XX-000

Enhancing Alzheimer's Diagnosis through Spontaneous Speech Recognition: A Deep Learning Approach with Data Augmentation

Abstract

Alzheimer's disease (AD) is a growing public health issue due to its progressive nature and rising prevalence. This study explores a neural network model trained on speech data from the ADReSS2020 Challenge dataset to distinguish AD patients from healthy individuals, using log-Mel spectrogram features. To improve accuracy, five data augmentation methods, including pitch and time shifting, were used. The results highlight deep learning, combined with data augumentation, as a promising, scalable, and noninvasive approach for early AD diagnosis.

Introduction

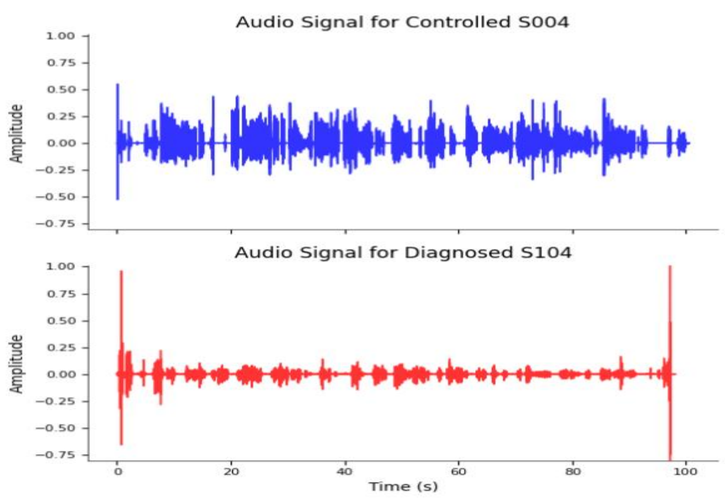
Alzheimer's disease (AD) is the most common cause of dementia, responsible for 60–70% of cases globally, and its irreversible progression presents a growing public health challenge. Early detection is critical for symptom management, especially given the lack of a cure.

Deep learning (DL) methods have shown promising results in early-stage AD detection by analyzing subtle changes in data, particularly in speech. This study builds on the baseline model by Meghanani et al., which used CNN-LSTM architecture and transfer learning to capture spatial and temporal features in speech data. To enhance performance further, we experimented with transfer learning models (ResNet50 and VGG16), which are highly effective for image processing tasks. Given that spectrograms resemble images, these models provided a refined approach to feature extraction, enabling more accurate AD classification. Additionally, we applied data augmentation techniques such as pitch shifting, time shifting, and background noise addition to improve model robustness. Our approach achieved a notable improvement in accuracy over the baseline, demonstrating the potential of speech-driven deep learning for scalable, non-invasive AD diagnosis.

Dataset Overview

The ADReSS2020 (Alzheimer's Dementia Recognition through Spontaneous Speech) Challenge dataset is designed to aid AD research. It includes audio recordings of spontaneous speech from participants describing the Cookie Theft picture, balanced by age and gender. This dataset provides high-quality data that enables reproducible research in AD classification. The below figure 1 shows the visual representation of audio signal.

Fig 1: Signals of audio recordings from ADReSS2020 Dataset



Methodology

We used the ADReSS2020 dataset, which includes balanced audio recordings of both Alzheimer's patients and healthy controls describing the "Cookie Theft" picture. This initial set provided a solid foundation, but to address data limitations, we applied several **data augmentation techniques (**See Figure2) to expand the dataset and improve the model's robustness. These augmentation methods increased the diversity of the audio samples, helping the model generalize better and preventing overfitting. See Figure 3 for the Spectrogram Generation Process

Fig. 2. Data Augmentation

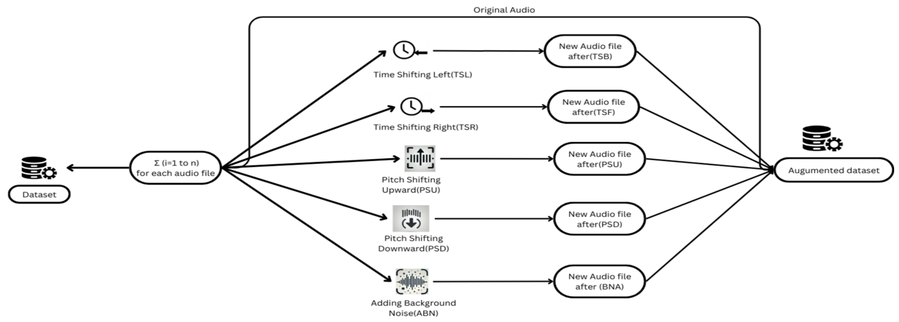
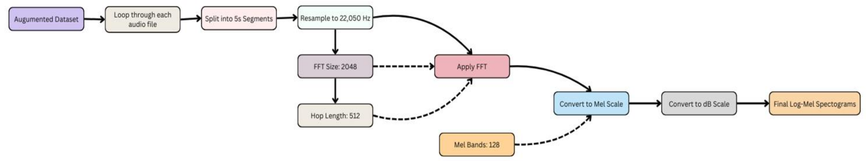


Fig. 3. Log-Mel Spectrogram Generation



Model Architecture: Expanding on the baseline CNN-LSTM and ResNet-LSTM models, we introduced an additional **VGG16-LSTM** model, applying transfer learning across all architectures. Each model was trained on the augmented dataset to maximize feature extraction and classification accuracy for Alzheimer's detection. This approach resulted in notable performance improvements, as shown in Figure 4.

CNN-LSTM: This architecture uses CNN layers to process each spectrogram frame and LSTM layers to capture sequence patterns across frames (See Table 1), making it well-suited for time-series data.

ResNet50-LSTM and **VGG16-LSTM**: For these models, we used transfer learning with ResNet50 and VGG16, which are pre-trained on Imagenet. Given that spectrograms resemble images, these models effectively handle complex feature extraction. We retained the lower layers of ResNet50 and VGG16, adding LSTM layers to process temporal information specific to AD classification.

TABLE 1

Summary of CNN-LSTM Model and Log-Mel spectrograms are the input to this CNN-LSTM model, with dimensions of 3 × 128 × 128

Layers	Input Dim	Output Dim	Operations
Conv1	128 × 128 × 3 (×10)	64 × 64 × 32 (×10)	Conv2D (32) ReLU, MaxPool
Conv2	64 × 64 × 32 (×10)	32 × 32 × 64 (×10)	Conv2D (64) ReLU, MaxPool
Conv3	32 × 32 × 64 (×10)	16 × 16 × 128 (×10)	Conv2D (128) ReLU, MaxPool
Flatten	16 × 16 × 128 (×10)	32,768 (×10)	Flatten
TimeDist	32,768 (×10)	32,768 (×10)	Apply CNN per frame
LSTM	32,768 (×10)	128	LSTM (128)
Dense	128	128	Dense, ReLU, L2 reg.
Dropout	128	128	Dropout (0.6)
Output	128	1	Dense, Sigmoid

Results

We compared our results with those from the baseline paper by Meghanani et al see figure 4, figure 5 and table2

Fig. 4. Proposed Model Accuracy for Different Algorithms

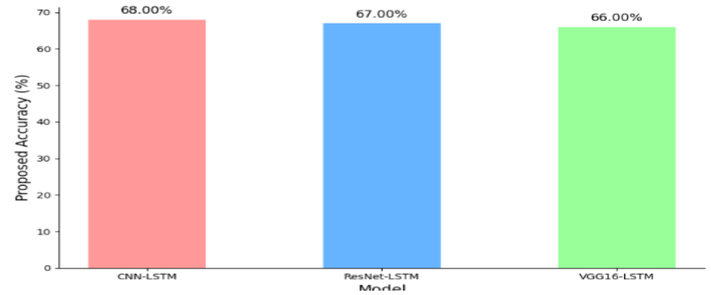


Fig. 5. Comparison of Baseline Model vs. Proposed Model Accuracy

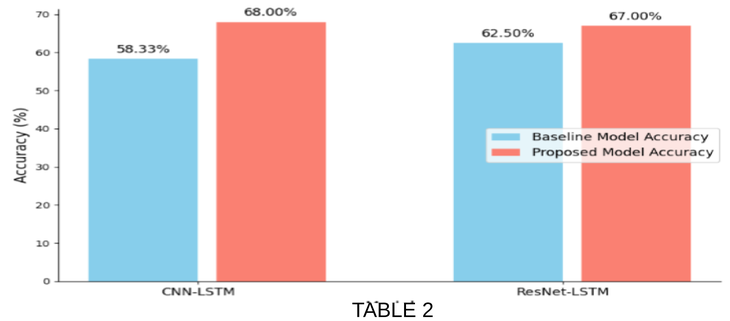


TABLE 2

Results on the test set, Comparison of Baseline and Our Results for CNN-LSTM, ResNet-LSTM with log-Mel Spectrograms, and VGG16-LSTM

Model	Features	Class	Baseline Metrics			Our Metrics		
			Precision	Recall	F1 Score	Precision	Recall	F1 Score
CNN-LSTM	log-Mel	Non-AD	0.57	0.62	0.60	0.65	0.60	0.62
		AD	0.59	0.54	0.56	0.71	0.75	0.73
ResNet-LSTM	log-Mel	Non-AD	0.62	0.62	0.62	0.65	0.55	0.59
		AD	0.62	0.62	0.62	0.69	0.77	0.73
VGG16-LSTM	log-Mel	Non-AD	-	-	-	0.56	1.00	0.72
		AD	-	-	-	1.00	0.39	0.57

Contact Information

Author - vmutala@students.kennesaw.edu – Venkata Sai Bhargav Mutala
Advisor – spouriye@kennesaw.edu – Seyedamin Pouriyeh

References

Meghanani, A., C. S., & Ramakrishnan, A. G. (2021). An exploration of log-Mel spectrogram and MFCC features for Alzheimer's dementia recognition from spontaneous speech. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 670-677). IEEE. https://doi.org/10.1109/SLT48900.2021.9383491

DementiaBank. (2020). *ADReSS 2020 challenge dataset*. TalkBank. https://dementia.talkbank.org/ADReSS-2020/

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Author(s) Venkata Sai Bhargav Mutala
Advisors(s) Seyedamin Pouriyeh