CNNs learn volumetric spatial features from MRI without hand-crafted descriptors, and effectively capture atrophy patterns, but treat the brain as a static snapshot. BiLSTM networks capture temporal dependencies in longitudinal clinical data in both directions, but ignore the anatomical structure. In this paper, we propose a hybrid architecture that combines spatial features from 3D-CNN and temporal features from BiLSTM at the feature level to classify subjects into AD, MCI and CN.

The major contributions are:

- A hybrid 3D-CNN + BiLSTM framework for automatic three-class AD classification using complementary neuroimaging and longitudinal data .
- Integrated approach to joint use of structural MRI, longitudinal cognitive scores and CSF biomarkers.
- Late fusion of spatial and temporal representations at feature level for improved discriminative power.

• Ablation studies and Wilcoxon signed-rank testing ($p<0.01$) demonstrating statistically significant improvements over baseline and unimodal deep learning models.

Related Work

Computational AD diagnosis has progressed from hand-engineered machine learning pipelines to end-to-end deep learning. Klöppel et al. [3] showed discrimination between AD patients and controls using SVM based on structural MRI. Sarica et al. [4] reviewed Random Forest approaches and reported the difficulty of these methods to model complex, non-linear multimodal relationships. These methods were shown to be viable, but still bottlenecked by manual feature engineering.

Deep learning eliminated this bottleneck. 3D-CNNs were directly used by Payan and Montana [5] on MRI volumes and achieved better performance than traditional machine learning. Deep networks trained on single modality MRI were able to classify AD, MCI and controls, and naturally attend to the hippocampus and temporal lobes, as shown by Basaia et al. [6]. But these models assume brain structure is static and do not incorporate progression of disease over time.

Recurrent architectures were proposed to model temporal dynamics. Ghazi et al. [7] proposed LSTM-based models robust to missing longitudinal records. Cui et al. [8] used BiLSTM to model cognitive trajectories and showed that bidirectional processing improves temporal representation learning, but imaging data was not available. Studies of multimodal fusion fuse both signal types – Liu et al. [9] fused MRI, PET and CSF biomarkers to outperform unimodal models, and Wang et al. [10] combined CNN and LSTM components on longitudinal MRI. However, most fusion work focuses on binary AD-vs-control classification as opposed to the clinically relevant three-class AD/MCI/CN problem, which is the gap this study addresses.

Table 1. Comparison of this work with related studies

Study	Imaging	Temporal	Multimodal	3-Class
Klöppel et al. [3]	Yes	No	No	No
Basaia et al. [6]	Yes	No	No	Yes
Cui et al. [8]	No	Yes	No	No
Liu et al. [9]	Yes	No	Yes	No
Wang et al. [10]	Yes	Yes	No	No
This work	Yes	Yes	Yes	Yes

Key Contribution

Unlike previous works which consider spatial and temporal architectures independently, we propose a joint 3D-CNN and BiLSTM framework with feature-level late fusion for three-class AD/MCI/CN classification, which is a more clinically demanding scenario than the binary comparisons considered in previous fusion studies. We contribute to the state of the art by showing through systematic ablation that the joint spatiotemporal representation yields consistent and statistically significant improvements over each unimodal branch.

IV. Method, Experiments and Results

Data were obtained from the public ADNI database consisting of 600 subjects equally distributed over AD, MCI and CN classes. Structural MRI volumes were skull stripped, spatially normalized to the MNI template, intensity normalized, and transformed to gray-matter density maps. “Longitudinal clinical variables (MMSE, ADAS-Cog-13, CDR-SB, Aβ42, p-tau) were z-score normalized and interpolated for missing values.

The framework is depicted in Fig. 1 and consists of 6 modules. The 3D-CNN branch uses three 3D convolutional layers (32, 64, 128 filters) with batch normalization, ReLU activation, and max pooling operations, which are then followed by a global average pooling operation to generate a 128-dimensional spatial feature vector. The BiLSTM branch processes the sequence of per-visit clinical feature vectors bidirectionally to obtain a 512-dimensional temporal feature vector. The two vectors are concatenated (640-d) and fed through two dense layers (512, 128 units, ReLU, dropout 0.5) to a 3-way softmax classifier. Training was done with weighted categorical cross-entropy, Adam (lr = 0.0001, batch size 16, max 100 epochs), learning-rate decay and early stopping. Evaluation was done using stratified 5-fold cross-validation.

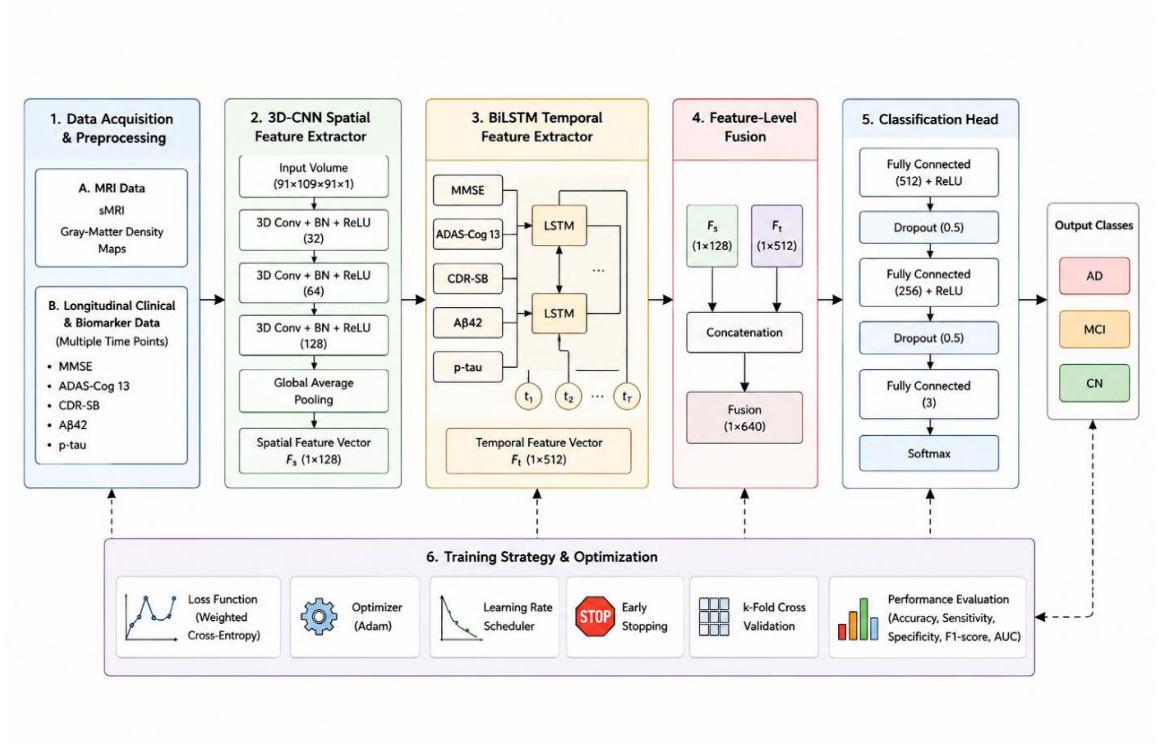

Fig. 1. System architecture of the proposed hybrid framework.

The proposed framework was compared with the SVM, Random Forest, standalone BiLSTM and standalone 3D-CNN baselines with same data splits. Table 2 summarizes the results in terms of accuracy, sensitivity, specificity, F1-score and AUC.

Table 2. Comparative performance against baseline models

Model	Acc. (%)	Sens. (%)	Spec. (%)	F1 (%)	AUC
SVM	85.4	83.2	87.1	84.1	0.892
Random Forest	87.2	85.6	88.9	86.3	0.914
Standalone BiLSTM	90.1	88.4	91.6	89.2	0.942
Standalone 3D-CNN	92.7	91.2	93.8	92.0	0.958
Proposed 3D-CNN+BiLSTM	94.8	93.6	95.9	94.3	0.971

The proposed framework achieved 92.7% accuracy, outperforming the strongest baseline (standalone 3D-CNN) by 2.1 points and SVM by 9.4 points. The sensitivity of 93.6% and specificity of 95.9% indicate good balance and ability to detect the disease with minimal false alarms. The AUC of 0.971 indicates strong discriminative power between classes.

An ablation study separating the components showed an accuracy increase from 86.3% (clinical biomarkers only) to 91.4% (MRI only) to 92.2% (MRI + BiLSTM without fusion optimization), and finally to 94.8% for the full framework. Wilcoxon signed-rank tests ($p < 0.01$) confirmed that the improvement at each step is statistically significant.

Discussions

These results suggest that spatial and temporal information are complementary, rather than redundant: MRI-derived atrophy patterns and longitudinal biomarker trajectories each capture distinct aspects of disease progression, and their fusion consistently improves three-class discrimination beyond either branch alone. This trade-off between sensitivity and specificity is particularly important for screening tests used in clinics, because both missing a diagnosis and false alarm can be very costly.

Conclusions

1. Problem statement/motivation: Automated three-class (AD/MCI/CN) diagnosis is a labour intensive and subjective process that limits population-scale screening.
2. Method: Hybrid 3D-CNN and BiLSTM architecture with feature-level late fusion of structural MRI and longitudinal clinical/CSF biomarkers.
3. Main results: 94.8% accuracy, 93.6% sensitivity, 95.9% specificity, 94.3% F1-score and 0.971 AUC on 600 ADNI subjects, with statistically significant gains ($p < 0.01$) over unimodal and conventional baselines. 4. Limitations and future work: Evaluation is limited to one cohort (ADNI); future work will explore attention mechanisms for interpretability, integration of PET and genomic data, and multi-site, multi-ethnic validation.

References

1. J. Frisoni, N. Fox, C. Jack Jr., P. Scheltens, and P. Thompson, "The clinical use of structural MRI in Alzheimer disease," Nature Reviews Neurology, vol. 6, no. 2, pp. 67-77, 2010.
2. H. Hampel, K. Bürger, S. J. Teipel, A. L. Bokde, H. Zetterberg, and K. Blennow, "Core candidate neurochemical and imaging biomarkers of Alzheimer's disease," Alzheimer's & Dementia, vol. 4, no. 1, pp. 38-48, 2008.
3. S. Klöppel, C. M. Stonnington, C. Chu, et al., "Automatic classification of MR scans in Alzheimer's disease," Brain, vol. 131, no. 3, pp. 681-689, 2008.
4. A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," Frontiers in Aging Neuroscience, vol. 9, p. 329, 2017.
5. A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," arXiv preprint arXiv:1502.02506, 2015.
6. S. Basaia, F. Agosta, L. Wagner, et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," NeuroImage: Clinical, vol. 21, p. 101645, 2019.
7. A. Ghazi, M. Nielsen, A. Pai, 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.
8. R. Cui, X. Liu, and L. Li, "Modeling longitudinal Alzheimer's disease progression using bidirectional LSTM networks," IEEE Access, vol. 8, pp. 29981-29990, 2020.
1. S. Liu, S. Liu, W. Cai, et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, pp. 1132-1140, 2015.
2. S. Wang, Z. Shen, and Y. Tang, "CNN-LSTM based deep learning framework for Alzheimer's disease diagnosis using longitudinal MRI," Computerized Medical Imaging and Graphics, vol. 73, pp. 1-10, 2019.