Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Decoding Alzheimer's: Interpretable Visual and Logical Attention in Picture Description Tasks

Ning Wang[1,*], *Bingyang Wen*[2,*], *Minghui Wu*[1], *Yang Sun*[3], *Zongru Shao*[4], *Haojie Zhou*[†1], *K.P. Subbalakshmi*[2]

[1]Jiangnan University, China; [2]Stevens Institute of Technology, USA; [3]Henan Polytechnic University, China; [4]Silicon Austria Labs, Austria

ningwang@jiangnan.edu.cn, zhouhaojie@jiangnan.edu.cn

Abstract

Recent studies have started to incorporate imagery information from picture-description tasks in clinical interviews to automate Alzheimer's disease detection in the elderly. However, the high-level logical flow of visual-attention cognition mechanisms has not yet been investigated for enhanced interpretability. In this study, we systematically analyze the elements of picture-description tasks and propose a set of top-to-bottom human-interpretable features to describe the cognitive behaviors of patients, focusing on visual attention patterns, description quality, and repetition characteristics. These features achieve 85% accuracy in AD detection without specialized equipment, offering valuable insights for clinical practices and non-expert caregivers. Our results demonstrate that these high-level descriptive features, particularly those related to visual attention and the logical flow of speech, serve as effective biomarkers for AD detection.

Index Terms: Alzheimer's Disease; Interpretable AI; Neural Additive Model; Visual Attention

1. Introduction

Dementia affects over 55 million individuals globally, with Alzheimer's disease (AD) accounting for 60-70% of cases, making it the seventh leading cause of death worldwide. In 2019, the cost of caring for patients with Alzheimer's Disease and Related Dementia (ADRD) in the United States reached \$244 billion[1][2]. Early detection and accurate diagnosis of AD remain critical challenges in healthcare, underscoring the need for more effective diagnostic tools and biomarkers. Previous studies [1, 2] have highlighted the pivotal role of language in detecting cognitive decline at various stages of ADRD, demonstrating that speech analysis can facilitate early identification of cognitive impairment. Research has revealed a significant correlation between lexical attributes in linguistic production and the integrity of medial temporal lobe regions in early AD patients [1]. Recent advancements in machine learning and deep learning techniques have achieved notable success in automated AD detection through linguistic and acoustic biomarkers [3, 4, 5, 6, 7]. While these approaches demonstrate the feasibility of using textual data for AD detection, the underlying mechanisms driving their success remain poorly understood, potentially hindering their adoption in clinical settings. Picture description tasks, particularly the *Cookie Theft* task, have become valuable tools for assessing cognitive impairment, with studies indicating that over 70% of AD patients encounter difficulties in such tasks [8]. Previous efforts for the automatic detection of AD have primarily focused on individual language and speech features [9, 10, 11, 12]. To enhance the understanding of cognitive decline in language, it is crucial to incorporate task-specified features that can detect subtle differences in cognition between AD patients and healthy controls. More recently, research has begun exploring visual attention patterns in these tasks [13, 14]. However, these studies often employed black-box approaches that obscured the underlying cognitive mechanisms. The complexity of black-box decision-making in AI poses significant challenges in AD detection, where interpretability is essential for transparency, trustworthiness, trust and clinical utilization. Some researchers have attempted to address model interpretability by analyzing switching of the described objects and text-image relevance [15, 16]. While these approaches have provided valuable insights, there remains an opportunity to develop fine-grained and interpretable features that effectively capture visual attention patterns and their relationships to cognitive decline.

Motivated by the need for deeper mechanistic understanding of how cognitive decline manifests in language production, we systematically devise a set of interpretable cognitive indicators through picture description tasks. While we demonstrate our method using the Cookie Theft task, the underlying principles of modeling visual-attention patterns and language production capacity can be generalized to similar cognitive assessment tasks. Our work hypothesizes that if these cognitive-based features can effectively distinguish AD patients, they may explain why text-based models achieve high accuracy across various assessment contexts. Models trained on our cognitively-motivated features achieve comparable performance to those trained directly on text or linguistic features, suggesting that we have identified the underlying cognitive factors that drive AD detection. Furthermore, by applying interpretable machine learning methods to analyze these features, we reveal specific mechanisms of how these cognitive factors contribute to AD prediction. Through this analysis, we align our findings with previous research observations while also uncovering new indicators for AD detection, establishing a framework that bridges machine learning performance and clinical understanding.

The main contributions of our work are: (1) We devise a set of features that capture multiple aspects of cognitive behaviors in picture-description tasks, with a particular emphasis on underexplored visual and logical attention patterns in a top-to-bottom scheme. (2) The 28 devised features are validated using interpretable machine learning models, demonstrating their effectiveness while maintaining clinical relevance. (3) A comprehensive analysis of feature contributions reveals both previously known and novel indicators of AD, providing valuable insights

[*] These authors contributed equally.

[†] Corresponding author.

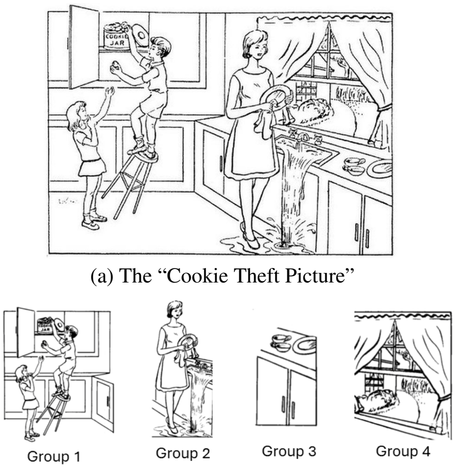
[1] https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf

[2] https://www.who.int/news-room/fact-sheets/detail/dementia

10.21437/Interspeech.2025-2596

for clinical validation and advancing our understanding of AD's impact on cognitive processes.

2. Methods



(a) The "Cookie Theft Picture"

(b) Picture segmented into four groups.

Figure 1: *(a) The "Cookie Theft Picture" from the Boston Diagnostic Aphasia Examination. The prompt employed in the study was, "Please tell me everything you see going on in this picture." (b) The picture is segmented into four element groups, labeled from 1 to 4 from left to right, where each group corresponds to a critical information content unit traditionally analyzed in clinical assessments [17].*

2.1. Dataset Description and Annotation Process

We use the *Cookie Theft* sub-dataset from the Pitt corpus [18], containing 243 control and 306 dementia samples, where participants described events shown in Figure 1 (a). Furthermore, We manually tag object mentions to develop features for visual attention and description quality. Challenges include synonyms (e.g., "son" vs. "boy") and ambiguous terms (e.g., "dish" for multiple objects). To enhance consistency, we standardize names using lexicon mapping rules (Table 1).

2.2. Proposed Feature Modeling Approach

We propose 28 hand-crafted features, categorized into five groups, to quantify participants' cognitive functions (perception, memory, and language) based on their descriptions of the "cookie theft" scene.

2.2.1. Region Coverage in Picture Description

"Region Coverage" (RC) quantifies a participant's attention to different areas of the scene. The picture is divided into four non-overlapping regions, each containing key elements of the "cookie theft" activity. The RC is calculated as the ratio of mentioned regions to all regions, $RC = M/R$, where R is the number of all regions and M is the number of regions mentioned. A higher RC suggests better cognitive and perceptual function, while a lower value may indicate deficits in attention or cognition.

2.2.2. Level of Attention

We introduce a metric to assess whether all objects are perceived similarly by both groups, categorizing objects into four frequency levels (high, mid-high, mid, and low) based on mention rates. Let O denote the set of objects mentioned in one description. Define O_{hf}, O_{mhf}, O_{mf}, and O_{lf} as the sets of objects classified into high frequency, mid-high frequency, mid-frequency, and low frequency categories, respectively. The attention to high frequency (Attention$_{hf}$), mid-high frequency (Attention$_{mhf}$), mid-frequency (Attention$_{mf}$), and low frequency categories (Attention$_{lf}$) are measured by:

$$\text{Attention}_i = \frac{|O \cap O_i|}{|O_i|}, i \in \{hf, mhf, mf, lf\} \tag{1}$$

These features reflect attention to different object categories.

High Freq		Mid-high Freq		Mid Freq		Low Freq	
Object	**Freq**	**Object**	**Freq**	**Object**	**Freq**	**Object**	**Freq**
Mother	522	Window	232	Kids	87	Table	8
Boy	514	Floor	216	Faucet	69	Bowl	4
Dish_g2	484	Cup	150	Cabinet_g3	66	Handle	3
Girl	478	Dish_g3	127	Path	52	Cabinet_g2	3
Stool	461	Curtain	120	Door	51	Corner	2
Water	434	Plant	100	Lid	44	Wall	2
Cookie	432	-	-	House	38	Mop	1
Sink	394	-	-	Cabinet_g1	32	Bird	1
Jar	389	-	-	Yard	31	Button	1
-	-	-	-	Towel	25	Board	1
-	-	-	-	Garage	17	Dish_g1	1

Table 1: *Frequency Distribution of Objects: Objects accompanied by a subscript "g" indicate their group affiliation. "Freq" indicates frequency.*

2.2.3. Major Object Description Frequency (MODF)

We introduce the "major object description frequency" (MODF) to assess attention to specific objects, focusing on high and mid-high frequency objects. For each description, MODF for a major object O_{major} is calculated as: $MODF_{O_{\text{major}}} = \text{Count}(O_{\text{major}}, D)$, where D is the description, and Count(O_{major}, D) counts the occurrences of O_{major}. MODF highlights individual focus and narrative priorities, reflecting cognitive and perceptual differences.

2.2.4. Description Repetition Score

The Description Repetition Score (DRS) quantifies repetitiveness in descriptions, potentially indicating cognitive decline in AD. It measures sentence similarity using sentence embeddings [19] and cosine similarity:

$$\mathrm{sim}(\mathrm{S_i}, \mathrm{S_j}) = \frac{\mathrm{S_i} \cdot \mathrm{S_j}}{||\mathrm{S_i}||\,||\mathrm{S_j}||} \quad (\mathrm{S_i}, \mathrm{S_j} : \text{sentence embeddings}) \tag{2}$$

The DRS is calculated as:

$$\mathrm{DRS} = \frac{2}{\mathrm{n(n-1)}} \sum_{\mathrm{i=1}}^{\mathrm{n-1}} \sum_{\mathrm{j=i+1}}^{\mathrm{n}} \mathrm{sim}(\mathrm{S_i}, \mathrm{S_j}) \cdot \log_{10}(|\mathrm{i-j}|) \tag{3}$$

where n is the number of sentences, i, j denote sentence indices, and $\log_{10}(|i - j|)$ account for the significance of repetition between non-adjacent sentences. This formula quantifies repetitiveness, giving more weight to distant sentence repetitions.

2.2.5. Description Quality Score

The Description Quality Score (DQS) evaluates how effectively a description captures an image's content using structural and semantic approaches. The structural approach analyzes sentence count (DQS_{SC}) and average word count per sentence (DQS_{WC}), assuming variations indicate description efficiency.

Higher sentence counts suggest less information delivery per sentence. The semantic approach measures the similarity between the description and the image using the CLIP model [20], comparing text and image content. For each group g, DQS_g is defined as the average sentence-image similarity:

$$\mathrm{DQS_g} = \frac{1}{|\mathrm{S_g}|} \sum_{\mathrm{i \in S_g}} \mathrm{sim(I_g, S_i)} \tag{4}$$

where S_g is the set of indices of sentences describing group g, S_i is the sentence embedding with index i, and I_g is the image embedding of sub-image of group g. sim is implemented by cosine similarity. The overall DQS is the average across all groups:

$$\mathrm{DQS_{overall}} = \frac{1}{\mathrm{G}} \sum_{\mathrm{g=1}}^{\mathrm{G}} \mathrm{DQS_g} \tag{5}$$

where G is the number of groups (4 in this case).

2.3. Neural Additive Models

Neural Additive Models (NAMs) combine neural networks with generalized additive models (GAMs), integrating feature-specific neural networks whose outputs are summed for final predictions [21]. NAMs offer explainability through shape functions, showing each feature's impact on the prediction. Density plots overlay these functions to display data distribution and model confidence. Global explanations are provided via feature importance scores based on shape function variability. NAMs were chosen for their explainability and strong predictive performance. Their design ensures accurate explanations, avoiding post-hoc method inaccuracies [22, 23, 24], while maintaining model effectiveness, making them ideal for applications needing both predictability and interpretability.

3. Results

3.1. Experiment Setup and Data Pre-processing

The data was split into 81% training, 9% validation, and 10% testing. Model robustness and generalizability were assessed through five runs with different random seeds, ensuring consistent performance evaluation across subsets. Further, we applied Z-Score[3] to our extracted feature inputs.

3.2. Baseline Models

In this section, we introduce seven baseline models including the first three models—Bouazizi et al. [25], Bouazizi et al. [16], and Zhu, Youxiang, et al. [15]—chosen for their strong performance in the same AD prediction task. We use their reported scores directly, as we are evaluating on the same dataset, rather than reproducing their methods. The remaining four models are widely-used machine learning algorithms: XGBoost [26], Random Forest [27], Support Vector Machine (SVM) [28], and Multilayer Perceptron (MLP) [29]. These models are trained on our crafted features to provide a comprehensive evaluation of feature effectiveness across different algorithmic approaches.

3.3. AD Prediction Performance

Table 2 shows the AD detection performance across all models. Both machine learning and deep learning models were trained on our features, achieving an average accuracy of 80%.

Model	Accuracy	AUC	F1	Recall	Precision
Bouazizi et al. [25]	0.815	-	0.852	0.831	0.873
Bouazizi et al. [16]	0.821	-	0.821	0.821	0.821
Zhu, Youxiang, et al. [15]	0.834	-	-	-	-
XGBoost [26]	0.789	0.861	0.811	0.832	0.807
Random Forest [27]	0.782	0.843	0.803	0.794	0.828
SVM [28]	0.778	0.844	0.811	0.826	0.799
MLP [29]	0.815	**0.878**	0.847	0.881	0.817
NAM	**0.851**	0.877	**0.876**	**0.884**	**0.872**

Table 2: *Performance of different models that are trained on our handcrafted features, with scores averaged over five trials employing varied random seeds for train-test splitting.*

Deep learning models like MLP and NAM outperform traditional models (e.g., Xgboost, SVM, Random Forest). Among all, NAM performs best, prompting further exploration of its learned insights through explanation analysis.

3.4. Feature Importance Analysis

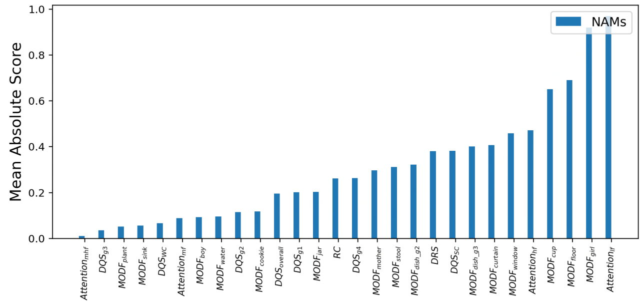


Figure 2: *Bar chart depicting the NAM's gloabl feature importance score. Features with higher importance scores contribute more to the variability of the model's output and, therefore, have a more significant impact on the predictions.*

We train five NAMs on different dataset splits and evaluate the consistency of their explanations by measuring pairwise Spearman's correlations between their feature importance scores. The analysis revealed high consistency across all model pairs, with a minimum correlation coefficient of 0.87. This agreement between models trained on different data splits suggests that NAMs are capturing genuine underlying patterns rather than spurious correlations. For subsequent analyses, to ensure that we present consistent patterns rather than potentially outlier explanations, we selected the model that exhibited the highest average correlation with the other models. Figure 2 shows the NAM's global feature importance scores. Features with higher scores have a greater impact on predictions. It shows that the NAM prioritizes high-frequency object attention (Attention$_{hf}$), object mention frequencies ($MODF$), and description repetitiveness (DRS) for AD detection, with objects like the *girl* and *cup* being most influential. Region coverage (RC) and description quality (DQS) have a lesser impact. Further analysis of feature contributions follows.

Region Coverage Figure 3 (28) shows the relationship between RC scores and AD prediction likelihood. As RC scores increase, AD prediction likelihood decreases, especially between 0.5 and 0.75. Beyond 0.75, RC's influence becomes slightly negative, suggesting that high RC leads the model to predict healthy control (HC). This trend implies that AD patients may have diminished visual perception. While RC is correlated with AD likelihood, its variability has a limited impact, indicating it is relevant but not highly influential.

Level of Attention Figure 3 (1)-(4) shows that attention to high and low-frequency objects significantly impacts AD prediction, while mid-high and mid-frequency objects have little

[3] we use the scikit-learn implementation: https://scikit-learn.org/stable/index.html

effect. The relationship between high and low-frequency objects is reversed: more high-frequency objects lead to lower AD prediction likelihood, while fewer low-frequency objects reduce the likelihood of AD prediction. This suggests that healthy controls focus more on high-frequency objects, while AD patients tend to focus on low-frequency objects.

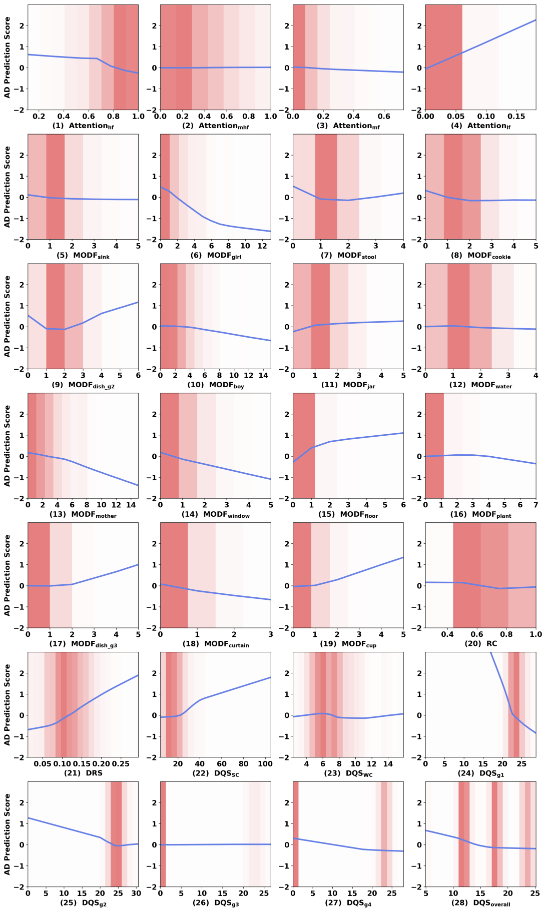


Figure 3: *NAM plots for 28 features: Each shape curve shows the relationship between input feature and model output, while background shading indicates data density to reflect prediction certainty in different regions.*

Major Object Description Frequency MODF measures attention at the object level, revealing differences in visual attention between AD patients and healthy controls (HC). Figure 3 (13)-(27) shows that the MODF of 12 out of 15 high and mid-high frequency objects influences AD prediction. Objects like the cup, dish, floor, and jar are positively correlated with AD prediction, while the curtain, girl, boy, and others are negatively correlated. The MODF feature reflects visual engagement, with high scores indicating detailed descriptions. AD-predictive objects are smaller and peripheral, while HC-predictive ones are larger and central, highlighting attention differences between AD and HC groups.

Description Repetition Score Figure 3 (12) shows a positive relationship between DRS and AD prediction likelihood, indicating that higher sentence repetitiveness is associated with increased AD prediction. AD individuals tend to repeat descriptions more than healthy controls (HC), suggesting that repetitive language patterns may signal cognitive decline.

Description Quality Score Figure 3 (5)-(12) shows that DQS_{SC} has a slight positive relationship with AD prediction, suggesting AD patients tend to use more sentences, possibly due to repetition or low language efficiency. However, sentence count has limited significance in AD detection. Similarly, DQS_{WC} does not significantly affect AD prediction. For semantic-based measurements, higher DQS values correlate with a reduced likelihood of AD prediction. Notably, $DQS_{overall}$ shows a significant drop in AD prediction between 5 and 18, suggesting HC individuals provide higher-quality descriptions. Among group-specific DQS values, DQS_{g1} has the strongest impact, with AD prediction dropping significantly between 20 and 30. This reflects that group 1 descriptions likely cover more detail, making language production deficits in AD patients more apparent.

4. Discussion

Through carefully designed interpretable features, we successfully captured multiple aspects of cognitive function, achieving an accuracy of 0.851 ± 0.027 with our best-performing NAM model. This suggest that speech transcripts from the Cookie Theft picture description task contain rich cognitive information and reveal that visual characteristics likely to be more crucial for AD detection than traditional linguistic features.

The explanations derived from the trained NAM model revealed several important features that align with previous clinical studies on visuospatial function in early Alzheimer's disease. The Description Quality Score (DQS) feature, which measures the overall quality of the picture description, was found to be an important predictor of AD. This finding can be explained by the theory of visual attention (TVA) [30, 31]. According to TVA, deficits in visual short-term memory (VSTM) may cause AD patients to provide relatively superficial descriptions of the scenes in the picture, as they have limited ”bandwidth” to allocate sufficient memory to every detail in the image. While this explanation is plausible, further clinical verification is needed to confirm this hypothesis. Another important feature identified by our model was the repetitiveness of description. This finding is consistent with previous studies [32, 33], which found that AD patients repeated words, phrases, and ideas more frequently than healthy controls in picture description tasks. However, our measure of repetitiveness differs from [10] by considering the relative position between two sentences, which better characterize the repetitiveness in real-world cases.

The simplicity of the picture description task, combined with our interpretable feature design, offers practical advantages for clinical implementation. While current object annotation requires manual effort, integration with established NER techniques could enable full automation. Beyond immediate diagnostic applications, our findings on visual attention patterns and cognitive markers open new avenues for investigating AD progression mechanisms and developing targeted assessment tools.

5. Conclusions

This study shows that interpretable, hand-crafted features from Cookie Theft picture descriptions are effective for AD detection. Our model reveals key cognitive markers and novel visual attention patterns. The task's simplicity and our approach's transparency offer a promising framework for enhancing AD diagnosis in clinical settings, contributing to a better understanding of cognitive patterns and paving the way for targeted diagnostic tools.

6. References

[1] A. Venneri, W. J. McGeown, H. M. Hietanen, C. Guerrini, A. W. Ellis, and M. F. Shanks, "The anatomical bases of semantic retrieval deficits in early alzheimer's disease," *Neuropsychologia*, vol. 46, no. 2, pp. 497–510, 2008.

[2] M. Boyé, N. Grabar, and T. M. Tran, "Contrastive conversational analysis of language production by alzheimer's and control people." in *MIE*, 2014, pp. 682–686.

[3] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Detecting signs of dementia using word vector representations." pp. 1893–1897, 2018.

[4] S. Karlekar, T. Niu, and M. Bansal, "Detecting linguistic characteristics of alzheimer's dementia by interpreting neural models," *arXiv preprint arXiv:1804.06440*, 2018.

[5] F. Di Palo and N. Parde, "Enriching neural models with targeted features for dementia detection," *arXiv preprint arXiv:1906.05483*, 2019.

[6] A. Khodabakhsh, S. Kuşxuoğlu, and C. Demiroğlu, "Natural language features for detection of alzheimer's disease in conversational speech," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2014, pp. 581–584.

[7] N. Wang and M. Chen, "Explainable cnn-attention networks (c-attention network) for automated detection of alzheimer's disease," *arXiv preprint arXiv:2006.14135*, 2020.

[8] K. E. Forbes-McKay and A. Venneri, "Detecting subtle spontaneous language decline in early alzheimer's disease with a picture description task," *Neurological sciences*, vol. 26, no. 4, pp. 243–254, 2005.

[9] S. K. Barnwal and U. S. Tiwary, "Using psycholinguistic features for the classification of comprehenders from summary speech transcripts," in *International Conference on Intelligent Human Computer Interaction*. Springer, 2017, pp. 122–136.

[10] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[11] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[12] N. Wang, Y. Cao, S. Hao, Z. Shao, and K. Subbalakshmi, "Modular multi-modal attention network for alzheimer's disease detection using patient audio and language data." in *Interspeech*, 2021, pp. 3835–3839.

[13] N. Heidarzadeh and S. Ratté, "'eye-tracking'with words for alzheimer's disease detection: Time alignment of words enunciation with image regions during image description tasks," *Journal of Alzheimer's Disease*, no. Preprint, pp. 1–14, 2023.

[14] H. Jang, T. Soroski, M. Rizzo, O. Barral, A. Harisinghani, S. Newton-Mason, S. Granby, T. M. Stutz da Cunha Vasco, C. Lewis, P. Tutt *et al.*, "Classification of alzheimer's disease leveraging multi-task machine learning analysis of speech and eye-movement data," *Frontiers in Human Neuroscience*, vol. 15, p. 716670, 2021.

[15] Y. Zhu, N. Lin, X. Liang, J. A. Batsis, R. M. Roth, and B. MacWhinney, "Evaluating picture description speech for dementia detection using image-text alignment," *arXiv preprint arXiv:2308.07933*, 2023.

[16] M. Bouazizi, C. Zheng, S. Yang, and T. Ohtsuki, "Dementia detection from speech: What if language models are not the answer?" *Information*, vol. 15, no. 1, p. 2, 2023.

[17] H. Goodglass and E. Kaplan, "The assessment of aphasia and related disorders," *(No Title)*, 1983.

[18] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis," *Archives of neurology*, vol. 51, no. 6, pp. 585–594, 1994.

[19] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[20] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.

[21] R. Agarwal, L. Melnick, N. Frosst, X. Zhang, B. Lengerich, R. Caruana, and G. E. Hinton, "Neural additive models: Interpretable machine learning with neural nets," *Advances in neural information processing systems*, vol. 34, pp. 4699–4711, 2021.

[22] Z. Carmichael and W. J. Scheirer, "Unfooling perturbation-based post hoc explainers," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, 2023, pp. 6925–6934.

[23] M. Moradi and M. Samwald, "Post-hoc explanation of black-box classifiers using confident itemsets," *Expert Systems with Applications*, vol. 165, p. 113941, 2021.

[24] S. Bordt, M. Finck, E. Raidl, and U. von Luxburg, "Post-hoc explanations fail to achieve their purpose in adversarial contexts," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 891–905.

[25] M. Bouazizi, C. Zheng, and T. Ohtsuki, "Dementia detection using language models and transfer learning," in *Proceedings of the 2022 5th International Conference on Software Engineering and Information Management*, 2022, pp. 152–157.

[26] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[27] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[28] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[29] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[30] C. Bundesen, "A theory of visual attention." *Psychological review*, vol. 97, no. 4, p. 523, 1990.

[31] P. Bublak, P. Redel, C. Sorg, A. Kurz, H. Förstl, H. J. Müller, W. X. Schneider, and K. Finke, "Staged decline of visual processing capacity in mild cognitive impairment and alzheimer's disease," *Neurobiology of aging*, vol. 32, no. 7, pp. 1219–1230, 2011.

[32] M. Nicholas, L. K. Obler, M. L. Albert, and N. Helm-Estabrooks, "Empty speech in alzheimer's disease and fluent aphasia," *Journal of Speech, Language, and Hearing Research*, vol. 28, no. 3, pp. 405–410, 1985.

[33] C. K. Tomoeda, K. A. Bayles, M. W. Trosset, T. Azuma, and A. McGeagh, "Cross-sectional analysis of alzheimer disease effects on oral discourse in a picture description task," *Alzheimer Disease & Associated Disorders*, vol. 10, no. 4, pp. 204–215, 1996.