

LangSonic Final Results

Henry Lefebvre, Basel Ahsan, Abrar Mohammad Fuad

November 21, 2023

1 Introduction

LangSonic is a simple Convolutional Neural Network designed for swift and accurate spoken-language classification. Our system analyzes audio spectrograms to identify languages with precision. This final report illustrates the project's culmination, where we present our final training results and the integration of LangSonic into a web application.

2 Final Training Results

2.1 Comparative Analysis

A thorough comparison between our preliminary and final results shows significant improvements in the model's performance. The modifications made to the model since the previous deliverable are discussed here, resolving how these adjustments have optimized the outcomes.

2.2 Final Model Performance

The final training of the LangSonic model used a robust subset of Mozilla Common Voice comprising over 90,000 spectrograms for each language, achieving a final validation accuracy of 76%. This is comparable to the 79% achieved with a similar CNN by [Sergey Vilor](#). Below, we report a variety of metrics demonstrating the model's efficacy.

2.2.1 Accuracy by Language

Language	Accuracy
English	0.83
German	0.82
Italian	0.74
French	0.71
Spanish	0.70

Table 1: Accuracy of LangSonic by language

2.2.2 Confusion Matrix

The confusion matrix, depicted below, illustrates the model’s performance across different classifications.

Actual label	de	0.82	0.11	0.02	0.04	0.02
	en	0.09	0.83	0.04	0.03	0.02
	es	0.04	0.09	0.70	0.05	0.12
	fr	0.10	0.08	0.04	0.71	0.06
	it	0.05	0.06	0.09	0.06	0.74
	Predicted label					
	de	en	es	fr	it	

Figure 1: Confusion matrix for LangSonic’s classification performance

2.2.3 Validation Accuracy and Loss

We present graphs depicting the validation accuracy and loss during the training process, offering insights into the model’s learning over time. Further training offered little benefit, while suffering from the drawbacks of overfitting.

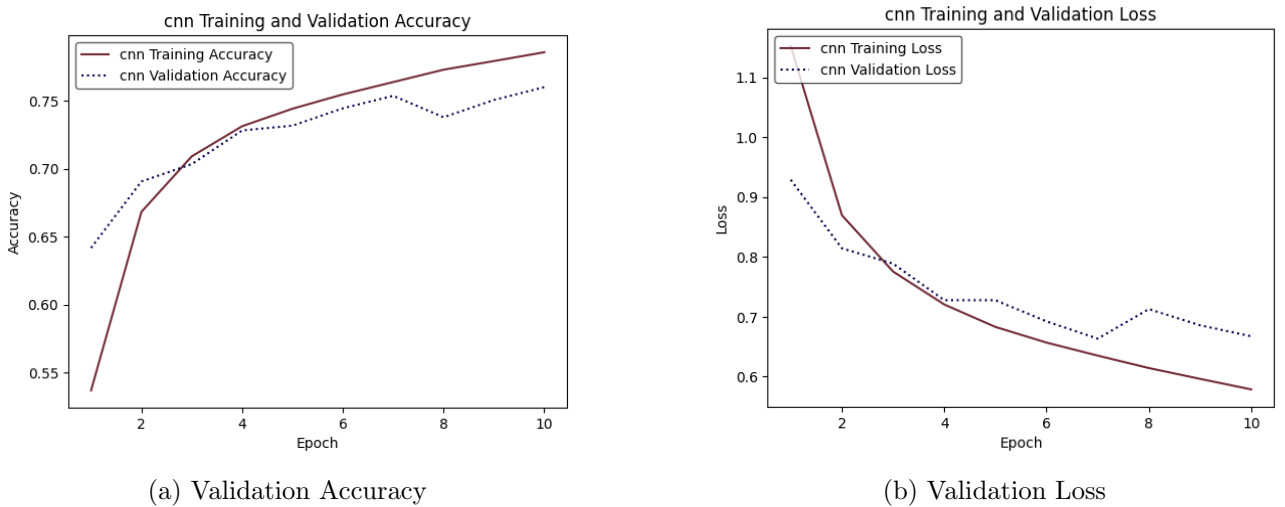


Figure 2: Validation accuracy and loss over training epochs

3 Final Demonstration Proposal

3.1 Application Integration

The final product is a user-friendly web application, accessible at [this URL](#). This section discusses the technological choices made for integrating the model into the app and justifies their appropriateness.

3.2 Tech Stack

Our team chose Flask for the web framework and Heroku for hosting. Flask's compatibility with Python allowed us to leverage existing audio processing code, ensuring consistently processed data. The familiarity with Flask reduced the learning curve and streamlined development.

3.3 Final Product

The web application serves as a landing page where users can record audio and receive instant language classification predictions.

4 Conclusion

LangSonic's final training and integration have proven the model's potential in delivering swift and reliable language classification. The web application serves as a testament to the accessibility and practicality of the model. Further enhancements and model adjustments can lead to even higher accuracy and an expanded range of languages supported.

5 References

1. [Pietz's Language Classifier](#)
2. [Spoken Language Recognition on Mozilla Common Voice](#)
3. [LangSonic Web Demo](#)