```
-- Summary and research questions answered
-- Explain how the model works in detail and the code implementation
```

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

1. **Tasks to Complete**

   a) Paper Analysis

   - Prepare a report (in PPT or Word format) addressing the following questions:
     1. **Summary:** What is the paper about?
     2. **Research Question:** What scientific question is the paper trying to address?
     3. **Proposed Solution:** What solution does the paper propose to address this question?
     4. **Potential Improvements:** How could you improve upon this work?

   b) Code Implementation

   - Reproduce the experiments or models described in the paper.
   - Share your source code via GitHub.
   - Provide a report (in PPT or Word format) covering:
     1. **Code Sharing:** A link to your GitHub repository.
     2. **Results:** The outcomes of your implementation.
     3. **Interpretation:** Your analysis and interpretation of the results.

Submission Guidelines:

- Please send your reports and GitHub link to me via email within one to two weeks.
- Ensure that all work submitted is your own and properly cites any external resources used.
- If you encounter any difficulties or have questions during the process, feel free to reach out for assistance.

Glossary

Acronym	Description

NC	Normal cognition
MCI	Mild cognitive impairment
DE	Dementia
AD	Alzheimer’s disease
LBD	Lewy body dementia, including dementia with Lewy bodies and Parkinson’s disease dementia
VD	Vascular dementia, vascular brain injury and vascular dementia, including stroke
PRD	Prion disease including Creutzfeldt-Jakob disease
FTD	Frontotemporal lobar degeneration and its variants, including primary progressive aphasia, corticobasal degeneration and progressive supranuclear palsy, and with or without amyotrophic lateral sclerosis
NPH	Normal pressure hydrocephalus
SEF	Systemic and environmental factors including infectious diseases (HIV included), metabolic, substance abuse / alcohol, medications, systemic disease and delirium
PSY	Psychiatric conditions including schizophrenia, depression, bipolar disorder, anxiety and posttraumatic stress disorder
TBI	Moderate/severe traumatic brain injury, repetitive head injury and chronic traumatic encephalopathy
ODE	Other dementia conditions, including neoplasms, Down syndrome, multiple systems atrophy, Huntington’s disease and seizures

Biomarkers -

Biomarker - Measures things like protein in blood . A biomarker may be used to see how well the body responds to a treatment for a disease or condition.

Diagnostic biomarkers confirm the presence of a disease , prognostic biomarkers predict the course of a disease

Dementia is a broad category that includes several distinct conditions, most notably

Alzheimer’s disease, vascular dementia, Lewy body dementia, and frontotemporal dementia

Some biomarkers - **Alzheimer’s Disease (AD)**

```
- **Amyloid Beta (Aβ) Peptides**: Detected in cerebrospinal fluid (CSF) and by PET imaging.
- **Tau Proteins**: Elevated total tau (t-tau) and phosphorylated tau (p-tau) in CSF.
- **PET Scans**: Visual confirmation of amyloid and tau protein accumulations.
```

- **Vascular Dementia (VD)**
    - **MRI Scans**: Detect cerebrovascular disease signs like stroke or chronic ischemic changes.
    - **Blood Vessel Biomarkers**: Elevated homocysteine levels and other vascular markers.
- **Lewy Body Dementia (LBD)**
    - **Alpha-synuclein**: Abnormal aggregates in brain cells.
    - **DaTscan**: Dopamine transporter imaging to observe brain dopamine activity.
- **Frontotemporal Dementia (FTD)**
    - **Protein TDP-43, Tau, and FUS**: Misfolded proteins found in the brain.
    - **Genetic Markers**: Mutations in genes like MAPT, GRN, and C9orf72.
- **General Neurodegenerative Markers**
    - **MRI and CT Scans**: Detect patterns of brain atrophy.
    - **Electroencephalogram (EEG)**: Shows general brain activity slowing, supporting diagnostic assessment.

1. There is a overlap of etiologies, crucial for formulating early personalised management strategies
2. There is multiple types of data to consider -
    - Medical History
    - Medication Use
    - Functional Evaluation
    - Neuroimaging
    - Datasets across have 51269 participants over 9 geographies
3. Microaveraged AUROC of 0.94 in diffrentiating
    - Normal Cognition
    - Mild Cognitive Impairment
    - Dementia
4. An AUROC of 0.94 indicates that the model has a high level of diagnostic

accuracy. The AUROC measures the ability of the model to distinguish between classes—the closer the AUROC is to 1, the better the model is at making these distinctions.

5. **Microaveraged** AUROC means that the metric is calculated globally by considering the aggregation of all individual true positives, false positives, true negatives, and false negatives across all classes.
6. This implies that the model performs well across the board, not just excelling in one category while underperforming in others. It maintains its high level of performance across all three states: normal cognition, MCI, and dementia.
7. The improvement is quantified as 26.25%, which suggests that the integration of AI into their assessment process results in a more than one-quarter increase in their ability to correctly diagnose cases compared to relying solely on their clinical judgment and traditional methods.(For a subset of 100 patients)

The AI used here is able to correlate (near to 1) or differntiate different proteinopathies

Why the need for this AI based diagnosis

13. Some patient with alzeimer may be candidate for anti amyloid Treatment
14. But if the patient has other types of dimentia like Vascular Dementia then Amyloid treatment can complicate things
15. It can create Amyloid related imaging anomalies
16. There is a lack of good standard testing diminishing pool of dimentia clinics
17. Regulatory labs have allowed to move some biomarkers test from research labs to clinics but the accesibility is still constrained
18. There is the problem of prolonged waiting times
19. Lack of appropriate care due to lack of formal training

Previous Works

Previous work utilises neuroimaging data to distinguish between - Normal Cognition - Mild Cognitive impairment - Dementia

Tried contrastive learning by contrasting Alzheimer disease with other dementia types

Just working and trying to distinguish wrt to Alzheimer can have limited application as it co occurs with other etiologies

Image data alone is insufficient to understand ones medical approach

What does this model does

Takes multimodal input with features that mimic clinical diagnosis in real life -

Demographics - Family medical history - Medication use - Neuroimaging - Neuro psychological assessments

simplified the classification 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

Vascular Dementia (VD), we included individuals showing signs of stroke, potential or confirmed VD, or vascular brain injuries. 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.

Data and all

3 Way classification Model - Normal Cognition - Mild Cognition - Dimentia

Performance using ROC(Receiver Operating Characteristic) and precision-recall Unused Data from National Alzeimer Coordination Centre (NAAC)

Some 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

Catboost was used as a baseline to compare the results

CatBoost is a supervised machine learning method that is used by the Train Using

AutoML tool and uses decision trees for **classification and regression**. CatBoost has two main features, it works with categorical data (the Cat) and it uses gradient boosting (the Boost).

Shapley Analysis

Tells which feature influences the features most in the actual output

a game theory approach used to determine the contribution of each feature in a predictive model towards the prediction for a particular instance.

Normal Cognition Shapley Analysis - - High score on neurophysical exams - High score on montreal cognition assesment - Better performance in memory tasks - all these are directly proportional to Normal cognition in shapley analysis

Mild cognitive Impairement - - Similar memory related features - Functional impairements and T1 weighted MRI - T1 weighted scans show brain atrophy

Dimentia - - Functional impairement has the most influence - MMSE assignment has low scores on this one

Missing Data Handling

Remove portions of data to mimic the missing data

Types of datasets simulated - MRIs - Unified parkinson disease - G Depression Scale - NP test

- Model told a strong correlation of pre Alzheimer disease and mild cognitive impairment
- Differentiates between AD and Non AD even at MCI stage which underlines its importance in clinical settings

POV of Clinical Dementia Rating (CDR)

The model compares the probability of dementia to the Clinical Dementia Rating (CDR) score, denoted as (P(DE) 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 the 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) 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.

Some Limitations

- difficulty in Distinguishing early stage of cognitive decline in FHS dataset
- Probably due to limited feature set available in FHS

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.

Limitations

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.

Methods and Models

1. Model Architecture: Swin UNETR
    - It's a 3D transformer-based architecture designed to process various types of MRI sequences (T1w, T2w, SWI, FLAIR).
    - The model consists of two main parts:
        a. Swin Transformer encoder: Processes 3D patches of the input image
        b. Convolutional Neural Network (CNN) decoder: Reconstructs the input from the encoded representation
2. Encoder Operation:
    - Input: A 3D volume X with dimensions H × W × D (Height, Width, Depth)
    - The encoder segments X into a sequence of 3D tokens with dimensions H'/H × W'/W × D'/D
    - These tokens are projected into a C-dimensional space (C = 48) using an embedding layer
    - Patch size: 2 × 2 × 2
    - Feature dimension: 2 × 2 × 2 × 1
3. Decoder Operation:
    - Connected to the encoder through multi-resolution skip connections
    - Forms a 'U-shaped' network structure
    - Combines encoder outputs at different resolutions
    - Performs upsampling using deconvolutions
    - Generates a reconstruction of the initial input volume
4. Pre-training:
    - The Swin UNETR encoder was pre-trained using a self-supervised approach
    - Pre-training data: 3D volumes of chest, abdomen, and head/neck regions
5. MRI Processing Pipeline:
    a. Pre-processing:
        - Resampling to standardized pixel dimensions
        - Foreground cropping
        - Spatial resizing to create 128 × 128 × 128 subvolumes
    b. Feature Extraction:
        - Subvolumes are input into the Swin UNETR model
        - The model extracts encoder outputs sized 768 × 4 × 4 × 4

c. Embedding Generation:
    - Extracted features undergo downsampling
    - Uses a learnable embedding module with four convolutional blocks
    - Results in 1D vectors of size 256 for each MRI scan
d. Integration:
    - These 256-dimensional vectors are combined with non-imaging features
    - The combined data is then input into the downstream transformer for further processing

6. Dataset:
    - The model was trained, validated, and tested on a dataset of 8,155 MRI volumes

This approach represents a state-of-the-art method for processing complex 3D medical imaging data. By using a transformer-based architecture (Swin UNETR), the model can capture intricate spatial relationships in the brain MRI scans. The combination of a powerful encoder with a CNN decoder allows for effective feature extraction and reconstruction.

The pre-processing steps ensure that all images are standardized before being fed into the model. The final output - 256-dimensional vectors for each MRI scan - provides a compact yet informative representation of the brain images, which can then be easily combined with other clinical data for comprehensive analysis in the main transformer model.

Tell how the data is processed in all these models

Can be done in future by Us

1. We can try contrastive learning - Indication of the same condition as positive pairs and those of negative pairs

   Co-occurring Etiologies: 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 in a standard scale Create postive pairs from patients with similar or same symptom severity Create Negative pairs from patients with different diagnosis or symptom presentations

Models we can 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.

We can integrate it with existing 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.

Challenge - Symptom data can be high-dimensional, especially when encoded from clinical notes. But we can use dimensionality reduction techniques or regularization methods to manage complexity.

2. Work on 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

Working with shapley values is computationally expensive as it requires us to retraining the model on all possible subsets of data

In-Run Data Shapley is a novel method designed to efficiently compute the Shapley values of data points during a single training run of a machine learning model

In-Run Data Shapley 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 In-Run Data Shapley breaks down the problem of attributing data contributions over the entire training process into smaller subproblems that focus on individual training iterations.

There is a local utility function that measures the impact of a subset of data points on the models performance at a specific iteration

By focusing on local utility , the method simplifies the coputation of data contribution at each step which can help us understand overall contribution

Uses the gradient of the loss function with respect to the model parameters to estimate how a small change in parameters affects the loss.

Second order approximation - Includes the Hessian matrix (second-order derivatives) to account for the curvature of the loss landscape. - **Computation:** 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.
- **Avoiding Redundant Computations:**
  - **Memory Efficiency:** Reduces the need to store large gradient or Hessian

matrices.
    - **Computational Speed:** Achieves significant speed-ups compared to naive implementations by minimizing additional backward passes.

The individual contributions computed at each iteration are summed to obtain the overall Shapley value for each data point. which then reflects te data points total contribution to the models performance

How does it helps

- Data points with high shapley value help in more signifacnt improvement of the model
- Data point with negative shapley value are noisy or harmful
- Copyright - Even if generated content doesn't closely resemble the training data, the data owners may deserve credit based on their data's contribution.

How it helps in dementia - We can identify patients contribution to the overal model which may help us remove noise errors Can help in clearing data points with negative contributions **Resource Allocation:** Focus on collecting high-impact data types or from specific patient groups.

to assess the impact of individual features or feature groups from various modalities (such as demographics or neuroimaging). The goal is to identify which data types are most informative for the classification task. This approach offers benefits in feature engineering and data collection strategies, allowing researchers to focus on the most impactful features and allocate resources more efficiently to gather high-quality data in crucial modalities.

- Provide insights into why the model makes certain predictions based on data contributions.
- **Method:** Use Shapley values to explain individual predictions by showing the influence of specific data points or features.
- **Outcome:** Increase trust in the model's decisions among clinicians and stakeholders.

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

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.
    - Try Longitunal data analysis - Where you collect data from the same

indivisuals over a perioid of time and see how the disease evolves . regularly collect data on various biomarkers, cognitive test scores, and other relevant factors

- This focuses on how different measurable aspects of the disease change over time. For Alzheimer’s, this might include:
    - Tracking changes in brain volume through regular MRI scans
    - Monitoring levels of certain proteins (like amyloid-β or tau) in cerebrospinal fluid
    - Assessing performance on cognitive tests over time This will help early indicator of disease onset potentially before clinical symptoms appear