FEATURE IMPORTANCE ANALYSIS FOR MINI MENTAL STATUS SCORE PREDICTION IN ALZHEIMER'S DISEASE

Anonymous authors
Paper under double-blind review

ABSTRACT

This research article proposes developing predictive models to forecast Mini-Mental State Exam (MMSE) scores using the 54 most important features identified from the current state-of-the-art model. The study employs the SHapley Additive exPlanations (SHAP) method to explore feature importance and interpret model performance. The analysis shows that the Automated Readability Index (ARI) is the most influential feature in predicting MMSE scores. This finding suggests that ARI's capability to capture language impairment and morphosyntax is valuable in predicting cognitive decline in dementia patients. Although the analysis could not evaluate all features, this study provides a foundation for future investigations into features that may assist in predicting MMSE scores and the onset of Dementia.

1 INTRODUCTION

According to the Alzheimer's Association, the prevalence of Alzheimer's Disease (AD) in adults aged 65 and above is estimated to be 1 in 9 individuals (Harun et al., 2016). Since AD can impact an individual's speech and cognitive function (Ferrer et al., 1983; Lira et al., 2014; Rudzicz et al., 2012), researchers have focused on exploring the association between acoustic-linguistic features and AD detection. While medical literature has extensively documented AD's acoustic and linguistic symptoms, their analysis through Machine Learning / Natural Language Processing (NLP) is still limited. Recent studies have compared features for AD classification (Li et al., 2021), but few have explored feature analysis for granular MMSE prediction. To address this gap, the present study adopts the top 54 features from the current State of The Art (SOTA) (Aryal et al., 2022) approach and trains multiple machine learning models on these features. The best-performing model is then scrutinized using SHapley Additive exPlanations (SHAP) values Lundberg & Lee (2017) to identify the most critical features for predicting Mini Mental Status Exam (MMSE) scores and to improve AD detection.

2 METHODOLOGY

This study utilizes the feature set from the work of Aryal et al., which achieved the best performance on the ADReSS Challenge dataset Luz et al.. The limited dataset comprises audio and transcript data from 156 patients performing the Cookie Theft task from the Boston Diagnostic Aphasia exam. After extensive analysis, Aryal et al. selected the top 54 features from over 13,000 extracted features.

We employ Light Gradient Boosted Machine (lightGBM) (Ke et al., 2017), Kernel-based Support Vector Machine (SVM), standard distance-based k-nearest neighbors (kNN), and Aryal et al.'s linear model with SGD optimization, which have been previously utilized in healthcare (Callahan & Shah, 2017). The models are fine-tuned through Bayesian Hyperparameter Optimization. The performance is evaluated using Root Mean Square Error (RMSE) and a 95% confidence interval over 20 random seeds. To gain a more comprehensive understanding of the importance and relevance of features in predicting MMSE scores, a feature analysis will be conducted using SHAPly values on the best-performing model tested on the dataset. SHAP values will be used to rank and quantify the significance of each feature. This process will enable a thorough examination of the features that offer the most value for accurately predicting MMSE scores through regression modeling.

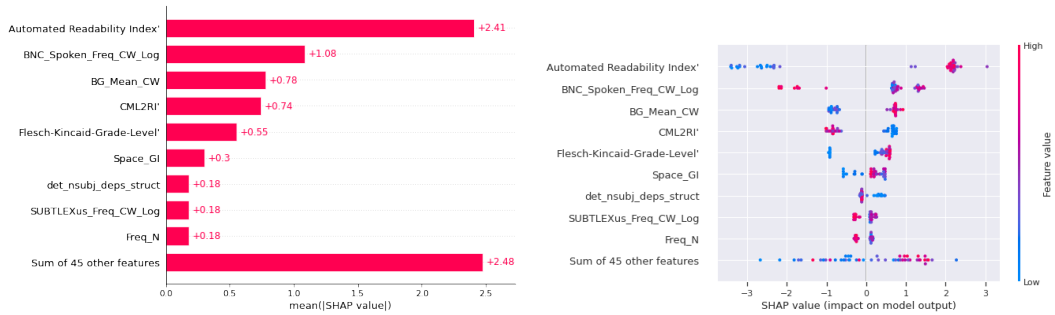
Feature Analysis will be performed on the best-performing model on the test dataset with SHAPly values. SHAP will be used to quantify features' importance and ranking. This will provide a deeper understanding of features that provide the most value for regression modeling MMSE scores.

3 RESULTS

	lightGBM	Aryal et al.	kNN	SVM
Train RMSE	**2.07 ± .24**	4.47 ± .44	4.33 ± .47	4.11 ± .36
Test RMSE	**2.91 ± .02**	4.1 ± .38	3.59 ± .05	3.84 ± .14

Table 1: Regression Results

In Table 1, all the models trained outperformed Aryal et al.'s model, suggesting the need for further feature-level analysis. The best-performing model, LightGBM will be analyzed using SHAP and TreeExplainer Lundberg et al. (2020) to determine which features contribute to its success.



(a) lightGBM absolute SHAP Values of the top 54 Features

(b) Summary plot of the top SHAP value features and their impact of predictions

Figure 1: Plot of SHAP Values

As seen in Figure 1 the top 6 features account for most of lightGBM predictions. Furthermore, Automated Readability Index(ARI) derived from text transcripts have the most significant impact. Feature analysis will be done on ARI to account for the brevity requirements, and further analysis will be considered in future work.

ARI quantifies text understandability by computing the average number of words in a sentence and characters per word (Kincaid et al., 1975). Since AD negatively impacts morphosyntax, which combines morphology and syntax (Altmann et al., 2001; Taler & Phillips, 2008), poor morphosyntax results in low-complexity words and sentences are reflected in ARI. Further analysis, utilizing Spearman's (0.788) and Pearson's correlation (0.741), shows that ARI is monotonically and linearly related to MMSE. Figure 3 demonstrates the strongest overall association between ARI and MMSE compared to the other 53 features analyzed. While trends of associations of features exist with MMSE, ARI shows the highest feature importance and associations.

4 CONCLUSION

This study analyzes the impact of the features on MMSE prediction using SHapley Additive exPlanations (SHAP) values. The results suggest that the ARI significantly predicts MMSE scores, as it can capture language impairment frequently observed with AD. However, it is essential to note that our work's limitation includes the incomplete analysis of the entire feature set and the reliance on manually annotated transcripts that may not be readily available. Moreover, ARI may not be generalizable to non-native or non-English speakers. Future research aims to expand upon this work by developing interpretable features that can effectively capture morphosyntactic impairments leading to better models to assist medical professionals in AD diagnosis and evaluation.

URM STATEMENT

The authors acknowledge that all key authors of this work meet the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Lori JP Altmann, Daniel Kempler, and Elaine S Andersen. Speech errors in alzheimer's disease. 2001.

Saurav K. Aryal, Howard Prioleau, and Legand Burge. Acoustic-Linguistic Features for Modeling Neurological Task Score in Alzheimer's, September 2022. URL http://arxiv.org/abs/2209.06085. arXiv:2209.06085 [cs, eess].

Alison Callahan and Nigam H Shah. Machine learning in healthcare. In *Key Advances in Clinical Informatics*, pp. 279–291. Elsevier, 2017.

I. Ferrer, A. Aymami, A. Rovira, and J. M. Grau Veciana. Growth of abnormal neurites in atypical Alzheimer's disease. *Acta Neuropathologica*, 59(3):167–170, September 1983. ISSN 1432-0533. doi: 10.1007/BF00703200. URL https://doi.org/10.1007/BF00703200.

Aisha Harun, Esther S Oh, Robin T Bigelow, Stephanie Studenski, and Yuri Agrawal. Vestibular impairment in dementia. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 37 (8):1137, 2016.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch, 1975.

Jinchao Li, Jianwei Yu, Zi Ye, Simon Wong, Manwai Mak, Brian Mak, Xunying Liu, and Helen Meng. A comparative study of acoustic and linguistic features classification for alzheimer's disease detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6423–6427. IEEE, 2021.

Juliana Onofre de Lira, Thaís Soares Cianciarullo Minett, Paulo Henrique Ferreira Bertolucci, and Karin Zazo Ortiz. Analysis of word number and content in discourse of patients with mild to moderate alzheimer's disease. *Dementia & neuropsychologia*, 8:260–265, 2014.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 30*, pp. 4765–4774. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):2522–5839, 2020.

Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney. Alzheimer's dementia recognition through spontaneous speech: The adress challenge. *arXiv preprint arXiv:2004.06833*, 2020.

F Rudzicz, G Hirst, P van Lieshout, G Penn, F Shein, A Namasivayam, and T Wolff. Torgo database of dysarthric articulation, 2012.

Vanessa Taler and Natalie A Phillips. Language performance in alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of clinical and experimental neuropsychology*, 30(5):501–556, 2008.

A APPENDIX

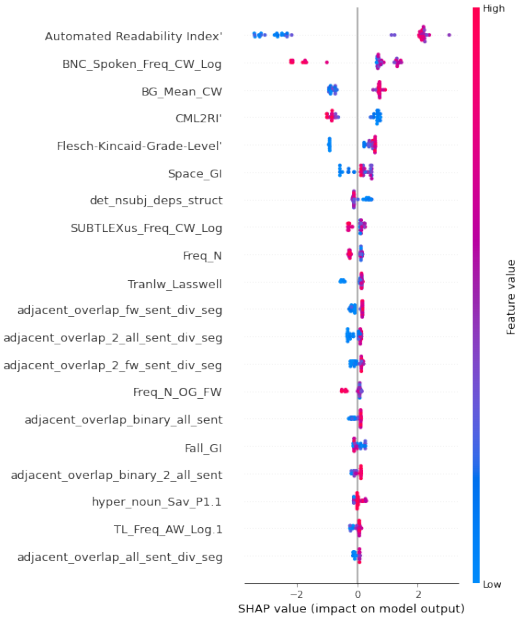


Figure 2: Expanded Impact of Features have on lightGBM prediction

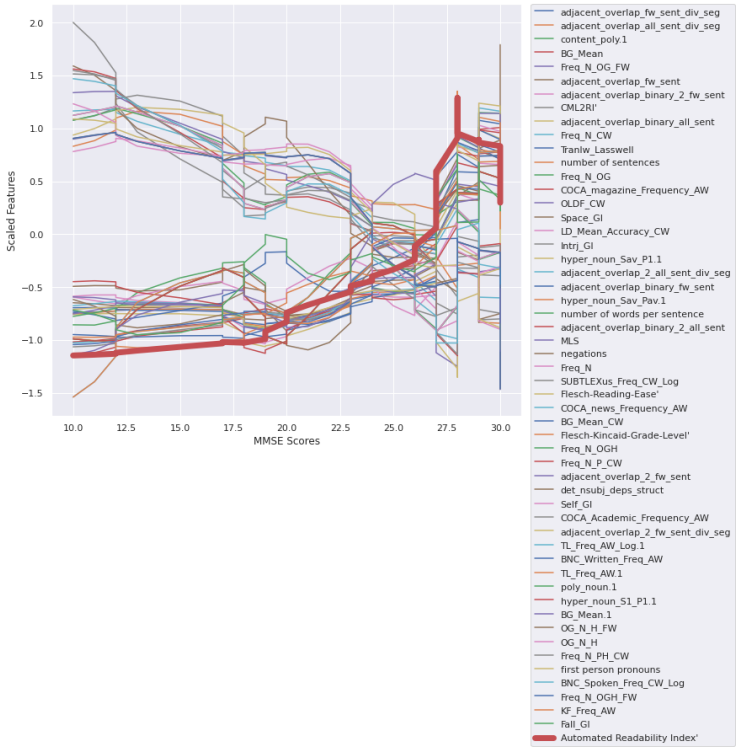


Figure 3: Plot of all 54 Features scaled for comparability as the Y and MMSE being the X. For a clearer picture of the trend of the data an Gaussian filter was applied to smooth out the data with a sigma of 2. The embolden line plot is the ARI plot.