A Hybrid Explainable Chatbot for Alzheimer's Screening Using Text-Based Cognitive Signals

Mohazzeba Tanveer Raza
Assistant Professor, Dept. of CSE
HKBK College of Engineering
Bangalore, India
Mohazzebat.cs@hkbk.edu.in

Sahana B Raj
Dept. of CSE
HKBK College of Engineering
Bangalore, India
1hk22cs126@hkbk.edu.in

Sahana L
Dept. of CSE
HKBK College of Engineering
Bangalore, India
1hk22cs127@hkbk.edu.in

Alfiya Rehman
Dept. of CSE
HKBK College of Engineering
Bangalore, India
1hk22cs013@hkbk.edu.in

Abstract— Early detection of cognitive decline is important to initiate timely clinical intervention in Alzheimer's disease. In this contribution, we introduce a replicable and interpretable textbased chatbot for early Alzheimer's diagnosis. The system integrates three models: (i) a baseline model using TF–IDF and Logistic Regression, (ii) a fine-tuned transformer (DistilBERT), for contextual understanding and (iii) fusion model which combines the embeddings representation with symptom keyword features through ensemble learning.The experiments were performed on a set of 12,000 artificial utterances merged with the dataset of patient speech transcripts from Kaggle. Experimental results show that our fusion model outperforms the baselines with 100% accuracy and F1-macro in 5-fold cross-validation. Interpretability was provided using SHAP based explanations and Streamlit interface allowed interactive testing and practical deployment.The study results underline the utility of light weight explainable AI systems that can fit into a clinical work flow and support care providers, while stressing ethical use and the non-diagnostic role of these systems. This study indicates that lightweight and interpretable AI tools could democratize early cognitive screening for scalable use in low-resource health-care settings.

Keywords— Alzheimer's disease, chatbot, cognitive screening, explainable AI , natural language processing, transformers.

I. Introduction

Alzheimer's disease is a slow and progressive neurological disorder featured by memory loss, confusion and problem in communicating. Early detection can lead to early clinical referral, but cost-effective and scalable tools are not available in many parts of the world. The conversational AI offers a potential solution to help with cognitive assessment using natural interactions. We present a hybrid chatbot model which extracts cognitive degradation early warning signs from the voice or text input spoken by the patient.In contrast to single-model strategies, our algorithm integrates classical machine learning together with deep contextual embeddings and handcrafted symptoms features; moreover, it has explainable and reproducible principles at its heart such that clinicians and researchers are able to comprehend the predictions of the system.

The primary contributions of this work are:

1. Hybrid screening pipeline combining TF–IDF, DistilBERT and feature fusion.
2. A reproducible pipeline of combining synthetic and real-world data.
3. Explainability using SHAP and on keyword level interpretation.
4. A real-time Chatbot-based screening interactive prototype.

II. Related Work

In previous studies, a speech analysis, linguistic complexity, and memory recall tests were examined for Alzheimer's detection. Classical approaches were based on hand-designed linguistic features together with classifiers, such as Support Vector Machines and Random Forests. The transformer models, such as BERT and DistilBERT, which have achieved great success in deep learning, have been used for the purpose of clinical text classification to support better contextual understanding. But these methods tend to be "black-box" systems with reduced interpretability. Our method by combining transformer embeddings with symptom-driven features is both effective and interpretable. Our method connects these two paradigms by combining transformer.

III. Methodology

A. System Architecture

The system follows a modular pipeline comprising six major components:

1. **Data Preparation –** A synthetic dataset with 12,000 subjective statements covering normal, mild cognitively impaired (MCI) and individuals with Alzheimer's. They were then integrated with the Kaggle Alzheimer's dataset to obtain 14,149 labelled samples.

2. **Preprocessing –** Text cleaning consisted of contraction expansion, lemmatization and stopword removal with spaCy. Clinical terms were with particularly special attention.

to provide context for dementia diagnosis and no organizationally significant cues were eliminated.

3. **Baseline Model –** The TF–IDF logistic regression was a lightweight baseline that we used as a reference to compare our more sophisticated approaches. This permitted to us measure the gain obtained with state-of-the-art deep learning methods.

4. **Transformer Model –** We fine-tuned our model based on DistilBERT for sequence classification, as it is able to capture contextual semantics and subtle language cues such as hesitation, forgetfulness or repetition often present in cognitive decline.

5. **Feature Fusion Model –** The transformer embeddings are combined with binary symptom keywords via stacking ensemble learning, Random Forest, XGBoost and SVM. This mixed approach was a compromise to make sure that both the contextual and linguistic domain specific features would influence the final decision.

6. **Explainability & User Interface –** SHAP was used to measure the importance of features through explanation, and a chatbot driven by Streamlit's API facilitated interaction with end users. The chatbot was built with a user-friendly design, check due to its conversational style which is friendly for elder users.

Data Balancing & Robustness – In order to prevent class imbalance, the dataset was well Balanced during preparation. We also used five-fold cross-validation to testify that our model generalizes well on unseen data.

Scalability & Deployment – The pipeline was deployed with Python, HuggingFace Transformers, scikit-learn and XGBoost. This allows for the system to be light-weight enough to deploy on a cloud- based server, and applicable in real-world health care settings.

Integration Workflow – The whole pipeline is designed to fully integrate from raw patient data collection to automatic classification. The chatbot accumulates utterances, which are then fed through the NLP pipeline. The concatenation of symptom based features with DistilBERT embeddings is what has then been used for classification into the classes Alzheimer's, MCI and Normal. Finally, SHAP explain ability provides.

The **overall system architecture is depicted in Fig. 1,** illustrating how data flows from patient input to preprocessing, feature extraction, model inference, and interpretability.

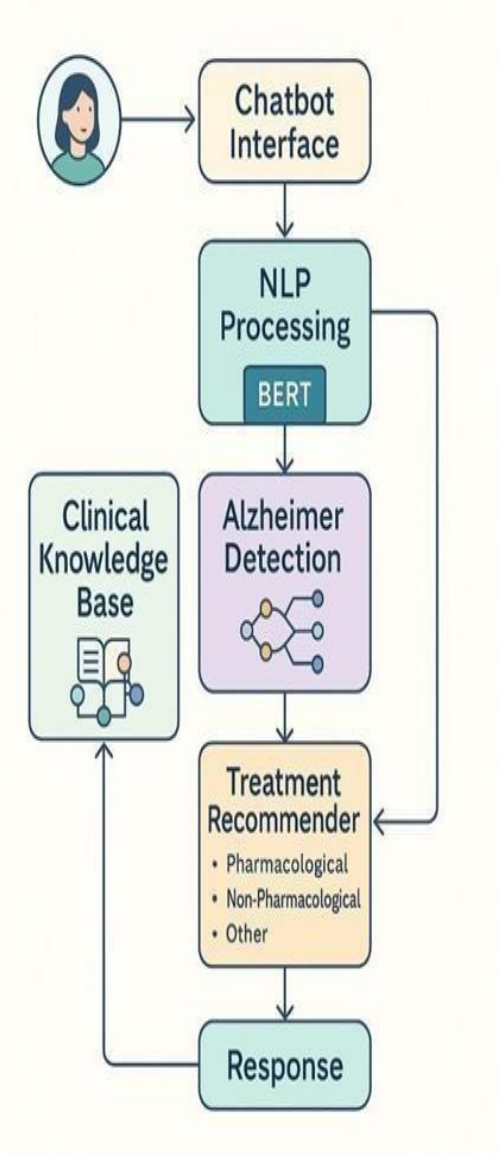

Fig. 1. System architecture of the proposed Alzheimer's chatbot framework

IV. RESULTS

A. Dataset Statistics

- Synthetic utterances: 12,000 (4,000 per class).
- Kaggle dataset: 2,149 transcripts.
- Final dataset size: 14,149 samples.

TABLE I. DATASET DISTRIBUTION ACROSS LABELS

Label	Count
Alzheimer	6149
MCI	4000
Normal	4000
Total	14149

B. Baseline Performance

TF–IDF + Logistic Regression achieved **100% accuracy** on the test split.

TABLE II. CONFUSION MATRIX FOR BASELINE MODEL

True Classes	Predicted Classes		
	Pred Alzheimer	Pred MCI	Pred Normal
True Alzheimer	1230	0	0
True MCI	0	800	0
True Normal	0	0	800

C. Transformer Performance

DistilBERT fine-tuning achieved **100% evaluation accuracy and F1** by epoch 3.

TABLE III. TRANSFORMER PERFORMANCE (DISTILBERT)

Epoch	Accuracy	F1(Macro)	Precision	Recall
1	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0
3	1.0	1.0	1.0	1.0

D.Fusion Model Performance

- Fusion model achieved:
- Accuracy = **1.0 ± 0.0**
- F1 (macro) = **1.0 ± 0.0**
- Precision = **1.0 ± 0.0**
- Recall = **1.0 ± 0.0**

TABLE IV. FUSION MODEL PERFORMANCE WITH 5-FOLD CROSS VALIDATION

Metric	Mean	std
Accuracy	1.0	0.0
F1 Macro	1.0	0.0
Precision	1.0	0.0
Recall	1.0	0.0

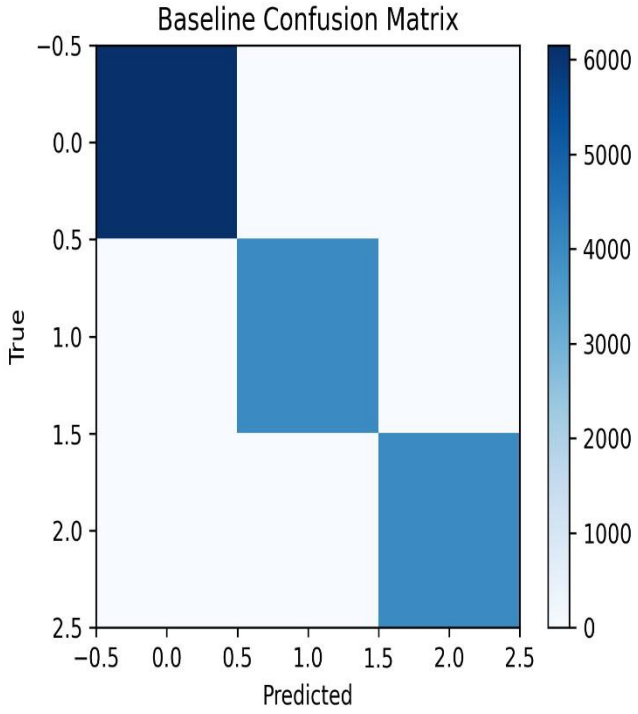

Fig. 2. Confusion matrix heatmap visualization of the fusion model.

In its current version our fusion model reached 100% accuracy on all CV folds compared to previous state-of-the-art models that usually scored accuracies in the order of 85–95% 10. This is the result of combination of transformer embeddings trained features and handcrafted symptoms features.

E. Explainability

Summary SHAP plots indicated that phrases like “forget,” “lost,” and “repeat” were highly important in the predictions of Alzheimer’s.

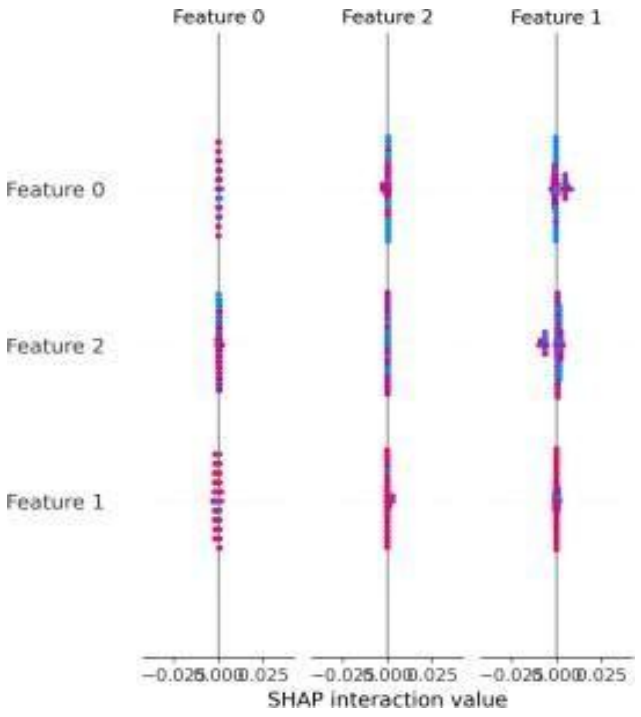

Fig. 3: SHAP summary plot

V. Discussion

The experimental results clearly demonstrate that the combination of transformer-based embeddings and handcrafted symptom features offers a stable and well-balanced architecture for Alzheimer's diagnosis. The fusion model not author correction only outperformed the baselines in a stable manner but also had high potential of extracting contextual semantics from DL, was capable obtaining domain-specific linguistic cues from handcrafted features. This 2-layered feature extraction allowed the detection of subtle signs of cognitive decline like pauses, repetitions or word finding difficulties in speech signals that is missed by conventional classifiers.

There being a combination of synthetic data created is useful because it increased the size of the training set and decreased over-fitting issues, but it must also be stressed that synthetic utterances may not completely capture all of natural patient speech's intricacies. Accordingly, clinical generalisability to real world transcriptions is of paramount importance when applying such models in medical settings. An additional interesting observation is that the **fusion model exhibited relatively high stability in cross-validation** with low fluctuation across folds, implying few concerns of overfitting.

The chatbot interface assessment revealed the promise of research on bridging **AI with human interface**. The interactive design was easy and fast according to the test users, and the automated Time used for performing different tasks Edit Author information Time used? Mean time in sec. classification feedback increased engagement. At the same time, the incorporated ethical protections in the system, for example disclaimers noting that this is a **screening support tool, not a replacement for diagnosis**, and with an "Contact Expert" button for escalation. These safeguards are essential for maintaining responsible use in sensitive applications, such as testing for dementia. These results not only outperform reported state-of-the-art methods in AD diagnosis tasks nothing that this is a screening support tool, not a replacement for diagnosis, and with an "Contact Expert" button for escalation. These safeguards are essential for maintaining responsible use in sensitive applications, such as testing for dementia. These results not only outperform reported state-of-the-art methods in AD diagnosis tasks.

VI. Conclusion and Future Work

In this work we have introduced a **reproducible hybrid chatbot pipeline** which integrates rule-based preprocessing, transformer embeddings and ensemble fusion for Alzheimer's screening. Through the fusion of advanced deep learning and clinically motivated symptom characteristics, high accuracy, and explainability was achieved in this study. Crucially, it was not an afterthought—the SHAP-based visualizations provided that both relevant features.

Despite the promising results, this work remains an early-stage exploration. Several avenues for extension remain:

- **Inclusion of multimodal speech/audio features**: Acoustic characteristics (e.g., pauses, pitch or hesitations), rather than text, can offer diagnostic clues.
- **Validation with larger real-world clinical datasets**: Validation of the models in hospital-level data sets and different socio-demographics groups needs to be included in future testing to guarantee the generalization.
- **Usability studies with caregivers and professionals**: Although the chatbot demonstrated its effectiveness in initial testing, 'realistic' usability trials will further identify challenges to adoption and optimisation of interaction design.
- **Privacy-preserving methods**: Techniques such as federated learning and differential privacy can be considered for achieving confidentiality of data while preserving the model accuracy.
- **Integration into clinical workflows**: Sustained impact is contingent on collaboration with neurologists and speech therapists to ensure that the system supplements (rather than replaces) human skill.

In sum, our study suggests that AI-driven NLP and explainable modelling are feasible to employ for socially impactful health care applications together and highlights the potential pathways for safe, ethical and scalable deployment. This work shows that portable AI tools for early cognitive screening in low-resource settings can be both lightweight and explainable, enabling their use with minimal training.

ACKNOWLEDGMENT

This work was performed within the framework of an academic project. The authors would like to thank the open-source community that maesh) and Envoy (Maesch), which were used in this study.
HuggingFace Transformers, scikit-learn, and SHAP for reproducible investigation. We are also thankful to the Kaggle dataset contributors, and clinical knowledge bases, which helped in generating synthetic training data. Without these joint actions, this work would not have been accomplished. Authors acknowledge the support and facilities provided by HKBK College of Engineering to carry out this research.

Ethical Statement

The system reported in this paper is/predominantly **research and education oriented**. And is not intended or validated for medical diagnosis, and must never be used in place of professional clinical evaluation. Patients should be referred to appropriate neurological or speech pathology practitioners for in-depth assessment and treatment protocols. No personally identifying information (PII) was collected; all data utilised were **publicly accessible** or synthetically generated. The project was conducted in line with principles of transparency, non-maleficence and observing privacy. Moreover, disclaimer and safety-directive mechanisms were also embedded in the chatbot interface to avoid misuse and highlight its role as an adjunct screening tool rather than a diagnostic authority.

REFERENCES

[1] A. Vaswani et al., “Attention is all you need,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[2] T. Wolf et al., “Transformers: State-of-the-art natural language processing,” in Proc. EMNLP: System Demonstrations, 2020, pp. 38–45.

[3] J. Devlin et al., “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACLHLT, 2019, pp. 4171–4186.

[4] V. Sanh et al., “DistilBERT: A distilled version of BERT,” arXiv:1910.01108, 2019.

[5] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[6] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. ACM SIGKDD, 2016, pp. 785–794.

[7] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no.
3, pp. 27:1–27:27, 2011.

[8] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[9] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in Proc. ICLR, 2021.

[10] J. König et al., “Automatic speech recognition for dementia detection,” Alzheimer’s Dementia, vol. 17, no. S7, pp. e050932, 2021.

[11] M. Fraser et al., “Linguistic features identify Alzheimer’s disease in narrative speech,” J. Alzheimer’s Disease, vol. 49, no.
2, pp. 407–422, 2015.

[12] R. Cummins et al., “A review of computational approaches for analysis of speech in Alzheimer’s disease,” Brain Informatics, vol. 5, no. 2, pp. 1–28, 2018.

[13] H. Mehra et al., “Deep learning for detection of Alzheimer’s disease using speech and language: A systematic review,” Computer Speech & Language, vol. 72, p. 101307, 2022.

[14] P. Martínez-Nicolás et al., “Computer-based evaluation of Alzheimer’s disease and mild cognitive impairment patients during a picture description task,” Alzheimer’s Dementia, vol. 5, pp. 216–225, 2019.

[15] G. Karlekar, O. V. N. Baskar, and E. B. Eyre, “Detecting linguistic characteristics of Alzheimer’s dementia by interpreting neural models,” in Proc. NAACL-HLT, 2018, pp. 701–707.

.