Report on AI Based Differential Diagnosis of dementia Etiologies on multi-modal data

Parth Sharma[1][0009−0006−3168−3643]

Manipal University Jaipur,India
parthshr370@gmail.com

1 Summary

1. The paper presents an AI-based diagnostic model designed to classify cognitive states—Normal Cognition (NC), Mild Cognitive Impairment (MCI), and Dementia—with high accuracy.
2. The AUROC score for this experiment is 0.94, which indicates that the model has a high level of diagnostic accuracy. An AUROC closer to 1 means it can distinguish between different dementia types easily. It maintains its high level of performance across all three states: normal cognition, MCI, and dementia.
3. This model tries to mimic the real-world diagnosis process and incorporates those features into the model, such as a normal patient journey of tests, medical history, and scans that confirm a particular disease.
4. There is an overlap of etiologies, which is crucial for formulating early personalized management strategies.
5. The model addresses the complexities arising from overlapping etiologies in dementia diagnosis.
6. The improvement is quantified as 26.25%, which suggests that the integration of AI into their assessment process results in more than a one-quarter increase in their ability to correctly diagnose cases compared to relying solely on clinical judgment and traditional methods (for a subset of 100 patients).
7. The model incorporates multiple aspects of data types, such as:
   – Medication Use
   – Medical History
   – Functional Evaluation
   – Neuroimaging
   – Datasets consisting of 51,269 participants over 9 geographies
8. The model utilizes a transformer architecture called SWIN UNETR [2] , which is generally used to process 3D volumetric data such as MRI or CT scans.

2 Research Questions

1. How can an AI model effectively improve early and accurate diagnosis of various dementia types by leveraging multimodal clinical data, given the overlap of etiologies and limitations in current diagnostic practices?

2. Many diseases have overlapping symptoms, which might indicate the treatment direction to doctors without wasting resources on recurring symptoms.
3. Some patients with Alzheimer's may be candidates for anti-amyloid treatment. However, if the patient has other types of dementia, like Vascular Dementia, amyloid treatment can complicate matters, creating amyloid-related imaging anomalies that may hamper assessment and diagnosis. Therefore, doctors can use this model to detect the early onset of diseases.
4. There is a lack of good standard testing, diminishing the pool of dementia clinics. Though regulatory organizations have allowed moving some biomarkers from research labs to clinics, accessibility is still constrained.
5. Adding to the problem of prolonged waiting times and lack of care for these diseases due to insufficient formal training, we need a solution that can clear a lot of overhead.

2.1 Previous Works

1. Previous works have utilized neuroimaging data to distinguish between normal cognition, MCI, and dementia.
2. Some efforts have been made to apply contrastive learning in this domain by contrasting Alzheimer's disease with other dementia types. However, just working and trying to distinguish Alzheimer's has limited application as it co-occurs with other etiologies.
3. Image data alone is insufficient to understand a patient's complete medical approach.

3 Proposed Solutions

3.1 3-Way Classification Model

- Normal Cognition
- Mild Cognition
- Dementia

Performance is measured with AUROC and PR methods to determine the accuracy of the predictions.

3.2 Dataset

The study used data from 9 different cohorts, totaling 51,269 participants:

1. National Alzheimer's Coordinating Center (NACC)
2. Alzheimer's Disease Neuroimaging Initiative (ADNI)
3. Frontotemporal Lobar Degeneration Neuroimaging Initiative (NIFD)
4. Parkinson's Progression Marker Initiative (PPMI)
5. Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL)
6. Open Access Series of Imaging Studies (OASIS)

7. 4 Repeat Tauopathy Neuroimaging Initiative (4RTNI)
8. Lewy Body Dementia Center for Excellence at Stanford University (LBDSU)
9. Framingham Heart Study (FHS)

The dataset included participants with normal cognition (19,849), mild cognitive impairment (9,357), and dementia (22,063).

3.3 Data Management in the Model (Features Selected and Preprocessing)

- **MRI Scans**:
  - Stored in NIFTI file format
  - Underwent skull stripping using SynthStrip
  - Linearly registered to MNI152 atlas
  - Normalized to intensity range [0,1]
  - Calibration, localizer, and 2D scans were excluded
- **Non-imaging Data**:
  - Standardized across cohorts
  - Formatted into numerical or categorical variables (Numerical encoding)
- **Model Training**:
  - Data from NACC, AIBL, PPMI, NIFD, LBDSU, OASIS, and 4RTNI were merged
  - Random feature masking was used to handle missing data
  - Label masking was used to handle missing diagnostic labels

  A portion of NACC data and the full ADNI and FHS datasets were used for testing.

3.4 CatBoost Library

The CatBoost library was utilized for baseline measurements. It is a supervised machine learning method that uses decision trees for **classification and regression**. CatBoost has two main features:

- It works with categorical data (the "Cat")
- It uses gradient boosting (the "Boost")

3.5 Combination and Grouping of Symptoms

The classification was simplified by grouping dementia with Lewy Body Dementia (LBD) and Parkinson's disease dementia under the broad category of LBD, recognizing that treatment for these conditions generally follows similar protocols and is managed by a team of movement disorder specialists.

For Vascular Dementia (VD), individuals showing signs of stroke, potential or confirmed VD, or vascular brain injuries were included. This group also covers cases with symptomatic stroke, brain infarcts affecting cognitive functions, significant changes in brain white matter, and executive dysfunction. These criteria are based on the expectation that such individuals would typically be treated by specialists in stroke and vascular disorders.

3.6 Shapley Analysis

Shapley Analysis determines which feature influences the output most, using a game theory approach to quantify each feature's contribution in a predictive model.

Normal Cognition Shapley Analysis

- High score on neurophysical exams
- High score on Montreal cognition assessment
- Better performance in memory tasks
- These are directly proportional to Normal Cognition in Shapley Analysis.

Mild Cognitive Impairment

- Similar memory-related features
- Functional impairments and T1-weighted MRI
- T1-weighted scans show brain atrophy

Dementia

- Functional impairment has the most influence
- MMSE assignment has low scores in this case

3.7 With the Perspective of CDR (Clinical Dementia Rating)

The model compares the probability of dementia to the Clinical Dementia Rating (CDR) score, denoted as $P(DE) \rightarrow CDR$. This comparison is crucial in assessing the correlation between model predictions and established clinical assessments.

Model Independence from CDR CDR scores were not used during training or as inputs for the model, ensuring:

- **Independence**: Predictions remain independent of CDR scores, preventing any biases that could influence the diagnostic output.
- **Avoidance of Circular Logic**: The model does not reinforce its own inputs, which enhances the reliability of its predictions.
- **Generalization Testing**: This approach tests the model's ability to generalize across different diagnostic settings without relying on CDR.

Correlation Analysis The relationship $P(DE) \rightarrow CDR$ shows an increase with higher CDR scores, suggesting that as the clinical assessment of dementia severity increases, so does the model's prediction of dementia likelihood.

Statistical Significance The model demonstrates statistically significant differences across the spectrum of cognitive impairment, indicating its effectiveness in distinguishing varying levels of cognitive decline.

Utility as a Diagnostic Tool Even without training on CDR scores, the model proves to be a robust tool for general diagnosis, supporting its use in clinical environments where diverse assessments are applied.

3.8 Limitations

- Difficulty in distinguishing early stages of cognitive decline in the FHS dataset
- Possibly due to the limited feature set available in the FHS

4 Evaluation and Results

- The study encompassed 10 different dementia etiologies.
- High micro-average AUROC and weighted average scores indicate good overall performance across all dementia types.
- Lower macro AUPR suggests varying performance across different diagnostics.
- AUROC for co-occurring dementia ranges from 0.63 to 0.97, defining the model's performance across various combinations of dementia types.
- An AUROC of 0.5 represents a random guess.
- Lewy Body Dementia (LBD) + Psychiatric Disorders (PSY) were the most accurately predicted combinations.
- The model demonstrates proficiency in predicting the co-occurrence of conditions.

4.1 Results

Metric Type	**AUROC**	**AUPR**	**Details**
Microaveraged	0.94	0.90	Overall performance for NC, MCI, and dementia
Macroaveraged	0.93	0.84	Average performance across categories, not accounting for class imbalance
Weighted-Average	0.94	0.87	Weighted by class sample sizes, showing overall effectiveness
Subgroup Performance	>0.88	>0.82	Consistent across different age, gender, and race subgroups
Comparison with CatBoost	Varied	Varied	Outperformed CatBoost on ADNI and FHS datasets with AUROC and AUPR improvements up to 0.21 and 0.17 respectively

Table 1. Model Performance Metrics

5 Potential Improvements

There were some limitations that the author discusses:

1. **Limited Ethnic Diversity in Testing:**
   - The study population was predominantly white.
   - Lack of substantial data samples from Asian countries limits the ability to make assumptions about these populations.
2. **Performance Variability:**
   - Model performance varies across different data cohorts (e.g., NACC, ADNI).
   - Potential bias towards Alzheimer's disease.
3. **Annotation Challenges:**
   - Uncertainty in model training annotations, particularly for conditions like Semantic Variant Primary Progressive Aphasia (SEF) and Traumatic Brain Injury (TBI), due to limited feature sets.
4. **Severity Classification:**
   - The model is unable to distinguish between mild, moderate, and severe forms of dementia.

6 My Inputs and Future Improvements

6.1 1. Contrastive Learning

Approach: Indication of the same condition as positive pairs and those of negative pairs.

Co-occurring Etiologies: This addresses the challenge of overlapping symptoms in conditions like Alzheimer's disease (AD) and vascular dementia (VD).

- **Mimicking Diagnostic Reasoning:** Reflects how clinicians contrast symptoms to arrive at a diagnosis.
- **Holistic Assessment:** Considers the full spectrum of patient symptoms rather than isolated features.

Encode the symptoms on a standard scale:

- Create positive pairs from patients with similar or same symptom severity.
- Create negative pairs from patients with different diagnoses or symptom presentations.

Models to Use:

- **Siamese Networks:** Utilize networks that process two inputs simultaneously to learn the similarity between them.
- **Transformer Models:** Apply transformers capable of handling sequential symptom data and capturing long-range dependencies.

Integration with Multimodal Inputs:

- **Combined Representations:** Merge symptom-based representations with other modalities like neuroimaging and neuropsychological assessments.
- **Cross-modal Contrast:** Contrast symptom profiles with corresponding imaging data to learn joint representations.

Challenges: Symptom data can be high-dimensional, especially when encoded from clinical notes. Dimensionality reduction techniques or regularization methods can be employed to manage this complexity.

6.2 2. In-Run Data Shapley Analysis

Traditionally, the Shapley value from cooperative game theory has been used as a principled way to attribute the overall performance (utility) of a model to individual data points.

In-Run Data Shapley: [1] A novel method designed to efficiently compute Shapley values of data points during a single training run of a machine learning model.

- Leverages the iterative nature of training algorithms—such as Stochastic Gradient Descent (SGD)—to estimate each data point's contribution to the model's performance as training progresses.
- Focuses on local utility, simplifying the computation of data contribution at each step.

Second-Order Approximation:

- Includes the Hessian matrix (second-order derivatives) to account for the curvature of the loss landscape.
- Involves more complex calculations like gradient-Hessian-gradient products, which capture interactions between data points.

Custom Tools:

- **Gradient Dot-Products:** Efficiently compute the dot product of gradients without computing per-sample gradients individually.
- **Gradient-Hessian-Gradient Products:** Use optimized methods to compute second-order terms with minimal overhead.

The individual contributions computed at each iteration are summed to obtain the overall Shapley value for each data point, reflecting the data point's total contribution to the model's performance.

Benefits:

- Data points with high Shapley values significantly improve the model.
- Data points with negative Shapley values are noisy or harmful, and may be removed.
- Even if generated content doesn't closely resemble the training data, data owners may deserve credit based on their data's contribution.

How It Helps in Dementia:

- Identifying patients' contributions to the overall model helps remove noise and errors.
- Helps clear data points with negative contributions.
- **Resource Allocation:** Focus on collecting high-impact data types or from specific patient groups.

Explaining Predictions:

- Use Shapley values to explain individual predictions by showing the influence of specific data points or features.
- Increase trust in the model's decisions among clinicians and stakeholders.

Additional Benefits:

- Improve model performance by utilizing the most valuable data.
- Reduce overfitting by removing noisy data.
- **Clinician Confidence:** Providing explanations for model predictions can increase acceptance among healthcare professionals.

6.3 3. Inclusion of Additional Dementia Types

- **Specific Subtypes:** Instead of grouping dementias under broad categories, develop models to distinguish between specific subtypes like Frontotemporal Dementia or Creutzfeldt-Jakob disease.
- **Co-morbid Conditions:** Address co-existing conditions that may affect cognitive function, such as depression or other neurological disorders.

Longitudinal Data Analysis:

- Collect data from the same individuals over a period of time to observe how the disease evolves.
- Regularly collect data on various biomarkers, cognitive test scores, and other relevant factors.
- Track changes in brain volume through regular MRI scans, monitor levels of proteins (like amyloid- beta or tau) in cerebrospinal fluid, and assess cognitive test performance over time.

Benefits: Early indicators of disease onset, potentially before clinical symptoms appear.

References

1. Wang, J.T., Mittal, P., Song, D., Jia, R.: Data Shapley in One Training Run. arXiv preprint arXiv:2406.11011 (2024)
2. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv preprint arXiv:2103.14030 (2021)