Investigating the Interpretability of Recurrent Neural Networks for Music Genre Classification

Leander Stephen
Kennesaw State University

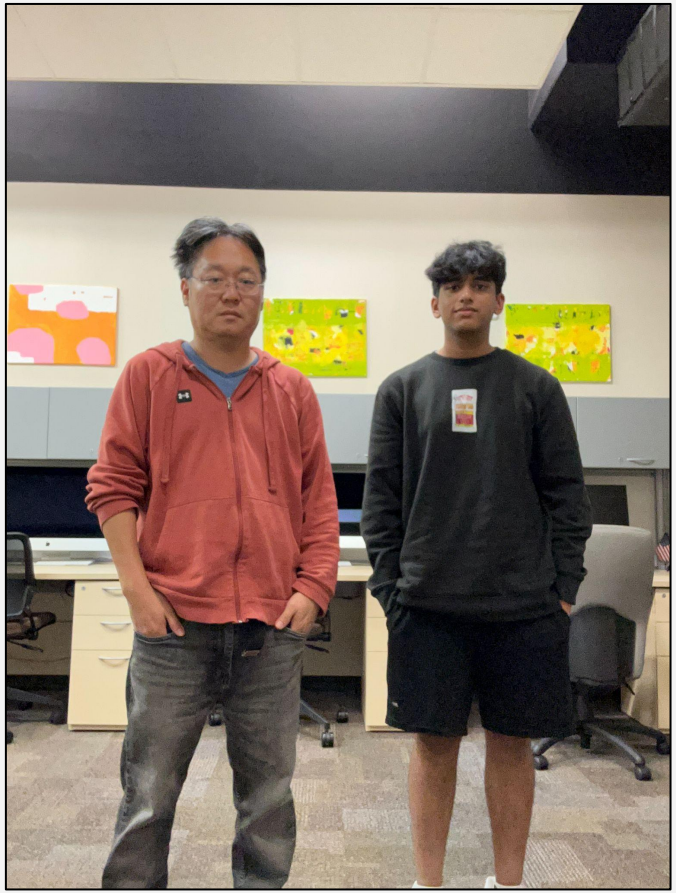
Mentor: *Arthur Choi; Computer Science Professor*

KSU Marietta Campus

- Department of Computer Science
- Prepares graduates for successful computer science careers
- Mentor: Professor Arthur Choi
  - Teaches computer science and conducts research in machine learning

Internship Duties

- Gain understanding of neural networks and proficiency coding in Python
- Review mentor's notes and other literature and apply to development process
- Biweekly status report of code to mentor



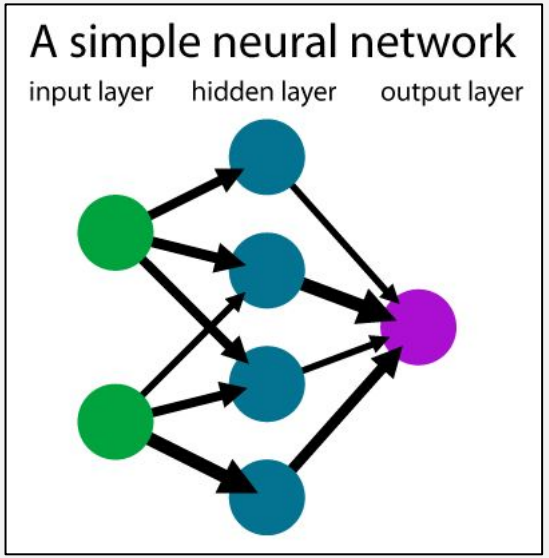
Me with my mentor at our meeting.

Neural Networks

- **Neural Network:** Class of machine learning algorithms inspired by the functions of neurons in the human brain (Bhardwaj et al., 2018)
- **Recurrent Neural Network (RNN):** Type of neural network that can process sequential data (Ghotra & Dua, 2017)

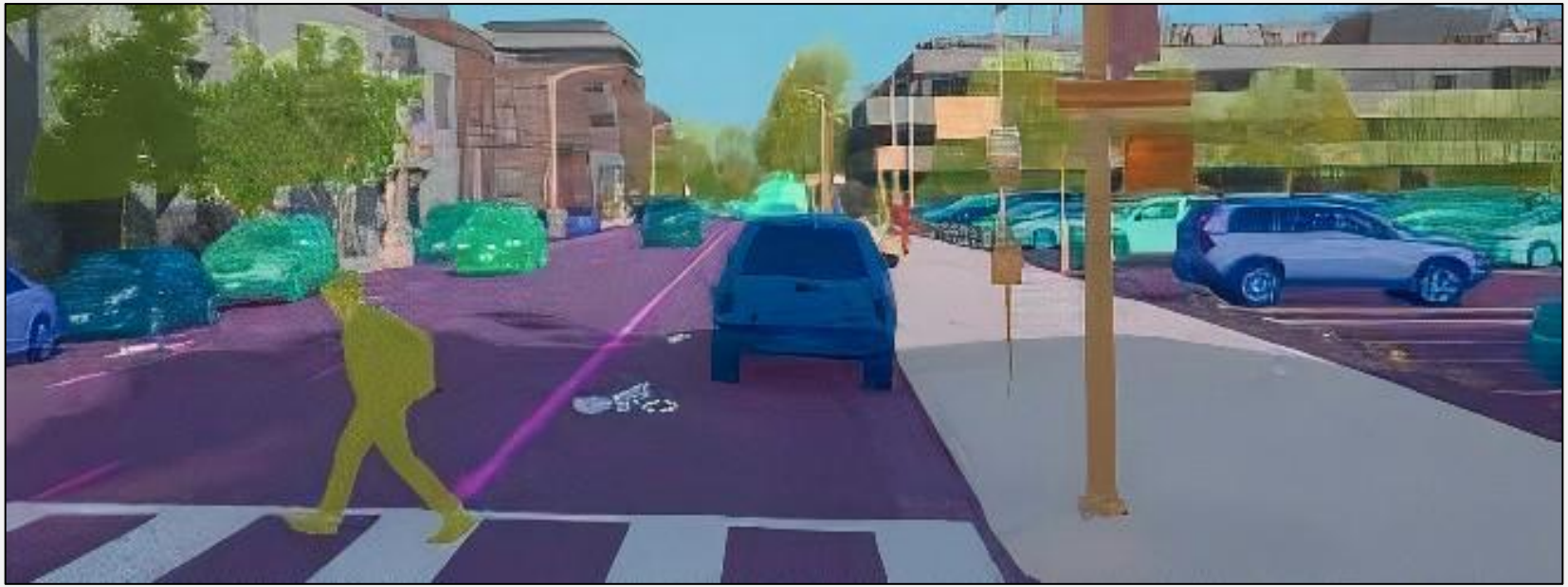
Applications:

- Music recommendation systems (He, 2022)
- Streaming services (Gao, 2022)
- Music analysis (Wang & Sohail, 2022)



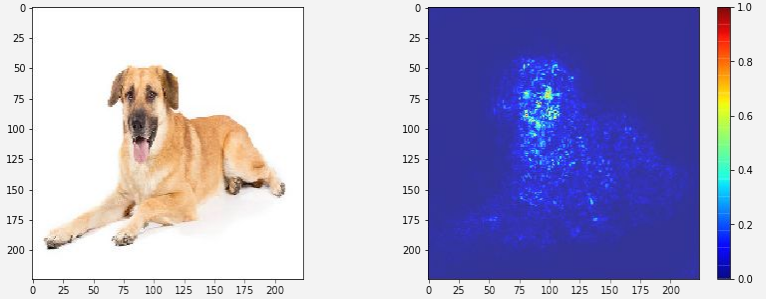
Source

Classification



Source

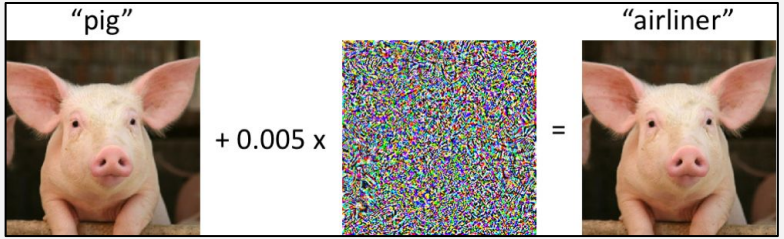
Input Analysis



Source

Saliency Mapping

- Used to highlight most important aspects of input data that affect output (Bhardwaj et al., 2018)

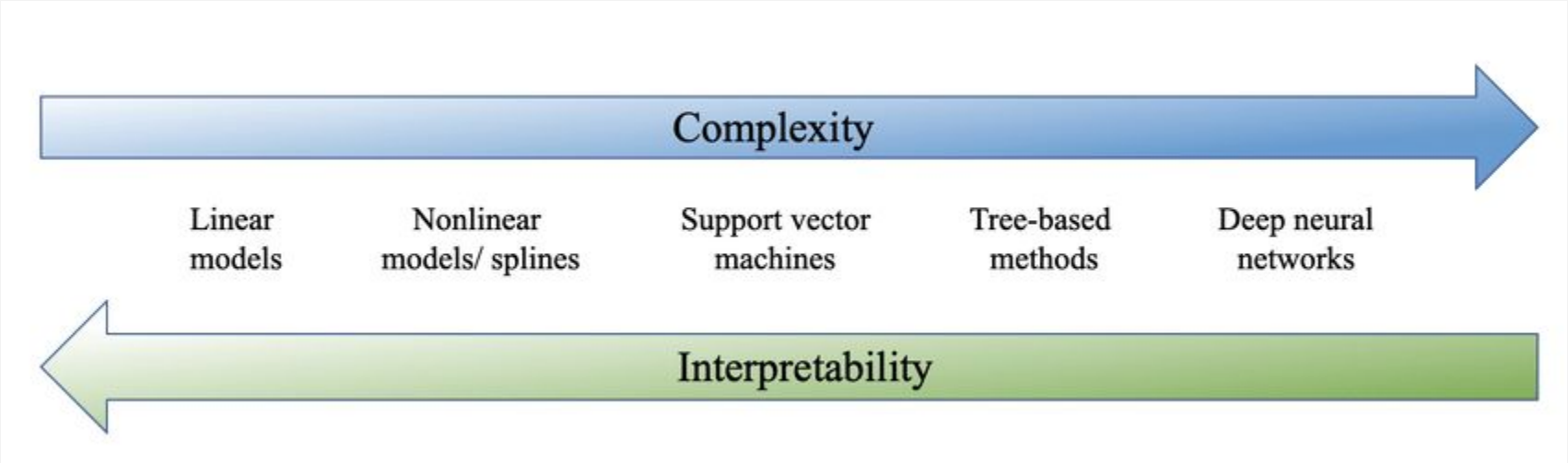


Source

Input Optimization

- Finding the optimal input for achieving a certain output (Ghotra & Dua, 2017)
- Modifying input data as little as possible to produce a classification flip

Complexity



Source

Neural Network for Music Genre Classification

Problem

Limited knowledge on interpretability of RNNs (for music genre classification)

Solution

Implement RNN and use saliency mapping and input optimization to investigate interpretability

Rationale

- Investigating how AI works is the key to creating reliable AI
- Development of more sophisticated music genre classification systems
- Contributes to void in literature
- Provides foundation for mentor's research

Key Terms

- Musical Instrument Digital Interface (MIDI) file
- Recurrent Neural Network (RNN)
- Loss Function
- Epoch
- Nodes & Layers

Source

Investigating the Interpretability of Recurrent Neural Networks for Music Genre Classification

What properties of music can an RNN detect in a MIDI file to classify music?

A musical piece's instruments, tempo, and note order (Bhardwaj et al., 2018), as well as the pitch, step, and duration of each note, are all detailed in MIDI files (Wang & Sohail, 2022). MIDI files' melodic motifs, rhythmic patterns, and song structures can all be recognized by an RNN (Bhardwaj et al., 2018).

What is a viable way to measure a neural network's effectiveness?

Accuracy is the most commonly used way to determine the accuracy of a neural network in classification tasks (Ghotra & Dua, 2017). The percentage of accurate predictions made by the network (Bhardwaj et al., 2018) will be used in this study due to its easy interpretability. The loss function can be used along with the accuracy to measure how sure the neural network is of its predictions (Ghotra & Dua, 2017).

Create and implement a recurrent neural network to perform music genre classification on a dataset of MIDI files.

A recurrent neural network that performs music genre classification is successfully implemented and has an accuracy of at least 90%.

The accuracy of the neural network will be used to evaluate each iteration of the design. The design will be coded in Python using a TensorFlow neural network implementation. The final iteration will be made when the neural network can successfully perform music genre classification with an accuracy of at least 90%.

Perform input optimization to flip MIDI file classification and generate note map of modified file. Print saliency map using gradient of MIDI file.

An input optimization note map that displays the altered MIDI file and a saliency map is created.

The design will be coded in Python. The final iteration will be made when the note map and saliency map are created.

BSP1 Analysis

What properties of music can an RNN detect in a MIDI file to classify music?

- A MIDI file contains pitch, step, and duration (Wang & Sohail, 2022)
- An RNN can detect melodic motifs, rhythmic patterns, and song structures (Ghotra & Dua, 2017)

BSP2 Analysis

What is a viable way to measure a neural network's effectiveness?

- Accuracy is the most commonly used metric (Ghotra & Dua, 2017)
- Loss function can be used alongside accuracy to measure neural networks certainty in predictions (Ghotra & Dua, 2017)

Scope of Study

Limitations/Delimitations

- Max of 100 notes per file
- Size of dataset
- Electronic/jazz
- MIDI to array conversion

Assumptions

- Saliency mapping and input optimization
- Training and compilation time
- Use of TensorFlow, Python, and Visual Studio Code

Source

Source

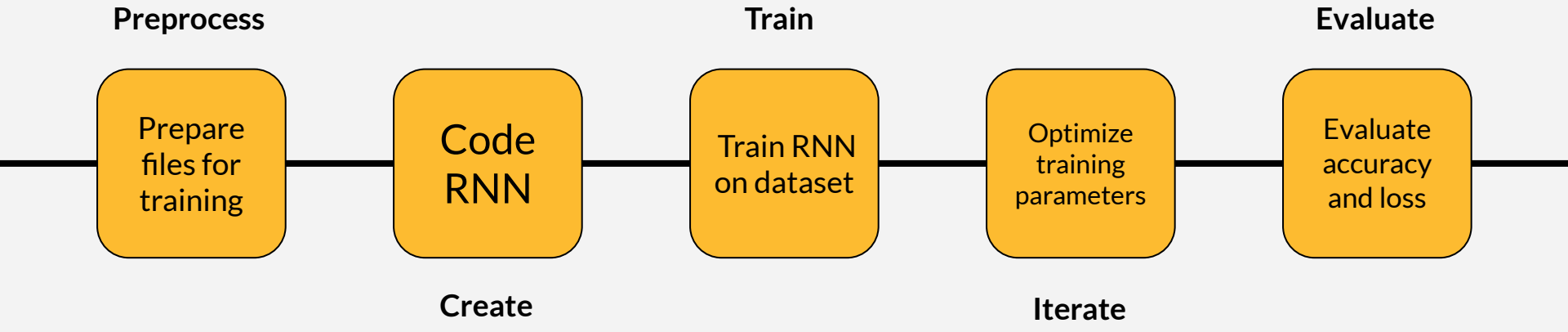
Source

Create and implement a recurrent neural network to perform music genre classification on a dataset of MIDI files.

Methodology (DSP 1)

Validity/Reliability:

- Large standardized dataset
- Accuracy is a widely used metric
- Well-established software



Iterations will continue until an accuracy of **≥ 90%** is reached.

Preprocessing

- Dataset shortened to electronic and jazz
- MIDI files converted into NumPy arrays
- Truncation to max of 100 notes

```
29  new_dataset = [song[:100] for song in dataset if len(song) > 100]
30  new_labels = [label for label, song in zip(labels, dataset) if len(song) > 100]
31  new_dataset = np.array(new_dataset)
32  new_labels = np.array(new_labels)
```

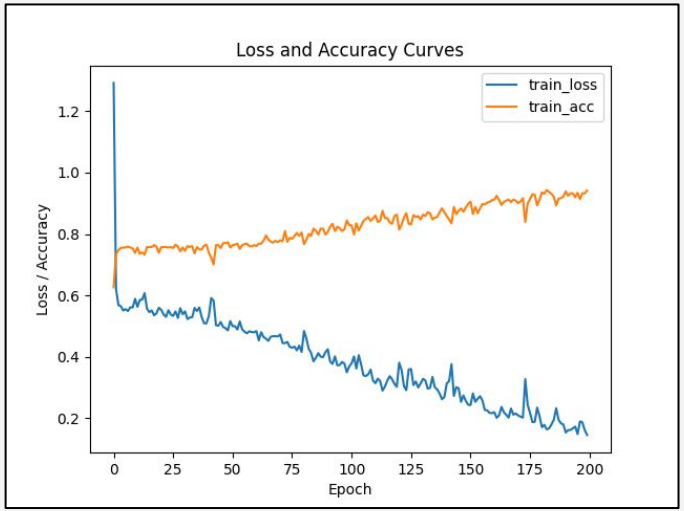
Truncation of notes.

Iterations & Results

Model	Nodes	Layers	Epochs	Accuracy	Loss
1	50	1	100	75.99%	3.7223
2	56	2	200	83.49%	0.3456
3	168	2	200	94.17%	0.1453

Iterated Parameters:

- # of Dense layers
- # of Nodes
- Epochs



Perform input optimization to flip MIDI file classification and generate note map of modified file. Print saliency map using gradient of MIDI file.

Methodology (DSP 2)

Validity/Reliability:

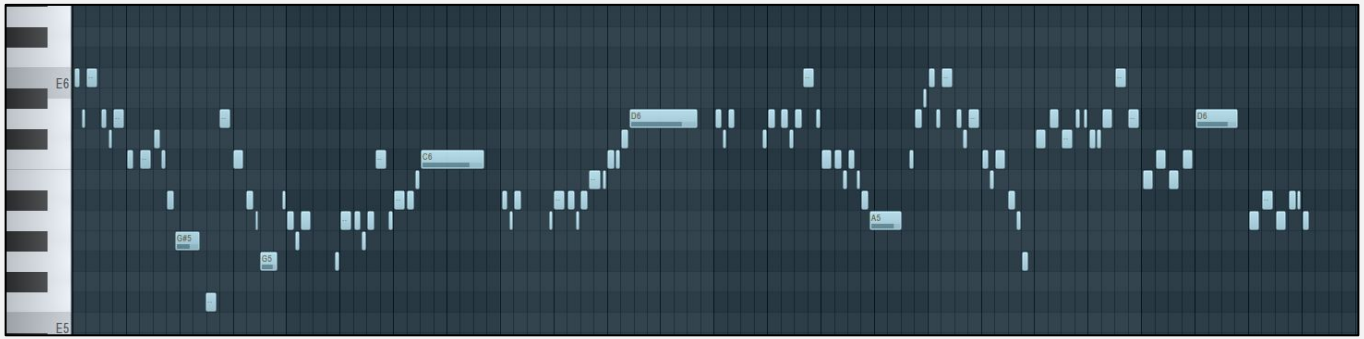
- State of the art interpretation techniques

Saliency Map

- Sample jazz MIDI file given to RNN and classified
- Gradient of model's output with respect to input

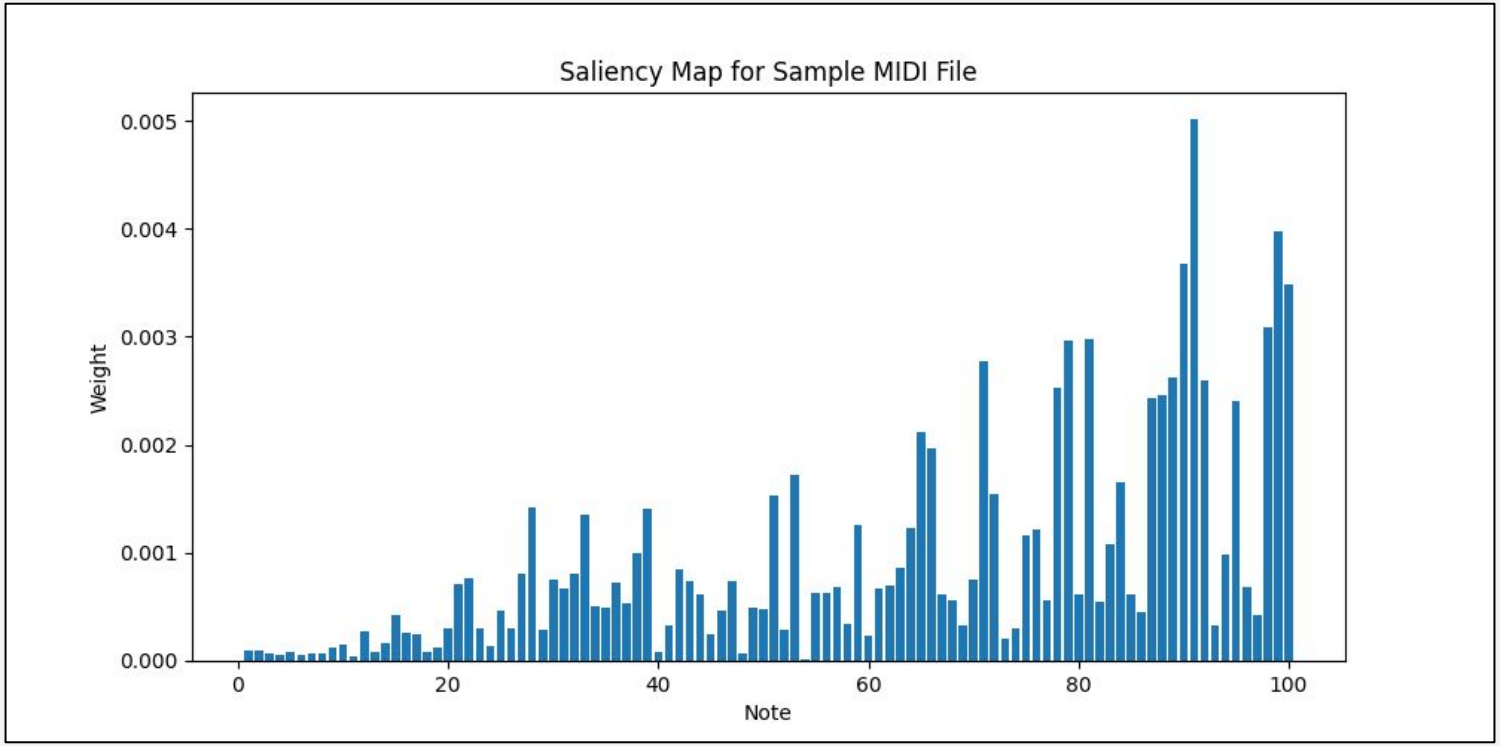
Input Optimization

- Optimized the loss function
- Created delta variable and appended to original MIDI

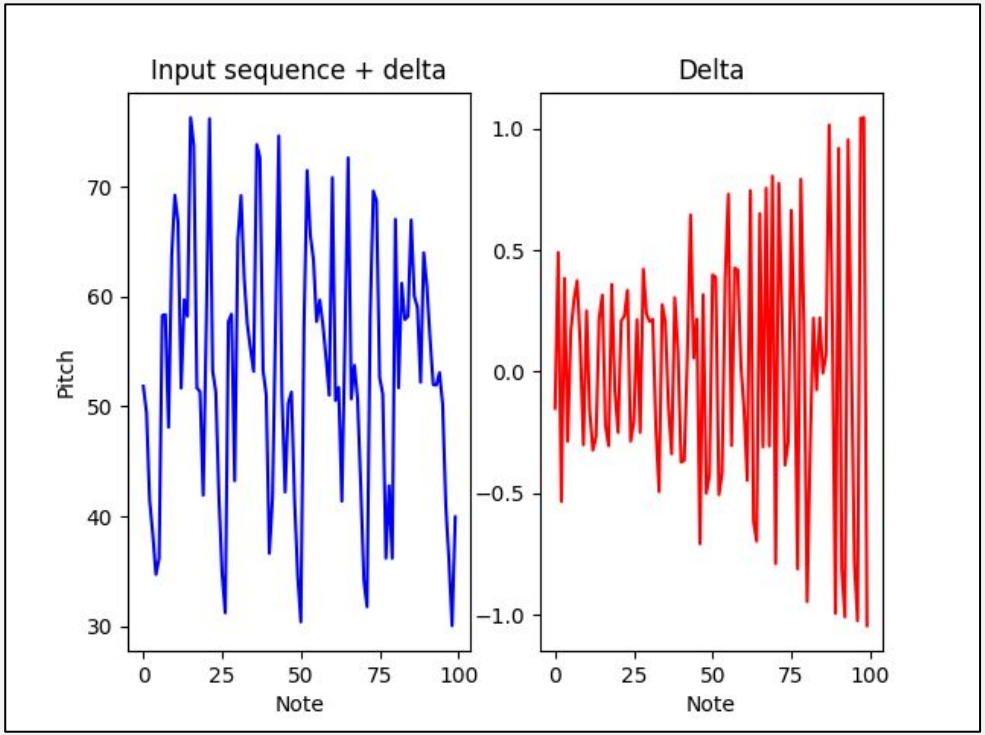


Sample jazz file

Saliency Map



Input Optimization Note Map



- Before: 1.00 (jazz)
- After: 0.45 (electronic)

Conclusions/Implications

- RNN performed music genre classification with >90% accuracy
    - Low loss
- Saliency mapping and input optimization were not interpretable
    - Complexity of model
    - Interpretation methods

Future Studies

- More genres
- More than one sample file
- Other audio files

Legacy and Impact

- Contributions to literature
- Foundation for future KSU research

Experience:

- Enhanced coding skills
- Experience in machine learning
- AI analysis skills

Acknowledgements

Thanks to:

- Professor Arthur Choi
- Dr. Cody
- Ms. Gainsford
- Classmates

Questions?

References

Bhardwaj, A., Di, W., & Wei, J. (2018). *Deep learning essentials : your hands-on guide to the fundamentals of deep learning and neural network modeling*. Packt.

Chowdhry, A. (2021, May 7). *Music genre classification using CNN*. Clairvoyant. https://www.clairvoyant.ai/blog/music-genre-classification-using-cnn

Cui, Y., & Wang, F. (2022). Research on audio recognition based on the deep neural network in music teaching. *Computational Intelligence & Neuroscience*, *2022*, 1-8. https://doi.org/10.1155/2022/7055624

Fan, M. (2022). Application of music industry based on the deep neural network. *Scientific Programming*, 1-6. https://doi.org/10.1155/2022/4068207

Gao, H. (2022). Automatic recommendation of online music tracks based on deep learning. *Mathematical Problems in Engineering*, *2022*.

Ghotra, M. S., & Dua, R. (2017). *Neural network programming with tensorflow: Neural networks and their implementation decoded with tensorflow.* Packt Publishing.

References

He, Q. (2022). A music genre classification method based on deep learning. *Mathematical Problems in Engineering*, *2022*,.

Wang, W., & Sohail, M. (2022). Research on music style classification based on deep learning. *Computational & Mathematical Methods in Medicine*, *2022*, 1-9. https://doi.org/10.1155/2022/3699885

Wang, X., Jin, C., & Zhao, W. (2018). Beijing opera synthesis based on straight algorithm and deep learning. *Advances in Multimedia*, *2018*.

Yu-Huei Cheng, Pang-Ching Chang, Duc-Man Nguyen, & Che-Nan Kuo. (2021). Automatic music genre classification based on CRNN. *Engineering Letters*, *29*(1), 1-5.