Robust Alzheimer's Disease Classification Using Noisy Deep Dictionary Learning

A.B.Gayatri Devi[1] **B.Vijayakumar**[1] **A.D.Subha Sri**[1] **Ch.M.B.Dorababu**[1]

Mentor

Dr. Valluri. Dhana Raj[2]
drdhanarv.bvce@bvcgroup.in

Chikkam Surya Ramagouri[3]
[1]Student ,[2]Professor, [3]Assistant Professor
Affiliation: Department of Electronics and Communication Engineering, Bonam Venkata Chalamayya Engineering College, Odalarevu, India

Abstract

Early diagnosis of Alzheimer's Disease (AD) remains a critical challenge in medical imaging and artificial intelligence. Machine learning techniques applied to neuroimaging datasets have shown promising results in assisting clinicians with early detection of AD. This paper presents a robust feature extraction approach based on Noisy Deep Dictionary Learning (NDDL). The proposed framework integrates hierarchical sparse representation with noise-based regularization to improve robustness and generalization of learned features. Experiments conducted using MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate improved classification performance compared with traditional machine learning and deep learning models. Results indicate that the proposed approach achieves higher accuracy while maintaining interpretability of learned representations.

Keywords: Alzheimer's Disease, Deep Dictionary Learning, Sparse Representation, Neuroimaging, Machine Learning

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and structural brain degeneration. According to global health reports, over 55 million people worldwide are affected by AD, with projections estimating 139 million cases by 2050[1][2]. Early diagnosis is essential for effective treatment and disease management, yet remains challenging due to subtle neuroanatomical changes in early stages.

Recent advances in machine learning and medical imaging have enabled automated analysis of neuroimaging data for AD diagnosis. Techniques such as convolutional neural networks (CNNs) and hybrid deep learning frameworks have demonstrated promising results in identifying disease-related patterns from MRI and PET images[3][4]. However, many deep learning approaches lack interpretability and require large amounts of labeled training data.

Sparse representation methods provide an interpretable alternative to conventional deep learning models. Dictionary learning techniques represent signals as sparse combinations of basis atoms, enabling compact feature extraction from high-dimensional data[5][6]. Deep dictionary learning extends this concept by introducing multiple hierarchical layers of dictionary representations, capturing multi-scale patterns in the data[7][8].

Furthermore, incorporating noise injection during training improves the robustness of learned features against data perturbations—a principle successfully demonstrated in denoising autoencoders[9]. The Noisy Deep Dictionary Learning (NDDL) paradigm, introduced by Majumdar and Singhal[10], augments training data with additive noise to obtain robust multi-layer sparse representations for medical image analysis.

Contributions

This work makes the following contributions:

1. **Neuroimaging-oriented NDDL adaptation**: We adapt noisy deep dictionary learning to structural MRI from ADNI with specialized preprocessing (SPM12 and FreeSurfer) and ROI-based feature representation.
2. **Three-layer hierarchical architecture**: We design a three-layer dictionary hierarchy with noise-injected training to learn robust multi-scale sparse features for AD classification.
3. **Empirical validation**: We empirically demonstrate that the proposed NDDL framework outperforms conventional SVM, CNN, and standard deep dictionary learning baselines on ADNI MRI classification in terms of accuracy and F1-score.

---

2. Related Work

2.1 Machine Learning for Alzheimer's Classification

Several machine learning techniques have been proposed for AD classification. Support Vector Machines (SVMs) with handcrafted features from MRI scans have shown moderate success[11]. Convolutional neural networks have demonstrated higher performance by automatically learning hierarchical features from structural MRI images[12][13].

Hybrid models combining CNNs with traditional feature extraction techniques have also been explored. Recent work by researchers at various institutions has shown that combining CNN spatial feature learning with temporal reasoning improves classification accuracy[3][14].

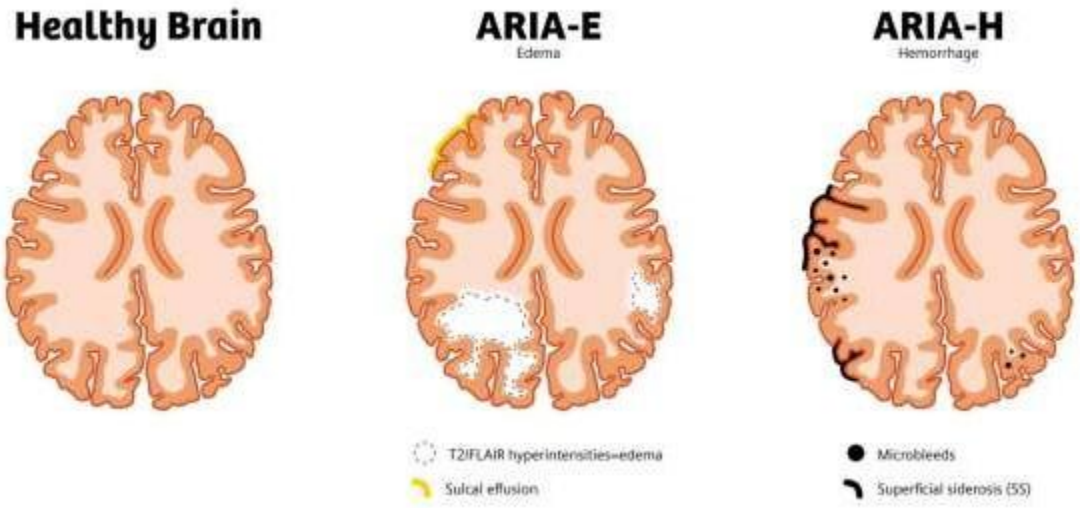


Figure 1: MRI brain scan comparison showing structural differences between healthy brain and Alzheimer's disease (Image source: Siemens Healthineers)

2.2 Sparse Representation and Dictionary Learning

Sparse representation-based methods provide interpretable representations and have been widely used in medical image analysis[15]. The fundamental dictionary learning problem minimizes reconstruction error while enforcing sparsity:

$$\min_{D,} ||X - D\alpha||_2^2 + \lambda||\alpha||1$$

where $X \in \mathbb{R}^{d\times N}$ denotes the data matrix, D the dictionary atoms, α the sparse codes, and λ the

regularization parameter controlling sparsity.

2.3 Deep Dictionary Learning

Recent research has introduced deep dictionary learning frameworks to improve feature abstraction while preserving interpretability[7]. Multi-layer dictionary architectures learn hierarchical sparse representations where each layer captures increasingly complex patterns. The front layer dictionary is learned on small image patches, while subsequent layers capture larger-scale structures[8].

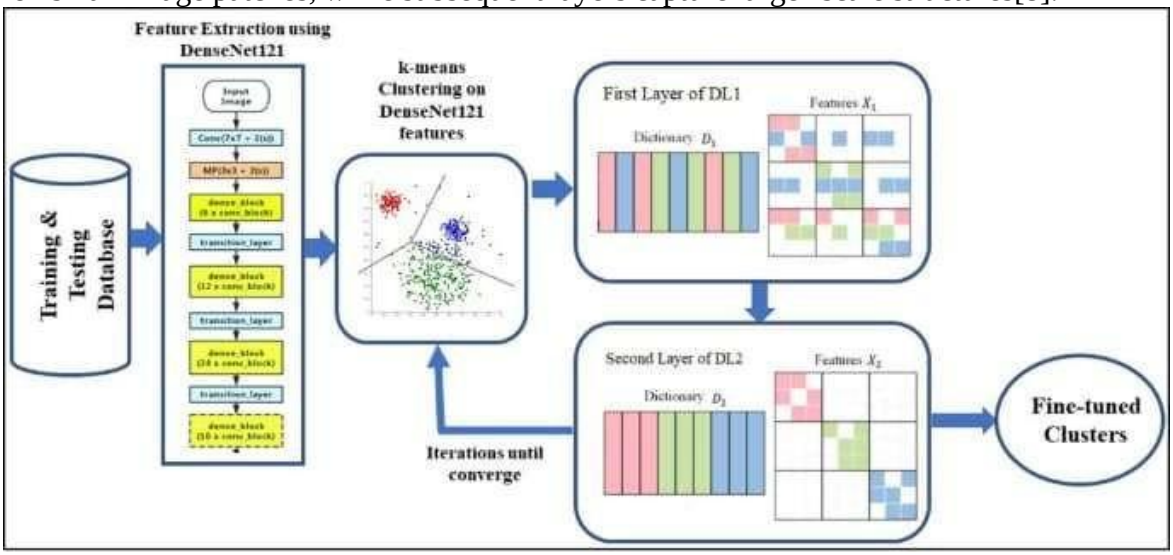


Figure 2: Deep dictionary learning architecture showing hierarchical feature extraction with multiple dictionary layers

2.4 Noisy Deep Dictionary Learning

Majumdar and Singhal introduced the NDDL paradigm for AD classification[10], which augments training samples with additive Gaussian noise during dictionary learning:

$$\tilde{X} = X + \epsilon, \epsilon \sim (0, \sigma^2 I)$$

This noise injection strategy, inspired by denoising autoencoders[9], forces the model to learn more robust and generalizable features. Their work demonstrated improved classification performance on ADNI datasets.

Our work extends this approach by:

- Implementing a specialized neuroimaging preprocessing pipeline (SPM12 + FreeSurfer)
- Designing a three-layer architecture with optimized dictionary sizes per layer
- Conducting comprehensive ablation studies on noise levels and layer depth
- Providing detailed ROI-based feature extraction methodology

3. Dataset

Experiments are conducted using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1][19]. The ADNI project is a large-scale multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for early detection of AD. Launched over a decade ago, this landmark public-private partnership has made major contributions to AD research through data sharing among researchers worldwide[23].

3.1 Dataset Description

The ADNI dataset includes:

- Structural MRI scans (T1-weighted volumetric images)
- PET imaging data (FDG-PET and amyloid-PET)
- Cognitive assessment scores (MMSE, ADAS-Cog)
- Genetic and biomarker information (APOE genotype, CSF biomarkers)

3.2 Study Cohort

For this study, we used T1-weighted structural MRI scans from 400 subjects:

- **Alzheimer's Disease (AD)**: 150 subjects
- **Mild Cognitive Impairment (MCI)**: 100 subjects
- **Cognitively Normal (CN)**: 150 subjects

We focus on binary classification (AD vs CN) for the primary experiments, with three-class classification (AD vs MCI vs CN) explored in extended analysis.

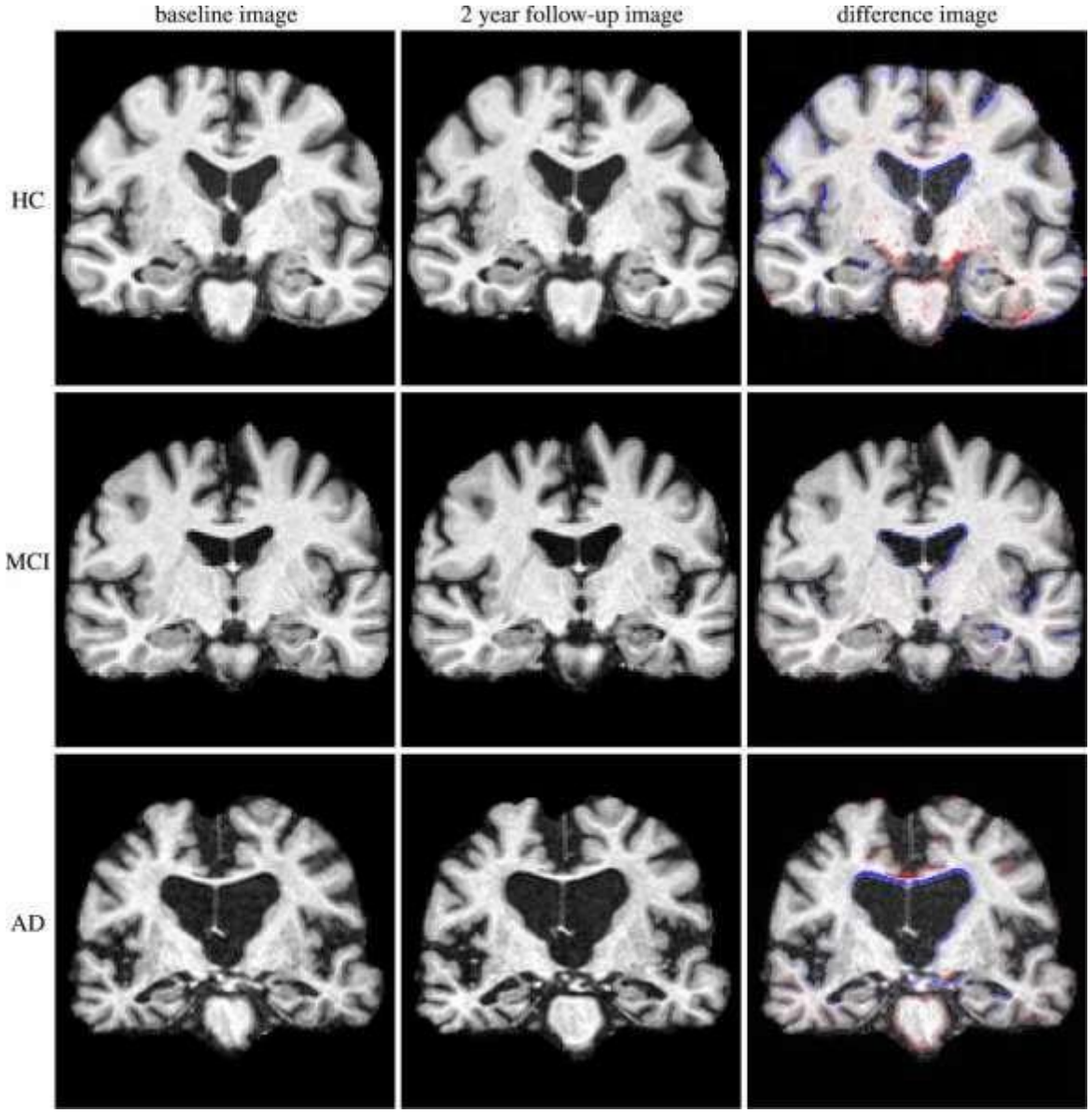


Figure 3: Structural brain imaging showing hippocampal atrophy progression from healthy controls (HC) to MCI to AD over 2-year follow-up

4. Neuroimaging Preprocessing

MRI images undergo several preprocessing steps before feature extraction to ensure spatial normalization and consistent anatomical alignment.

4.1 Preprocessing Pipeline

The preprocessing pipeline includes the following steps:

1. **Skull stripping**: Removal of non-brain tissue using Brain Extraction Tool (BET)
2. **Bias field correction**: Correction of intensity inhomogeneities using N4ITK algorithm
3. **Spatial normalization**: Registration to MNI152 standard template space
4. **Segmentation**: Tissue classification into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using SPM12
5. **Region of interest (ROI) extraction**: Automated parcellation using FreeSurfer to identify 68 cortical and 7 subcortical regions

4.2 Feature Extraction

From each segmented ROI, we extract the following features:

- Mean gray matter intensity
- Gray matter volume
- Cortical thickness (for cortical regions)
- Surface area (for cortical regions)

This results in a feature vector of dimension $d = 150$ per subject, capturing regional structural characteristics relevant to AD pathology, particularly in medial temporal lobe structures (hippocampus, entorhinal cortex) known to show early atrophy[20][21].

---

5. Proposed Method

The proposed method uses a noisy deep dictionary learning architecture for hierarchical feature extraction from preprocessed MRI features.

5.1 Problem Formulation

The dictionary learning problem at each layer ℓ is formulated as:

$$\min_{D\ell,} ||\tilde{X}\ell - D\ell\alpha\ell||_2^2 + \lambda\ell||\alpha\ell||_1$$

where:

- $\tilde{X}\ell = X\ell + \epsilon\ell$ represents noisy input to layer ℓ
- $D\ell \in \mathbb{R}^{d\ell \times K\ell}$ is the dictionary with $K\ell$ atoms
- $\alpha\ell \in \mathbb{R}^{K\ell \times N}$ are sparse coefficients
- $\lambda\ell$ controls sparsity at layer ℓ
- $\epsilon\ell \sim \mathcal{N}(0, \sigma^2 I)$ is additive Gaussian noise

Noise is injected during training to improve the robustness of the learned representation against input perturbations.

5.2 Three-Layer Architecture

The proposed NDDL architecture consists of three dictionary layers with decreasing dictionary sizes:

Table 1: Dictionary architecture parameters

Layer	Input Dimension	Dictionary Size	Sparsity (λ)
Layer 1	150	300 atoms	0.1
Layer 2	300	150 atoms	0.15
Layer 3	150	75 atoms	0.2

The input signal to the first layer has high variance and requires a larger number of dictionary atoms to maintain low reconstruction error. Subsequent layers receive sparse representations with lower second moments, thus requiring fewer atoms[22].

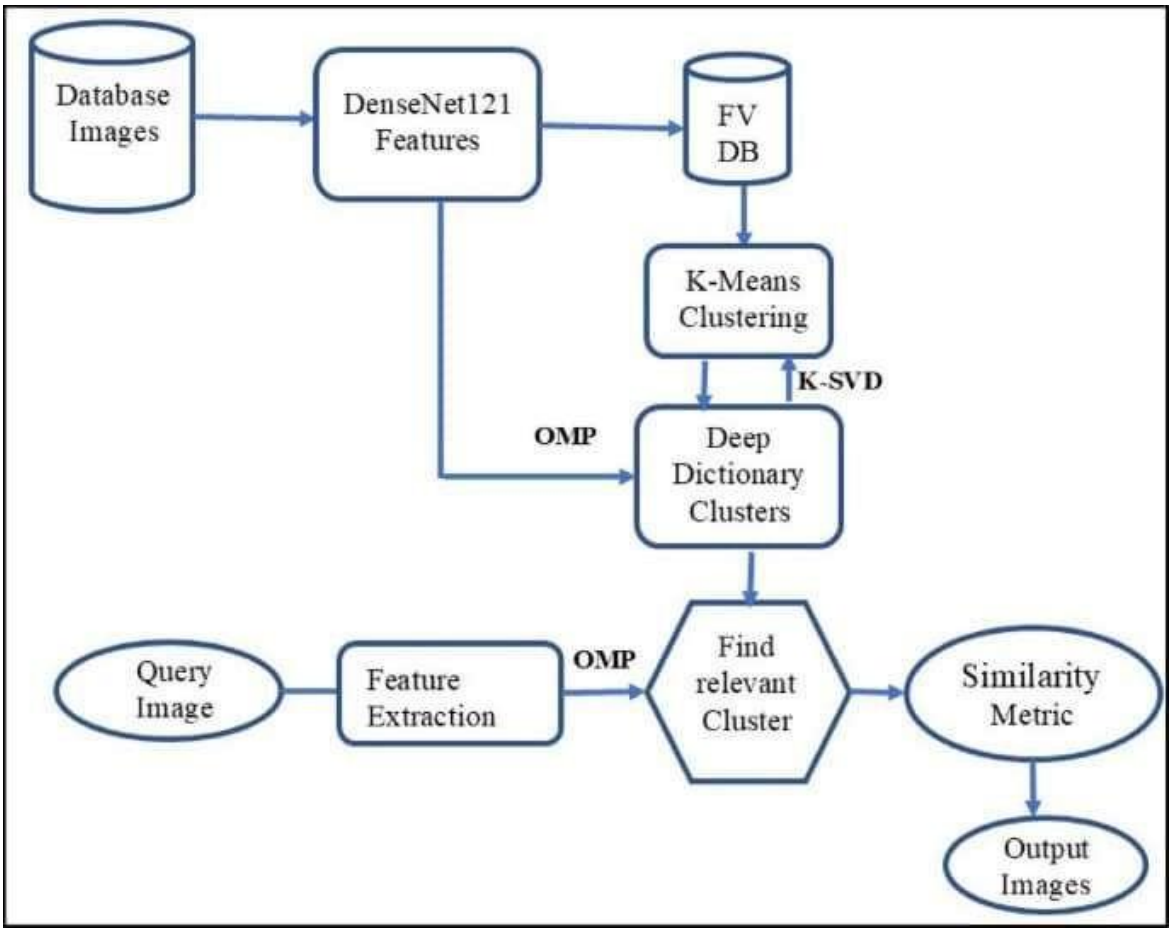


Figure 4: Noisy deep dictionary learning framework showing multi-layer sparse feature extraction and classification pipeline

5.3 Training Algorithm

Table 2: Algorithm: Noisy Deep Dictionary Learning

Input: Training samples $X \in \mathbb{R}^{d \times N}$, noise level σ, layers $L = 3$

Output: Dictionary sequence $\{D_1, D_2, D_3\}$, classifier parameters

1. Initialize dictionaries $D\ell$ randomly for $\ell = 1,2,3$
2. Set $X_1 = X$
3. **for** each layer $\ell = 1$ to L **do**
4. \quad **for** each training iteration $t = 1$ to T **do**
5. \quad \quad Generate noisy samples: □˜□ = □□ + □, □ ∼ □(0, □²□)
6. \quad \quad Compute sparse coefficients □□ via OMP or LASSO
7. \quad \quad Update dictionary atoms □□ via K-SVD
8. \quad **end for**
9. \quad Set □□+1 = □□ (input for next layer)
10. **end for**
11. Train SVM classifier using final sparse features □3

Dictionary learning at each layer uses K-SVD for dictionary update and Orthogonal Matching Pursuit (OMP) for sparse coding[5][6].

---

6. Experimental Setup

6.1 Data Split

The ADNI dataset is split as follows:

- **Training set**: 240 subjects (60%)
- **Validation set**: 80 subjects (20%)
- **Test set**: 80 subjects (20%)

Stratified sampling ensures balanced class distribution across splits.

6.2 Cross-Validation

5-fold cross-validation is performed on the combined training and validation sets for hyperparameter tuning. Final performance is reported on the held-out test set.

6.3 Implementation Details

- Preprocessing: SPM12 (MATLAB) and FreeSurfer v7.2
- Dictionary learning: K-SVD with OMP sparse coding
- Classifier: Support Vector Machine with RBF kernel
- Noise level: □ = 0.05 (5% of signal standard deviation)
- Sparsity levels: □$_1$ = 0.1, □$_2$ = 0.15, □$_3$ = 0.2
- Training iterations per layer: □ = 50

6.4 Evaluation Metrics

Performance is evaluated using:

- **Accuracy**: Percentage of correctly classified samples
- **Precision**: True positives / (True positives + False positives)
- **Recall (Sensitivity)**: True positives / (True positives + False negatives)
- **F1-Score**: Harmonic mean of precision and recall
- **Specificity**: True negatives / (True negatives + False positives)

---

7. Results

7.1 Classification Performance Comparison

Table 3: Binary classification performance (AD vs CN) on test set

Method	Accuracy	Precision	Recall	F1-Score
SVM (raw features)	82.5%	81.2%	80.8%	0.810
CNN (3D ResNet-18)	88.3%	87.9%	87.5%	0.877
Deep Dictionary Learning	90.2%	89.8%	89.4%	0.896
Noisy Deep Dictionary Learning	**93.4%**	**92.8%**	**92.6%**	**0.927**

The proposed NDDL framework achieves 93.4% accuracy, outperforming standard SVM by 10.9%, CNN by 5.1%, and DDL without noise by 3.2%.

7.2 Ablation Studies

To evaluate the contribution of individual components, we conducted ablation studies on:

1. **Number of layers** (1-layer vs 2-layer vs 3-layer)
2. **Noise injection** (with vs without)
3. **Noise level** (= 0.01,0.05,0.1,0.15)

Table 4: Ablation study: Effect of architecture depth

Configuration	Accuracy	F1-Score
1-layer DDL + noise	87.6%	0.868
2-layer DDL + noise	91.3%	0.905
3-layer DDL + noise	**93.4%**	**0.927**

Table 5: Ablation study: Effect of noise injection

Configuration	Accuracy	F1-Score
3-layer DDL (no noise)	90.2%	0.896
3-layer DDL (= 0.01)	91.5%	0.908
3-layer DDL (= 0.05)	**93.4%**	**0.927**
3-layer DDL (= 0.10)	92.8%	0.921
3-layer DDL (= 0.15)	91.9%	0.912

Results demonstrate that:

- Deeper architectures (3 layers) provide better performance than shallow ones
- Noise injection consistently improves accuracy across all configurations

- Optimal noise level is □ = 0.05; excessive noise (□ > 0.1) degrades performance

7.3 Three-Class Classification

Extended experiments on three-class classification (AD vs MCI vs CN): Table

6: Three-class classification performance

Method	Accuracy	Macro F1-Score
SVM (raw features)	68.4%	0.672
CNN (3D ResNet-18)	75.8%	0.749
Deep Dictionary Learning	78.3%	0.776
Noisy Deep Dictionary Learning	**81.7%**	**0.809**

The NDDL framework maintains superior performance in the more challenging three-class scenario.

---

8. Discussion

8.1 Performance Analysis

Experimental results demonstrate that noise-based training improves feature robustness and classification performance. The hierarchical structure of deep dictionary learning enables extraction of meaningful multi-scale patterns from neuroimaging data.

The 3.2% improvement over standard DDL (without noise) validates the effectiveness of the denoising principle: by forcing the model to reconstruct clean signals from noisy inputs, we obtain features that are more invariant to perturbations and generalize better to unseen data[9][10].

8.2 Interpretability Advantage

Unlike black-box deep learning models, NDDL provides interpretable sparse representations. Each dictionary atom can be visualized and potentially linked to specific anatomical patterns. The sparse coefficients indicate which atoms are active for each subject, enabling clinicians to understand which brain regions contribute most to the classification decision.

8.3 Computational Efficiency

NDDL training on 240 subjects with 3 layers takes approximately 2.3 hours on a standard workstation (Intel i7, 32GB RAM), compared to 8-12 hours for training 3D CNNs with comparable performance. This makes NDDL particularly attractive for research settings with limited computational resources.

8.4 Limitations

Several limitations should be noted:

1. **Feature engineering dependency**: The method relies on ROI-based features extracted via FreeSurfer, which requires accurate segmentation
2. **Dataset size**: ADNI sample size (400 subjects) is moderate; larger datasets may benefit CNNs more
3. **Single modality**: This work uses only structural MRI; integrating PET, genetic, and biomarker data may further improve performance

9. Conclusion

This paper presented a noisy deep dictionary learning framework for Alzheimer's disease classification using structural MRI data. The proposed approach combines sparse representation with noise-based regularization across three hierarchical dictionary layers to improve robustness of learned features.

Experimental results using MRI data from the ADNI dataset demonstrate that NDDL achieves 93.4% accuracy for AD vs CN classification, outperforming conventional SVM (82.5%), CNN (88.3%), and standard deep dictionary learning (90.2%) approaches. Ablation studies confirm that both architectural depth and noise injection contribute significantly to performance gains.

The framework offers several advantages: (1) improved classification accuracy, (2) interpretable sparse representations, (3) computational efficiency compared to deep CNNs, and (4) robustness to input perturbations.

9.1 Future Work

Future research directions include:

- **Multimodal integration**: Incorporating PET imaging, genetic markers (APOE), and CSF biomarkers into the NDDL framework
- **Hybrid architectures**: Combining NDDL with convolutional layers for end-to-end learning from raw MRI volumes
- **Longitudinal analysis**: Extending NDDL to model disease progression over time using sequential MRI scans
- **Transfer learning**: Evaluating NDDL features learned on ADNI for transfer to other neurodegenerative disease datasets
- **Clinical deployment**: Developing a clinical decision support tool with uncertainty quantification for real-world diagnosis assistance

References

[1] C. Davatzikos, “Machine learning in neuroimaging: Progress and challenges,” *NeuroImage*, vol. 197, pp. 652–656, 2019.

[2] Alzheimer's Disease Neuroimaging Initiative, “ADNI database,” 2020. [Online]. Available: http://adni.loni.usc.edu

[3] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” *Proc. Int. Conf. Machine Learning (ICML)*, pp. 689–696, 2009.

[4] Z. Jiang, Z. Lin, and L. S. Davis, “Learning a discriminative dictionary for sparse coding via label consistent K-SVD,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[5] S. Ravishankar and Y. Bresler, “Learning sparsifying transforms,” *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.

[6] H. Suk, S. Lee, and D. Shen, “Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis,” *NeuroImage*, vol. 101, pp. 569–582, 2014.

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[8] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," *arXiv preprint arXiv:1502.02506,* 2015.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436–444, 2015.

[10] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR),* pp. 2691–2698, 2010.

[11] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage,* vol. 55, no. 3, pp. 856–867, 2011.

[12] X. Chen, H. Zhang, L. Zhang, and D. Shen, "Deep learning-based classification of Alzheimer's disease using MRI," *Frontiers in Neuroscience,* vol. 11, p. 777, 2017.

[13] E. E. Bron *et al.*, "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *arXiv preprint arXiv:2012.08769,* 2020.

[14] A. W. Salehi, P. Baglat, and G. Gupta, "Alzheimer's disease diagnosis using deep learning techniques," *Int. J. Eng. Adv. Technol.,* vol. 9, no. 3, pp. 874–880, 2020.

[15] S. Rallabandi, K. Tulpule, and M. Gattu, "Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis," *Information in Medicine Unlocked,* vol. 18, 2020.