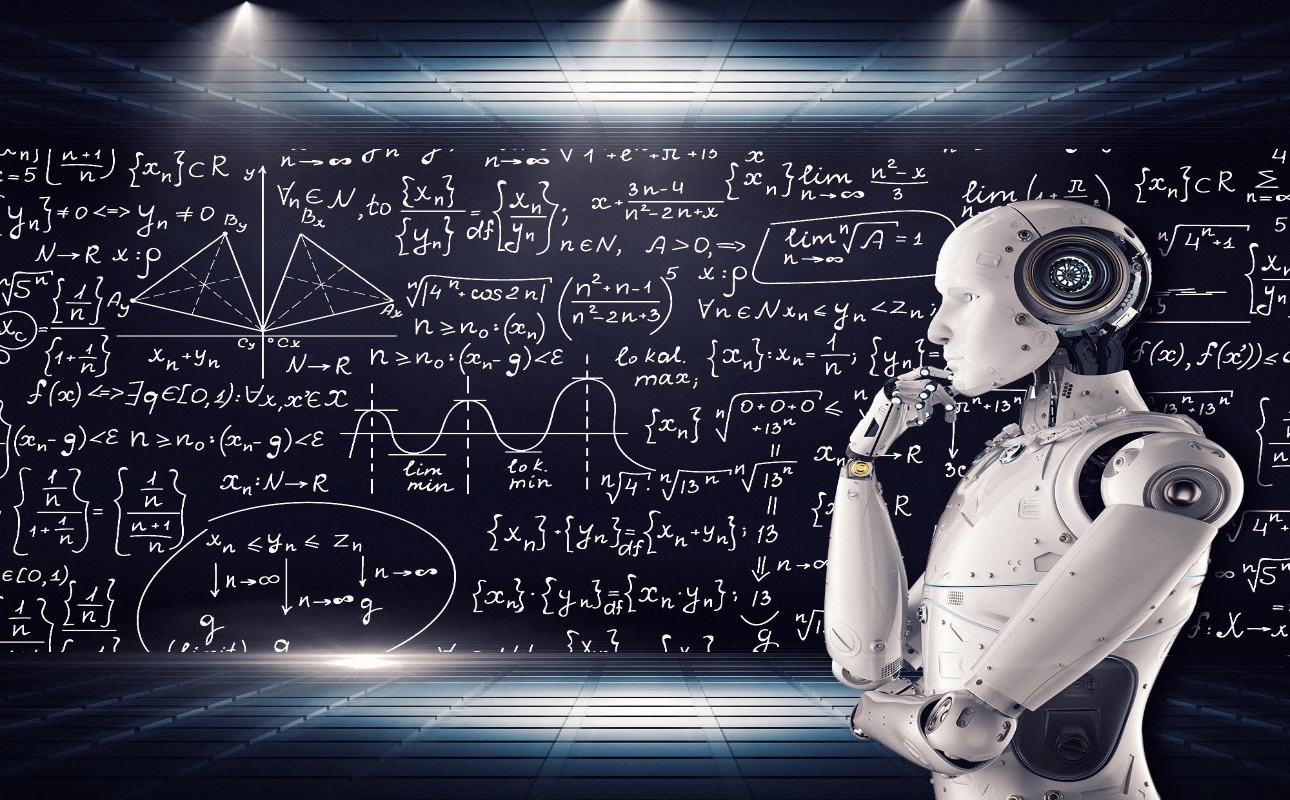
## P1#yIS1



AAI-511: [Music Genre and Composer Classification Using Deep Learning](https://ole.sandiego.edu/webapps/assignment/uploadAssignment?content_id=_3097238_1&course_id=_111306_1&group_id=&mode=view)

July 2023

Group 8: Christopher Teli, Adam Graves, Ikenna Opurum

Github Link to Model: <https://github.com/cteliStolenFocus/aai-511-team-8/tree/main>

Table of Contents

[P1#yIS1 0](#_Toc142735389)

[Abstract 3](#_Toc142735390)

[Goals/Strategies 4](#_Toc142735391)

[Data Processing 7](#_Toc142735392)

[Extracting Musical Features from MIDI Files 7](#_Toc142735393)

[Processing Composer Data into a DataFrame 7](#_Toc142735394)

[Data Fields of the generated dataset 8](#_Toc142735395)

[Spectrograms Generation from Composer Midi files for CNN model 9](#_Toc142735396)

[Phase 1: MIDI to WAV Conversion 9](#_Toc142735397)

[Phase 2: Spectrogram Image Creation 9](#_Toc142735398)

[Model Architecture 11](#_Toc142735399)

[LSTM Models 11](#_Toc142735400)

[Model 1 11](#_Toc142735401)

[Model 2 11](#_Toc142735402)

[Model 3 12](#_Toc142735403)

[Convolutional Neural Network (CNN) Model for Composer Detection Using LSTM Data 12](#_Toc142735404)

[Model Architecture 13](#_Toc142735405)

[Convolutional Neural Network (CNN) Model for Composer Detection 14](#_Toc142735406)

[Model Architecture 15](#_Toc142735407)

[Results Analysis 16](#_Toc142735408)

[Summary 16](#_Toc142735409)

[Appendix A: 17](#_Toc142735410)

[Code Implementation Details 17](#_Toc142735411)

[LOADING THE DATASET 17](#_Toc142735412)

[**1.3.1 The process:** 17](#_Toc142735413)

[**1.4. INITIAL DATA VERIFICATION AND NORMALIZATION** 18](#_Toc142735414)

[**1.5. OPTIMIZING THE DATASET** 19](#_Toc142735415)

[**1.6. DATA CLEANING** 19](#_Toc142735416)

[**2.** **BUILDING THE LSTM MODEL (Long Short-Term Memory)** 19](#_Toc142735417)

[**2.1.** **LSTM Model Version 2** 22](#_Toc142735418)

[**2.2.** **LSTM Model Version 3** 24](#_Toc142735419)

[**3.** **CNN MODEL (Convolutional Neural Network)** 25](#_Toc142735420)

# Abstract

In computational musicology there is often a demand to answer conceptually a simple questions like ” Who composed this piece?” As an example, melodic lines, rhythmic pattern, chords and chord progressions, tonality, and cadenzas are used.

This project focuses on the analysis of MIDI music data and the development of machine learning models for composer prediction based on the extracted features from MIDI files. MIDI (Musical Instrument Digital Interface) files encode musical information, making them suitable for exploring musical patterns and characteristics. The primary objective of this project is to investigate the potential of utilizing MIDI data for composer prediction and to compare the performance of different machine learning algorithms.

The composers are:

**Composer #of records**

bach 42

bartok 41

byrd 42

chopin 41

handel 41

hummel 42

mendelssohn 41

mozart 41

schumann 38

MIDI files serve as digital representations of musical compositions, storing information about pitch, timing, and other musical attributes. This project explores the application of machine learning techniques to analyze MIDI data and predict the composer of a piece based on extracted features.

With a limited amount of records we are expected to have reduced scores of prediction success.

“In the past few years, several open-source libraries such as Keras, PyTorch Lightning, Hugging Face Transformers, and Ray Train have been attempting to make DL training more accessible, notably by reducing code verbosity, thereby simplifying how neural networks are programmed. Most of those libraries have focused on developer experience and code compactness.” (AWS Machine Learning Blog, taken from: <https://aws.amazon.com/blogs/machine-learning/reduce-deep-learning-training-time-and-cost-with-mosaicml-composer-on-aws/>)

The Keras libraries. (TensorFlow) is the chosen library to use in this code.

# Goals/Strategies

The primary objective of this project is to develop a deep learning model that can predict the composer of a given musical score accurately.

The project aims to accomplish this objective by using two deep learning techniques: Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN).

The strategy includes:

1. Feature Engineering of the Dataset

* Preprocess and normalize the input data, such as image or audio files.
* Extract relevant features that are necessary for the training of both LSTM and CNN models.
* If handling MIDI files, convert them into an appropriate format, like WAV, for further processing.

1. Determination of the Final Dataset

* Select and organize the final set of features for extraction.
* Load the processed data into a suitable structure, such as a dataframe, for efficient handling.

1. Building an LSTM Model

* Research and identify the best architecture for the LSTM model.
* Select the most suitable optimizer and evaluation metrics for the model.
* Train the LSTM model with the processed data, tuning hyperparameters as needed.

1. Building a CNN Model for Character-Based Data

* Design and construct a Convolutional Neural Network that is suitable for character-based data.
* Optimize the architecture to achieve the desired performance.

1. Generating a Training Dataset from Composer MIDI Files by Converting to Spectrogram Images

* Convert MIDI files into WAV format.
* Transform the WAV files into PNG graphical files, preserving relevant information for the task.
* Prepare the PNG dataset for training with a CNN.

1. Building a CNN Model for the composer spectrogram Graphical-Based Dataset

* Design a CNN architecture that is tailored to the graphical representation of the data (PNG files).
* Train the CNN model with the PNG dataset, adjusting the structure and parameters as necessary.

1. Evaluation and Comparison

* Evaluate the performance of the LSTM and CNN models using suitable metrics.
* Compare the results and identify the strengths and weaknesses of each approach.
* Determine the best model(s) based on the project's specific requirements and goals.

The final outcome of the project will be the development and evaluation of three distinct models:

1. MODEL-1: Utilizing Long Short-Term Memory (LSTM) Networks for Composer Detection
2. MODEL-2: Implementing a Convolutional Neural Network (CNN) with Time Series Data for Composer Detection
3. MODEL-3: Developing a CNN Based on Spectrograms for Composer Detection from Musical Recordings

The final analysis will reveal the most suitable deep learning model for composer detection within this specific dataset.

# Data Processing

## Extracting Musical Features from MIDI Files

The function calculate\_features takes a MIDI file as input and computes various musical characteristics, such as pitch, note density, volume, rhythmic complexity, and tempo.

* Pitch: The average pitch of the notes playing at a given time.
* Note Density: The number of notes per second.
* Volume: The average volume of the notes playing at a given time.
* Rhythmic Complexity: The variance in the intervals between note onsets.
* Tempo: The musical tempo at each point in time, interpolated from tempo changes within the MIDI file.

These features are computed at one-second intervals over the duration of the MIDI file, resulting in time series data for each feature.

## Processing Composer Data into a DataFrame

The function process\_composer\_data iterates over a set of composer directories, each containing MIDI files for a particular composer. It leverages the calculate\_features function to compute the aforementioned features for each MIDI file, appending the results to a DataFrame along with the corresponding composer's name.

Composer: The name of the composer.

Times: The timestamps for the extracted features.

Pitch: Time series data for pitch.

Note Density: Time series data for note density.

Volume: Time series data for volume.

Rhythmic Complexity: A single value representing the rhythmic complexity.

Tempo: Time series data for tempo.

Once all the files have been processed, the DataFrame is serialized into a pickle file. This file serves as a compact and efficient way to share the preprocessed data across different parts of the project or with other team members.

The ETL process embodied in this code not only facilitates the creation of a structured dataset tailored for training deep learning models but also promotes collaboration by allowing for the seamless sharing of preprocessed musical features extracted from the raw MIDI files of different composers. It decouples the often computationally expensive feature extraction step from model building, thus enhancing the overall efficiency and flexibility of the modeling process."

## Data Fields of the generated dataset

Columns **=** ["Composer","Times",

"Pitch",

"Note\_Density",

"Volume",

"Rhythmic\_Complexity",

"Tempo" ]

## Spectrograms Generation from Composer Midi files for CNN model

The process of generating spectrograms from composer MIDI files for the Convolutional Neural Network (CNN) model is a critical step in feature extraction for our project. Spectrograms are visual representations of the frequency spectrum of audio signals, which contain valuable information about the characteristics of sound. In our context, these spectrograms serve as the primary input data for the CNN model to detect different composers. The entire procedure can be divided into two main phases: MIDI to WAV conversion and spectrogram image creation.

### Phase 1: MIDI to WAV Conversion

1. MIDI to WAV Transformation: Utilizing FluidSynth and sound font files, the MIDI files corresponding to different composers' works were converted into WAV audio format. The choice of WAV files stems from their lossless nature, preserving all the musical information without compression.
2. Automating the Conversion: To efficiently handle multiple files across various directories, a shell script was created to iterate through the directories and perform the conversion. This streamlined the process, allowing for batch processing of files.

### Phase 2: Spectrogram Image Creation

1. Loading WAV Files: Using Librosa, a popular library for analyzing and processing audio files, the WAV files were loaded into the Python environment. Librosa's load function was employed to read the audio data.
2. Short Time Fourier Transform (STFT): The Short Time Fourier Transform was applied to the audio data using Librosa's stft function. This method breaks down the audio signal into small windows and computes the Fourier Transform for each, allowing for the frequency analysis of the signal.
3. Amplitude and Image Resizing: The amplitude of the STFT was extracted and resized to a consistent 224x224 dimension using SciPy's ndimage module. This ensured uniform input size for the CNN model.
4. Spectrogram Visualization: Librosa's specshow function was used to visualize the resized amplitude as a spectrogram. The visualization parameters included both time and logarithmically-scaled frequency axes.
5. Image Saving: Finally, the spectrogram images were saved as PNG files in corresponding directories using Matplotlib's savefig function. PNG was chosen as the format for its lossless compression and wide support in image processing.

A blue and orange grid

Description automatically generated

The conversion from MIDI files to spectrogram images was a critical preprocessing step for our project. It transformed complex musical information into a visual form that could be directly fed into the CNN model. By automating the process through scripting and leveraging specialized libraries like Librosa, this phase successfully prepared the data for machine learning analysis, focusing on the unique attributes of each composer's works. The resulting spectrograms provide a rich and consistent dataset that plays a vital role in the subsequent modeling and classification tasks.

# Model Architecture

## LSTM Models

Three LSTM Models were evaluated.

### Model 1

Architecture:

* LSTM layer with 50 units, ReLU activation function.
* Dense layer with num\_classes units, softmax activation function.

Input Shape: (num\_steps, num\_features).

Optimizer: Adam.

Special Features: A simple, single LSTM layer model.

### Model 2

Architecture:

LSTM layer with 128 units, ReLU activation function, returns sequences.

LSTM layer with 64 units, ReLU activation function, returns sequences.

LSTM layer with 32 units, ReLU activation function.

Dense layer with num\_classes units, softmax activation function.

Input Shape: (num\_steps, num\_features).

Optimizer: Adam.

Special Features: A deeper model with 3 LSTM layers.

### Model 3

Architecture: Same as Model 2.

Input Shape: (num\_steps, num\_features).

Optimizer: Adamax.

Special Features: 3 LSTM layers, similar to Model 2, but specifically using the Adamax optimizer and Early Stopping.

## Convolutional Neural Network (CNN) Model for Composer Detection Using LSTM Data

The task of detecting composers is accomplished using a Convolutional Neural Network (CNN) designed to recognize patterns in the data processed by Long Short-Term Memory (LSTM) networks. The combination of LSTM and CNN architectures allows the model to capture both sequential dependencies and local patterns in the musical data, making it a powerful tool for this specific application.

### Model Architecture

Input Layer: The input to the model consists of sequences generated by previous LSTM processing. The shape of this input (num\_steps, X\_train.shape[2]) reflects the temporal structure and feature dimensions of the LSTM-processed data.

Convolutional Layer 1: The first convolutional layer, containing 64 filters of size 3 and using the ReLU activation function, is designed to detect local patterns within the temporal sequence. These filters work on segments of the LSTM-processed data to identify relevant local features.

Max Pooling Layer 1: The max pooling layer of size 2 helps reduce the dimensionality of the data, preserving the most salient features. It aids in reducing computation and focuses on the dominant patterns within the local segments.

Convolutional Layer 2: A second convolutional layer with 128 filters of size 3 continues to build on the extracted local patterns, further refining the features that characterize the composers.

Max Pooling Layer 2: Another max pooling layer further compresses the spatial representation, emphasizing the most significant local features.

Flatten Layer: The flatten layer transforms the spatially structured data into a flat vector, preparing it for the dense layers. It maintains all the spatial relationships identified by the previous layers.

Dense Layer: A dense layer with 128 neurons and ReLU activation builds higher-level abstractions from the flattened features. It integrates the local patterns into a global understanding of the data.

Dropout Layer: A dropout layer with a rate of 0.3 is used to mitigate overfitting, ensuring that the model generalizes well to unseen data.

Output Layer: The final dense layer maps the integrated features to the classes representing different composers, using a softmax activation to provide probability scores for each class.

Model Compilation and Summary: The model is compiled with the Adam optimizer and uses categorical cross-entropy loss, reflecting the multi-class nature of the classification task. Accuracy is chosen as the evaluation metric.

The integration of LSTM-processed data with a CNN model presents a novel approach to composer detection. By leveraging the sequential understanding provided by the LSTM and the pattern recognition capabilities of the CNN, the model offers a nuanced and robust means of analyzing complex musical data. The architecture is tailored to exploit both the temporal and spatial dimensions of the data, offering a sophisticated solution to a challenging problem.

## Convolutional Neural Network (CNN) Model for Composer Detection

The Convolutional Neural Network (CNN) designed to recognize patterns in the spectrogram images generated from the composer's MIDI files. CNNs are particularly well-suited for image classification tasks, as they can learn spatial hierarchies of features directly from the data. The architecture of the CNN model for our project is detailed below:

### Model Architecture

Input Layer: The input to the model consists of a sequence of spectrogram images, each represented as a fixed-size matrix. The input\_shape argument defines the dimensions of these matrices, corresponding to the number of time steps (num\_steps) and the number of features in the training set (X\_train.shape[2]).

Convolutional Layer 1: The first convolutional layer consists of 64 filters, each of size 3. These filters slide over the input data to detect local patterns, such as edges and textures. The activation function used is the Rectified Linear Unit (ReLU), a popular choice for introducing non-linearity into the model.

Max Pooling Layer 1: Following the first convolutional layer is a max pooling layer with a pool size of 2. Max pooling helps in reducing the spatial dimensions of the input, retaining only the most important information and thereby reducing computation.

Convolutional Layer 2: The second convolutional layer has 128 filters, each of size 3, and also uses the ReLU activation function. This layer continues the process of feature extraction, learning more complex and abstract patterns.

Max Pooling Layer 2: A second max pooling layer further reduces the spatial dimensions and emphasizes the dominant features.

Flatten Layer: The flatten layer reshapes the pooled feature maps into a single continuous vector, making it suitable for input to the dense layers.

Dense Layer: A fully connected dense layer with 128 neurons and ReLU activation is used to perform higher-level reasoning on the extracted features.

Dropout Layer: To prevent overfitting, a dropout layer is introduced with a rate of 0.3. This layer randomly sets a fraction of the input units to 0 during training, which helps in achieving a more robust model.

Output Layer: The final layer is a dense layer with as many neurons as there are classes (num\_classes, representing different composers). The softmax activation function is used to transform the raw output into probabilities, indicating the likelihood of each class.

# Results Analysis

# Summary

# Appendix A:

### Code Implementation Details

|  |  |  |
| --- | --- | --- |
| IMPORTING LIBRARIES | Importing the necessary libraries to read the dataset: Most are standard to the work done in this class regarding deep machine learning. The new ones that are loading deal with the MIDI file formatting: they are: | **import** os  **import** glob  **import** pretty\_midi  **import** numpy **as** np  **import** matplotlib.pyplot **as** plt  **import** pandas **as** pd  *# Ignore warnings*  **import** warnings  warnings**.**filterwarnings('ignore')  **from** sklearn.metrics **import** accuracy\_score, precision\_score, recall\_score, f1\_score  **import** tensorflow **as** tf  **from** tensorflow.keras **import** layers, models, optimizers  **from** tensorflow.keras.preprocessing.image **import** ImageDataGenerator  **from** sklearn.metrics **import** classification\_report  **from** keras **import** models, layers  *# Check if pickle file exists and use the file for dataset*  **import** pickle |
| Data Processing Functions | Functions created to iterate through composer midi files and create dataset. | #@title 2.1: Extract features using librosa for further feature extraction  def calculate\_features(midi\_file):  # Load MIDI file  midi\_data = pretty\_midi.PrettyMIDI(midi\_file)  # Time interval for calculating features  interval = 1.0 # 1 second  times = np.arange(0, midi\_data.get\_end\_time(), interval)  # Create arrays for storing time series data  pitch = np.zeros(len(times))  volume = np.zeros(len(times))  note\_density = np.zeros(len(times))  tempo = np.zeros(len(times))  # Calculate time series data for each feature  for i, t in enumerate(times):  # Get notes that are playing at this time  notes = [note for note in midi\_data.instruments[0].notes if note.start <= t < note.end]  # Calculate average pitch  if notes:  pitch[i] = np.mean([note.pitch for note in notes])  # Calculate note density (notes per second)  note\_density[i] = len(notes) / interval  # Calculate average volume  if notes:  volume[i] = np.mean([note.velocity for note in notes])  # Calculate rhythmic complexity (variance in inter-onset intervals)  inter\_onset\_intervals = np.diff([note.start for note in midi\_data.instruments[0].notes])  rhythmic\_complexity = np.var(inter\_onset\_intervals)  # Calculate tempo for each moment in time  tempo\_changes = midi\_data.get\_tempo\_changes()  tempo = np.interp(times, tempo\_changes[0], tempo\_changes[1])  return times, pitch, note\_density, volume, rhythmic\_complexity, tempo  #@title 2.2: Process composer data to df  def process\_composer\_data():  # Initialize DataFrame  df = pd.DataFrame(columns=["Composer","Times", "Pitch", "Note\_Density", "Volume",  "Rhythmic\_Complexity", "Tempo"])  # Iterate over all composer directories  for composer\_dir in glob.glob(os.path.join(base\_dir, '\*')):  # Get the composer's name  composer\_name = os.path.basename(composer\_dir)  print(f"Processing {composer\_name} MIDI files...")  # Iterate over all MIDI files in composer's directory  for midi\_file in glob.glob(os.path.join(composer\_dir, '\*.mid')):  print(f"Processing {midi\_file}...")  try:  times, pitch, note\_density, volume, rhythmic\_complexity, tempo = calculate\_features(midi\_file)  # Append to DataFrame  df = df.append({"Composer": composer\_name, "Times": times, "Pitch": pitch,  "Note\_Density": note\_density, "Volume": volume,  "Rhythmic\_Complexity": rhythmic\_complexity,  "Tempo": tempo},  ignore\_index=True)  except Exception as e:  print(f"Error processing {midi\_file}: {str(e)}")  # Write the DataFrame to a pickle file  df.to\_pickle(base\_dir + "/" + pickle\_file\_name)  return df |
| Pickle File Use | Using pickle files to allow one teammate to generate a pickle file and allow of the team not to have repeat work | #@title 2.3: Data Processing - Feature extraction  pickle\_file = base\_dir + "/" + pickle\_file\_name  # Check if the pickle file exists  if not os.path.exists(pickle\_file):  print("Music Data not Pickled, creating dataset using feature extract.")  df = process\_composer\_data()  else:  # Open the pickle file in binary mode and load the data  with open(pickle\_file, 'rb') as file:  data = pickle.load(file)  # Create a DataFrame from the loaded data  df = pd.DataFrame(data)  # Now you have your DataFrame ready for use  print(df.head()) |
| Data Preparation | Preparing the data for LSTM  Split train and test data sets (80-20)  Transform the data for the LSTM | # Convert all other features to have an extra dimension for LSTM  def transform\_series(series, num\_steps):  # Reshape series to (samples, time\_steps, features)  X = np.zeros((len(series), num\_steps, 1))  for i in range(len(series)):  X[i,:,0] = series.iloc[i][:num\_steps]  return X  --------------------------------------------------------------------------------------------------  #@title 3.1: Split train and test data sets (80-20)  # Using stratify to ensure the datasets have same prorportions of each composer as original dataset  df\_train\_val, df\_test = train\_test\_split(df, test\_size=0.2, random\_state=42, stratify=df['Composer'])  # Second, we separate the remaining data into the train and validation sets (75-25)  df\_train, df\_val = train\_test\_split(df\_train\_val, test\_size=0.25, random\_state=42, stratify=df\_train\_val['Composer'])  # The train/val/test split is now 60%/20%/20%  # Encode the labels  encoder = LabelEncoder()  encoder.fit(df['Composer']) # Fit on the whole dataset  # Transform the labels to one-hot encoded form for each subset  y\_train = np\_utils.to\_categorical(encoder.transform(df\_train['Composer']))  y\_val = np\_utils.to\_categorical(encoder.transform(df\_val['Composer']))  y\_test = np\_utils.to\_categorical(encoder.transform(df\_test['Composer']))  --------------------------------------------------------------------------------------------------  #@title 3.2: Apply transform\_series on each feature for each subset  def prepare\_data(df, num\_steps):  pitch = transform\_series(df['Pitch'], num\_steps)  note\_density = transform\_series(df['Note\_Density'], num\_steps)  volume = transform\_series(df['Volume'], num\_steps)  rhythmic\_complexity = np.array([df['Rhythmic\_Complexity'].values]\*num\_steps).T[:,:,np.newaxis]  tempo = transform\_series(df['Tempo'], num\_steps)  X = np.concatenate([pitch, note\_density, volume, rhythmic\_complexity, tempo], axis=-1)  return X  num\_steps = 27  X\_train = prepare\_data(df\_train, num\_steps)  X\_val = prepare\_data(df\_val, num\_steps)  X\_test = prepare\_data(df\_test, num\_steps) |
| Results Visualization | Plotting function to allow visualization of data.  Print the true class labels and predicted class labels  Plot  confusion Matrix  Print Classification Report | #@title 5: Plot the training and validation loss V1  def plot\_learning\_curves(history):  plt.plot(history1.history['loss'], label='Train Loss')  plt.plot(history1.history['val\_loss'], label='Validation Loss')  plt.legend()  plt.show()  # Plot the training and validation accuracy  plt.plot(history1.history['accuracy'], label='Train Accuracy')  plt.plot(history1.history['val\_accuracy'], label='Validation Accuracy')  plt.legend()  plt.show()  return plot\_learning\_curves  --------------------------------------------------------------------------------------------------  #@title 6.1v: Print the true class labels and predicted class labels  for true\_label, pred\_label in zip(y\_test, y\_test\_pred\_class):  true\_class = encoder.inverse\_transform([np.argmax(true\_label)])[0]  pred\_class = encoder.inverse\_transform([pred\_label])[0]  print(f"True Class: {true\_class}, Predicted Class: {pred\_class}")  --------------------------------------------------------------------------------------------------  #@title 6.2: Print a confusion matrix V1  from sklearn.metrics import confusion\_matrix  import seaborn as sns  #mport matplotlib.pyplot as plt  # Generate the confusion matrix  cm = confusion\_matrix(np.argmax(y\_test, axis=1), y\_test\_pred\_class)  # Create a heatmap of the confusion matrix  plt.figure(figsize=(8, 6))  sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=encoder.classes\_, yticklabels=encoder.classes\_)  plt.xlabel('Predicted Labels')  plt.ylabel('True Labels')  plt.title('Confusion Matrix')  plt.show()  --------------------------------------------------------------------------------------------------  #@title 6.3: Report the confusion matrix V1  # Reshape the input data for prediction  X\_test\_scaled\_reshaped = X\_test\_scaled.reshape(X\_test\_scaled.shape[0], num\_steps, num\_features)  # Make predictions on the reshaped test set  y\_test\_pred = model.predict(X\_test\_scaled\_reshaped)  y\_test\_pred\_class = np.argmax(y\_test\_pred, axis=1)  # Convert predicted class labels back to original composer labels using the encoder  lstm\_pred\_labels = encoder.inverse\_transform(y\_test\_pred\_class)  lstm\_true\_labels = encoder.inverse\_transform(np.argmax(y\_test, axis=1)) # Convert true class labels back  # Print the classification report  lstm\_classification\_report\_v1 = classification\_report(lstm\_true\_labels, lstm\_pred\_labels, target\_names=encoder.classes\_)  print("Classification Report - LSTM Model ver 1:\n", lstm\_classification\_report\_v1) |
| Spectrogram generation Code | Convert midi files to wav files using fluidsynth  Generate png files from wav files using librosa | #!/bin/bash  # Set the location of your soundfont file  sound\_font="../TimGM6mb.sf2"  # Iterate over all directories in the current directory  for dir in \*/  do  # Go inside each directory  cd "$dir"  # Iterate over all .mid files in the current directory  for midi\_file in \*.mid  do  # Replace the file extension from .mid to .wav  wav\_file="${midi\_file%.mid}.wav"  # Use fluidsynth to convert the midi file to a wav file  /mnt/host/c/tools/fluidsynth-2.3.2-win10-x64/bin/fluidsynth.exe -ni "$sound\_font" "$midi\_file" -F "$wav\_file" -r 44100  done  # Go back to the parent directory  cd ..  done  -----------------------------------------------------------------------  def generate\_images(dataset\_path):  X = []  y = []  composers = os.listdir(dataset\_path)  for i, composer in enumerate(composers):  composer\_path = os.path.join(dataset\_path, composer)  # Check if it is a directory  if os.path.isdir(composer\_path):  for filename in os.listdir(composer\_path):  if filename.endswith('.wav'):  print(dataset\_path + "/" + composer + "/" + filename)  %matplotlib inline  x , sr = librosa.load(dataset\_path + "/" + composer + "/" + filename)  X = librosa.stft(x)  img\_filename = change\_extension(filename, ".png")  # Generate spectrogram  D = np.abs(X)  # Resize to 224x224  D\_resized = ndimage.zoom(D, (224.0/D.shape[0], 224.0/D.shape[1]))  # Generate the image  plt.figure(figsize=(5, 5))  librosa.display.specshow(librosa.amplitude\_to\_db(D\_resized, ref=np.max), sr=sr, x\_axis='time', y\_axis='log')  plt.tight\_layout()  plt.savefig(dataset\_path + "/" + composer + "/" + img\_filename)  print(img\_filename)  plt.close() |