Emotion Recognition Using Sensor-Driven AI

Danielle Del Barrio, Sadie Shudde, Garvin Francois, Dr. Honghui Xu

Abstract

This project aims to develop an emotion recognition system based on audio data, leveraging sensor-driven AI techniques for downstream tasks. In real-world applications, microphone sensors capture audio signals that contain rich emotional cues such as tone, pitch, and intensity. The challenges have arisen due to environmental noise, speaker variability, and the complexity of emotional expression across different contexts. To conquer these challenges, we will build a model that can tune out extra noise and distinguish between a variety of emotions regardless of the situation. With this it can be used for real-world applications where emotions are quickly changing.

Introduction

As AI continues to evolve, it's essential that it is adaptable to human interactions across a wide range of cultures and vocal variations. Using audio data from the CREMA-D dataset, we trained our model to account for differences in environmental noise, tone, and speaker intensity before determining the emotion the speaker may be expressing.

CNNs are particularly skilled in having consistent accuracy, while LSTMs excel in handling sequential data. Thus, our group chose to combine the two to get the benefits of both architectures. Additionally, the use of a CNN-LSTM hybrid model enhances the system's memory capabilities, further improving its ability to recognize emotional cues in speech.

Research Questions

- What emotions can the AI detect?
- How efficient can the AI model be?
- What different features increase the accuracy the most?
- How well will it detect emotions overall?

Materials and Methods

Over the course of this project, we used 3 different models. We started with the standard LSTM model, but this suffered greatly from overfitting and lacked the ability to differentiate between fear and disgust. We then switched to a hybrid model of a CNN-LSTM. This helped greatly with overfitting and started to raise the accuracy of disgust and fear; however, it greatly impacted the model's ability to detect happiness. We switched from having multiple features such as MFCCs, contrast, and chroma to the Wav2Vec2 transformer by Facebook. This helped raise and even out the accuracies of the individual emotions as well as increase the total accuracy. At the end we used an MLPClassifier model to double-check that our results were compatible with other models.

Results

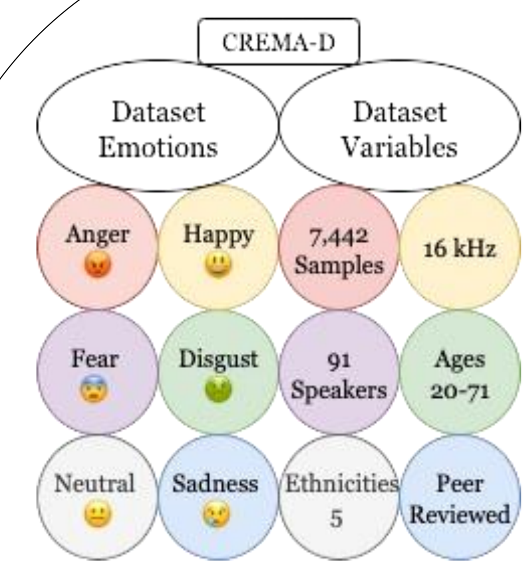

Figure 1

```
Final Training Accuracy: 76.71%
Test Accuracy: 66.11%

Per-Emotion Accuracy:
Anger    : 85.13
Disgust  : 62.56
Fear     : 63.08
Happy    : 51.28
Neutral  : 74.85
Sadness  : 61.03

F1 Score: 0.6600
```

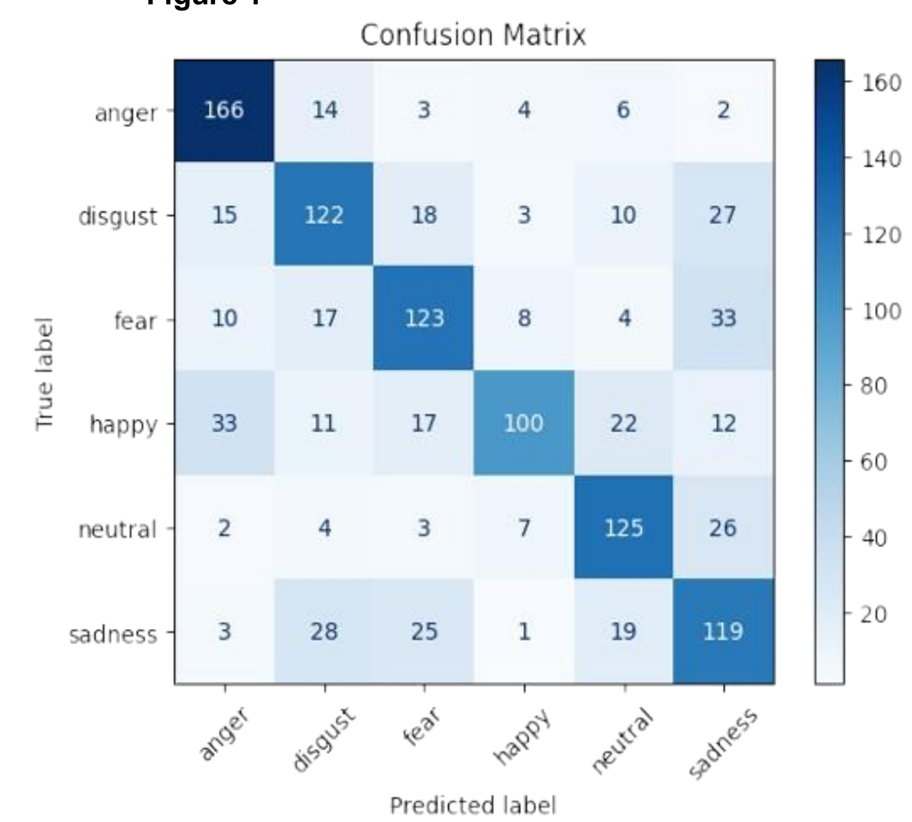

Figure 2

Model Architecture

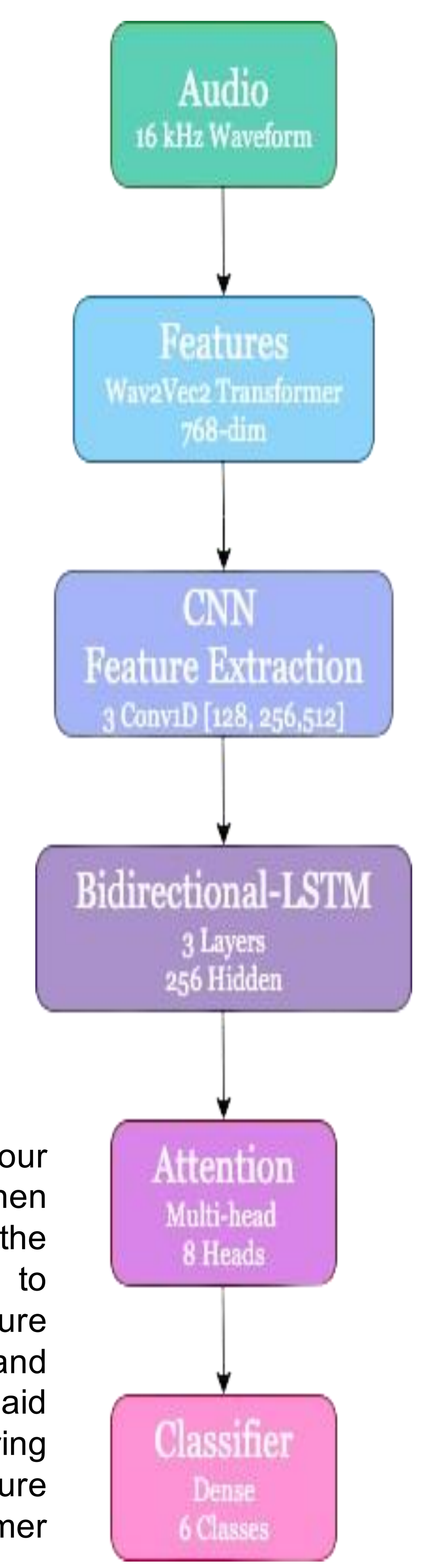

Figure 3

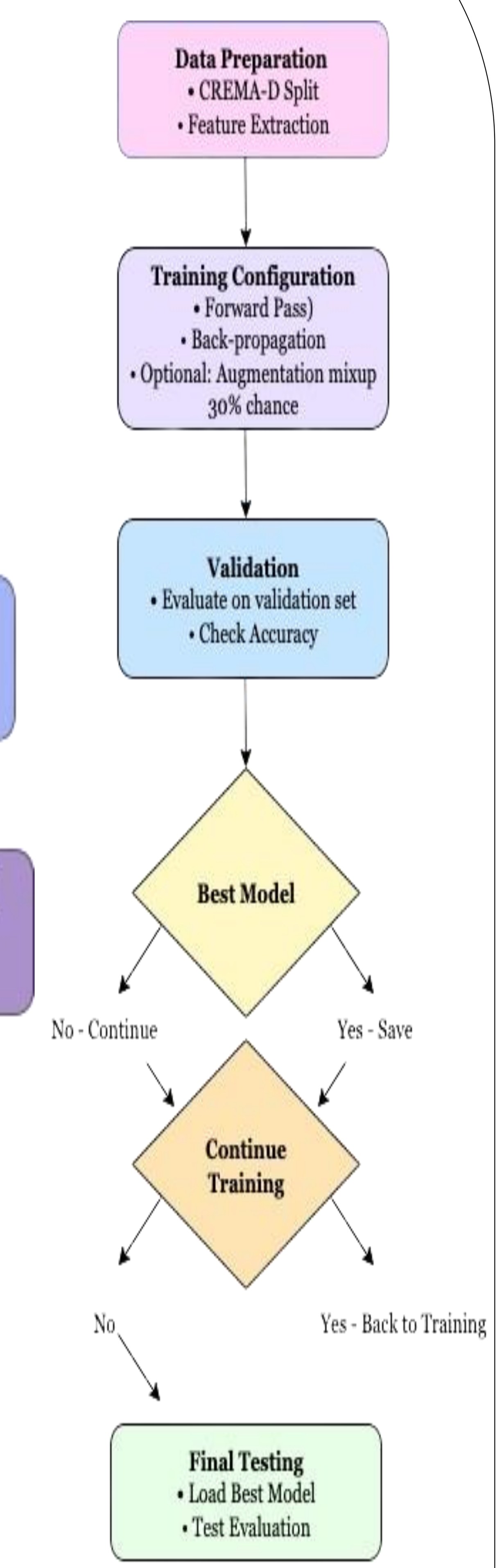

Figure 4

Highlights

Some things that were meaningful during our project include our overall decision to use a CNN-LSTM model. In the beginning, when we only used a LSTM model, we struggled with overfitting and the overall accuracy of certain emotions (disgust, fear). Choosing to adopt CNN architecture as well helped with enhancing feature extraction, improving sequential learning, better performance, and adaptability. Additionally, our accuracy improved more with the aid of the Wav2Vec2 transformer, as it was helpful in considering contextual dependencies. Since emotions and sentence structure are heavily dependent on context clues, this led to the transformer being useful for our model's emotion recognition.

Experience

This project helped our group with understanding neural networks and strengthened our skills in creating an AI model. The need for a better understanding is important, as the demand for AI skills in our respective careers(neurology, computer science), becomes more prevalent. Additionally, This project helped us in figuring out our challenges, such as overfitting and underfitting. Our group figured out that transformers were especially helpful with these issues, as the Wav2Vec2 was able to assist with accuracy and stability. Overall, this project not only expanded our knowledge of AI, but also our understanding of experimentation and applying theoretical concepts to our code.

Future Plans

Future plans for this project include the final publication of our research project, and submission to conferences. Our research paper is currently in its final draft and can be prepared for publication soon. And after publication, we wish to perfect and fine-tune our project a bit more before we submit our work to various conferences such as the IEEE Conference, an organization which represents electrical and electronic engineers. We believe that this conference would be most beneficial to us because our project is heavily related to computer engineering and machine learning, two traits that relate to the IEEE's goal of advancing technology for society.

Conclusions

Here are the results of the 3 models we implemented during the project. The LSTM proved to be unable to be generalized. The MLPClassifier was a solid model that worked well across the board. The CNN-LSTM greatly improved in all areas and was the least likely to over or underfit on the data.

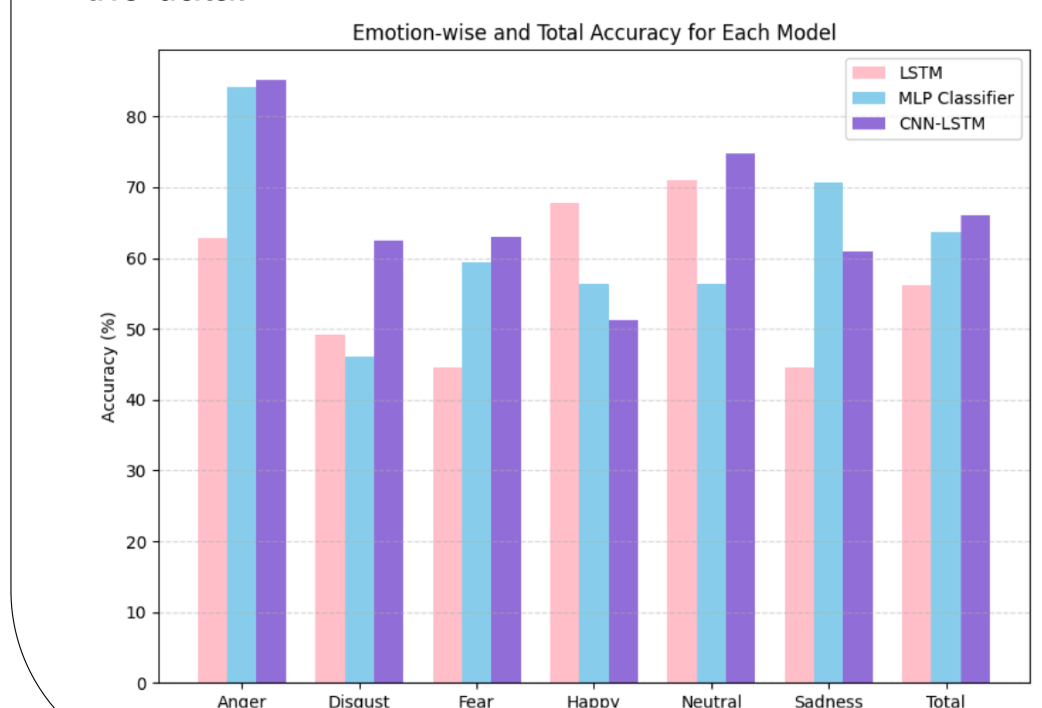

Figure 5

Acknowledgments

A special thank you to Dr. Xu for guiding us through this project and providing the advice necessary to better develop our model's accuracy. As well as Dr. Tian and Dr. He for providing us the resources needed in order to move forward with this project

Contact Information

danidbarrio@gmail.com
sadieshudde@gmail.com
garvinfrancois@gmail.com
hxu10@kennesaw.edu

References

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING