Performing binary classification for dementia detection using machine learning models
Qinzhii(Louis) Li-lee178lee@gmail.com
USC Viterbi Department of Computer Science, SHINE 2024
Arcadia High School, Class of 2025

Introduction
Dementia is a group of cognitive disorder often experienced by older individuals char-
acterized by decreased memory, thinking, and reasoning abilities. The present study
to determine dementia involves cognitive testing and neuromaging. The cookie theft
image is commonly used to diagnose cognitive disorders. Patients are asked to describe
the image during the tests. Although humans can classify dementia based on the pa-
tient’s response to an image, their decision might be biased by age and speech patterns.

Methodology

Why does this matter?
• About 12% of women and 6.3% of men develop Alzheimer’s disease (a type of
dementia) between 65 and 80 years old [5].
• Early diagnosis allows individuals to manage medications better [1].
• Long-term costs are lower with individuals who identify dementia in the early stage
compared to the later stage [1].

Data
All the data is from the Pitt corpus dataset. In the research, we only use transcripts
related to the cookie theft scenario. We filtered the transcripts according to the following
rules: (1) remove all transcripts not spoken by the participant; (2) remove all tabs,
commas, and other punctuation marks except periods; (3) remove all numbers unrelated
in the speech. The final dataset to train the model includes 553 total data points, with
244 negative and 309 positive cases. All the models use an 80/20 train/validation split.
Our train/validation sets remain consistent across all models.

Figure 1. Cookie Theft Image: Dementia

A person who has dementia tends to describe the image or add unnecessary relationships
mistakenly.

Dementia: The boy is falling off the stool while reaching for a cookie. His girlfriend
stands nearby, beckoning or reaching for him...

Healthy: The boy has climbed up the stool to get cookies and is giving one to his
sister, who isn’t eating it yet. The stool is about to fall...

Figure 2. BERT Confusion Matrix

We proposed the following structure in Figure 3 to test image captioning model: The
feature extracted from either VLIP or BLIP will be all feed into either a neural network
or a logistic regression. The final results will be compared to BERT.

Figure 3. Experiment Model

We also reevaluate the trained Logistic Regression and Neutral Network with a black
image to validate if the model accurately learned from the image.

Results

Table 1. Model Accuracies

<table>
<thead>
<tr>
<th></th>
<th>BERT (NN)</th>
<th>BLIP (Black Image)</th>
<th>ViLT (Black Image)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>75%</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>FPR</td>
<td>15%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>TNR</td>
<td>80%</td>
<td>85%</td>
<td>95%</td>
</tr>
<tr>
<td>FNR</td>
<td>25%</td>
<td>20%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Models trained on ViLT’s feature on the other hand performs significantly better than
that of BLIP but still has worse accuracy than fine-tuned BERT. We figure out that
model trained using ViLT’s feature tend to have less false negative but also more false
positive. The only benefit is training models using ViLT’s feature takes less time and
computing power than fine-tuning BERT. Our Conclusion is that considering both the
cookie theft image and the transcript when classifying dementia might be unnecessary.

Figure 4. Training Graph of Blip (NN)

Figure 5. Training Graph of ViLT

Figure 6. Training Graph of BERT

Limitations
In our research, the only method we use to improve model performance is hyper-
tune parameters due to limited data. Models presented in this research might perform
differently on more generalized data. Furthermore, the Pitt Corpus datasets have limited
control data compared to dementia data, which might make model performance lean
towards predicting dementia.

Acknowledgment
I would like to thank my mentor Leticia Pinto-Alva, professor Jesse Thomason, and
professor Maja Mataric for guiding me on this research.

References