Oriental Musical Instruments Identification by Selecting Optimized Features and Suitable Classifier

Peranantham Uruthiran

(168075C)

Degree of Master of Philosophy

Department of Information Technology

University of Moratuwa

Sri Lanka

June 2023

DECLARATION

I declare that this is my own work, and this thesis does not incorporate, without acknowledgment, any material previously submitted for a Degree or Diploma in any other University or institute of higher learning, and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment made in the text.

Similarly, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic, or other media. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: UOM Verified Signature Date: 05.06.2023

The above candidate has carried out research for the M Phil. thesis under my supervision.

Name of the supervisor: Dr. Lochandaka Ranathunga

Signature of the supervisor: UOM Verified Signature Date:05/06/2023

ACKNOWLEDGMENT

First and foremost, I would like to thank my supervisor Dr. Lochandaka Ranathunga for his productive support, proper guidance, and supervision, he placed in my research. I express my heartfelt thanks to my progress review panel members Dr. Jayathu Samarawickrama and Dr. C.R.J. Amalraj; they always judge my research work and give feedback to improve my studies.

I convey my sincere thanks to the Vice-Chancellor of the University of Moratuwa, the Dean of the Faculty of Information Technology, the Head of the Department of Information Technology, the non-academic staff members of the Faculty of Information Technology, University of Moratuwa for the opportunity given to commence my research work at the University of Moratuwa and facilitating me to carry out the same successfully.

Further, I extend my sincere gratitude to the Vice-Chancellor of the University of Vocational Technology, and the Head of the Department of Software Technology, and my colleagues in the academic staff for their assistance in many ways, which was a blessing in completing this research study.

I am thankful to the CEO of LK Domain Registry for granting me the Prof. V.K. Samaranayake Research Conference Publication Grant in order to present my research paper at the International Conference on Artificial Intelligence and Data Science (AiDAS 2019) at Casuarina Hotel, Ipoh, Perak, Malaysia.

The support and motivation provided by my postgraduate colleagues, Mr. K.A.S.H. Kulathilake, Mr. V. Senthooran, and Dr. A.M.R. Ravimal Bandara, who are affiliated with the Faculty of Information Technology, University of Moratuwa too deserve mentioning with a debt of gratitude.

I reveal my salutation and admiration for all the valued authors, researchers, and philosophers for their great theories, research, publications, and ideas, which have been an enormous support to enhance this research work.

Finally, yet importantly, the commitment of my loving wife, daughter, and son who were bearing all the burdens without passing them to me throughout the past few years made me complete this research.

ABSTRACT

The research field of Music Information Retrieval is a particular subcategory, which brings out data from the audio signal by the expedient of digital signal analysis. This thesis deals with temporal and spectral features of music instruments. Particularly, the formant concept of timbre is the main subject all through. This theory expresses that auditory musical instrument sounds may be classified with the aid of their formant structures. Ensuring this concept, our method aims to suggest a computer based implementation for constructing tools for musical instrument recognizable proof and grouping systems.

One of the most crucial aspects of musical instrument classification is selecting the relevant set of features, which are very important steps in musical instrument identification. Feature selection is an important task in musical instrument identification. Feature selection is one routine of reaching dimension reduction, and after an ephemeral debate of various feature selection techniques, the study endorses a derived technique for predominant feature selection in a sequential forward feature selection with a greedy algorithm. This technique is empirically selected to optimize the best set of features by using train data and it is displayed to gain classification accuracy with a diminished predominant set of features much like that gained with a complete set of features. This study extracted the 44 features from 20 musical instruments with three musical families.

The three classifiers used in this task, were Decision Tree, kNN, SVM and CNN. The best-selected features have been used in the classification. The confusion matrix got from each classification for evaluation to the performance of the classifiers. The SVM classifier contains the lowest error rate, and the highest AUC scores most values are 1 and a few are within the range of 0.99 - 0.98. Finally, the approval results are finished. SVM classifier is found to be the best classifier among the four classifiers. The predominant features are selected by the Greedy algorithm with SFFS technique for individual musical instrument and selected features are used for polyphonic music identification.

Keywords: Predominant features, Feature selection techniques, Musical instrument. Polyphonic music

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronym	Definition
AC	Autocorrelation Coefficients
AUC	Area Under the receiver operating characteristics Curve
BT	Binary Tree Classifier
CNN	Convolutional Neural Network
CPNN	Counter Propagation Neural Network
DNN	Deep Neural Network
DT	Decision Tree
FFT	Fast Fourier Transformation
HC	Hierarchical Cluster
kNN	k-Nearest Neighbors
LDB	Local Discriminant Bases
LiFT	likelihood-frequency-time
LPC	Linear Prediction Coefficient
LPCC	Linear Predictive Coding coefficients
MFCC	Mel-Frequency Cepstral Coefficient
MIMN	Multiple Instrument Multiple Note
MIR	Musical Information Retrieval
MPEG-7	Moving Picture Exports Group - Multimedia Content Description Framework

Acronym	**Definition**
MUMS	McGill University Master Sample
NMF	Non-negative Matrix Factorization
OMII	Oriental Musical Instruments Identification
PDF	Probability Density Function
RANSAC	Random sample consensus
ROC	Receiver Operating Characteristics
SFFS	Sequential Forward Feature Selection
SIMN	Single Instrument Multiple
SISN	Single Instrument Single Note
SVM	Support Vector Machine
VZCR	Variance of Zero Crossing Rates
ZCR	Zero Crossing Rates

CHAPTER 1

INTRODUCTION

The classification of a musical instrument by using multimedia data processing techniques is an emerging field. Music analysis research has been done in the music industry for the past 28 years because the World Wide Web opened in 1993 to free the public [1]. In 1996, the internet also became for daily usage and information shared with others, and YouTube also launched its platform in 2005 to upload multimedia content like audio and video files [2]. Therefore, huge multimedia contents are available on the World Wide Web now. However, research exists in the area of Music Information Retrieval (MIR) is specially focused on voice recognition, singer identification, and monophonic music identification; there is a few research attempt at polyphonic music identification. Automatic musical instrument identification is a process, in which the computer attempts to identify which portions of the music are played by which musical instruments in the given music. To identify such sounds one should first apprehend what a musical sound or note awareness includes a component of precise sounds that lead to recognizable. Therefore, sound has many characteristics or features that can be analyzed in this study. Many feature schemes were proposed and followed within the literature on musical instrument analysis.

One of the most important aspects of musical instrument classification is finding the right feature extraction technique and selecting the relevant set of features are very important stages in instrument recognition. The pitch, loudness, duration, and timbre are four perceptual attributes of the musical sound [3]. Pitch is directly associated with the fundamental frequency of musical sounds. Loudness is intensity, which is corresponding to the square of the amplitude of the acoustic strain. Duration corresponds with tones. Timbre was called the sound quality or tone color of music [4].

One of the most important and continuous activities in the multimedia data processing field is probably feature selection. For each instrument, a certain collection of attributes is required for sound processing and the teaching process [5]. It diminishes time and increments exactness.

1.1. Music

Music is the art of organizing sounds in a period to produce an arrangement over the features of melody, rhythm, harmony, and timbre [6]. Music is composed by composers and performed by a vast range of musical instruments played by musicians. Music is a mechanical wave. Its oscillation travels through space and energy travels from one point to another [7].

1.1.1. Music from the Human's Point of View

Some musicologists say that music origin from nature. Music is a pattern. It creates mood, emotion, and feeling in the human brain [8]. Its direct effect on the soul. Music has been relieved human stress and used in cures and medicine for some illnesses [9]. It can hear the human auditory system. It senses vibration. Humans can understand rhythm, beat, tone, and tempo with their pre-gained knowledge but they cannot measure them [10]. Musicians study music theory and produce music. Composers study music theory to understand how to compose music and write the chord and score of the music [11].

1.1.2. Music from the Machine's Point of View

Music varies with time always. Machine refers to the identification of content-based patterns in musical information [12]. Signal processing and data mining are involved in analyzing music content [13]. Artificial Intelligence did not produce and identify music without a human structured set of parameters and instructions. The machine can learn the process but the human is intelligent [14].

1.1.3. Musical Instrument

Physically a sound is delivered by waveform dislodging particles in any medium, most commonly air. The current study displays the perception of sounds created by a method of acoustic musical instruments. Three essential capabilities of musical instruments are to yield the sound, to manipulate the frequency content material of the sound, and to venture the sound. Thus, each acoustic instrument can be described in phrases of a generator, resonator, and radiator [15].

There are some types of musical instrument families. These are string, woodwind, brass, percussion, and electronics [16]. This study used string, brass, and woodwind. The string instrument is bowing or plucking the string to produce the music. The vibration of the string causes this sound. Woodwind and brass are holes instruments that blow a stream of air to produce music. All woodwind instruments yield sound by splitting the air blow into them on a pointy edge. The instrument causes the air column in the instrument to vibrate and convey its specific sound. Each of these has some unique characteristics. This study finds the predominant features of each musical instrument. Music of 20 musical instruments have used for this study.

1.2. Audio Features

The audio features defined many numerical values extracted from the audio signal. The sound features are categories of main types spectral shape features, cepstral features, temporal features, amplitude, and frequency modulation features [17].

1.3. Scope of the Study

Automatic musical instrument identification tasks depend on the features. Which needed relevant features and a suitable classifier. There are five research problems identified in this study, which are described below.

- What are the most significant features relevant for oriental musical instrument identification?
- How to drive a feature selection approach to obtain the most sensitive features for selected oriental musical instruments?
- How to select the best set of predominant features for each individual musical instrument identification?
- What is the most suitable classifier for the selected feature sets?
- How to identify musical instruments in presence of polyphonic music?

This study first focuses on relevant feature extraction techniques. There are some audio characteristic extraction toolboxes accessible, acquainted with the open use in varying configurations, however generally as at least one of the following formats:

- Standalone solicitations or specific applications
- Plug-ins for a host application
- Software function library

Input query to any records retrieval system includes relevant and beside-the-point information. Functions choice is a technique for eliminating as many irrelevant facts as is practical while presenting crucial facts in a condensed and significant way. At some stages in the technique of offering MIR deals with analyzing, indexing, and searching in a large number of digitized music, it has investigated a selection of functions consisting of some statistical characteristics and a few primary signal properties. Additionally, the study extracted features including FFT, Spectral Chroma Coefficients, and MFCC from musical melodic information using some of the virtual signal processing algorithms [18].

1.4. Aim and Objectives of the Study

The study aims to develop a computerized system to automate the identification of oriental musical instruments based on the music source with a suitable filter and classifier.

The study focuses on the following objectives to fulfill the aim of the research. The study focused,

- To form an Experimental framework for musical instrument identification.
- To find the potential feature extraction techniques.
- To identify a suitable feature selection method for the best feature selection.
- To identify the most effective set of features for a group of musical instruments.
- To identify the most suitable classifier for musical instrument identification.
- To evaluate the new music (audio) file with the selected features and classifier.

1.5. Organization of the Thesis

This thesis is presented in six chapters to apprise the noteworthy facets and outcomes of this research study. The sound properties and feature categories, background of the study, problems in feature extraction, feature selection, and feature learning in past research works have been elaborated in the literature review and the latest works attempt to enrich the above-stated activities have been discussed in Chapter 2.

The low-level features, their definition and mathematical formulations, feature extraction techniques, feature selection methods, feature learning methods, machine learning algorithms, and source separation have been deeply explained in Chapter 3.

Chapter 4 describes the proposed theatrical framework as the Oriental Musical Instruments Identification (OMII) process, dataset preparation, data task, and details of extracted features.

Chapter 5 elaborates on the experimental task, the best set of feature selection, and musical instrument identification determines the predominant feature of each musical instrument, and discussion of the results, validation, and decision making. Finally, the conclusion of this study has been presented in chapter 6.

1.6. Summary

This chapter presents a spacious overview of the research. It covers the concept of the music, describes two types of points of view of the music, and defines the functionality of the musical instruments. It explains the types of musical instrument families and which musical instruments are fallen into the categories. It states specific problems, the scope of the study, and the aims and objectives of the study.

CHAPTER 2

BACKGROUND OF THE STUDY

Sound has different perceptual properties. Researchers had focused on capturing the perceptual properties.

2.1 Timbre Features

Timbre is likewise recognized for the sound quality or tone coloration of the track. It's defined as when sounds had been heard that suit a similar pitch, similar loudness, and similar duration, and a distinction can still be heard between the two sounds, that difference is alluded to as timbre [19]. There are two physical correlates of timbre: temporal envelope, which has attack time characteristics, and amplitude modulation. Other spectral distributions are spectral envelop, number of partials and energy distribution of partials. The following set of features is dominant in the most sound property identifications [20].

2.1.1 Spectral Shape Features

Most of the audio signal data analysis starts from the time domain to the frequency domain by extracting Short Time Fourier Transform (STFT). The spectral function computed from the Short Time Fourier rework of the signal. These encompass centroid, spread, slope, skewness, kurtosis, roll-off, and so on. [21].

2.1.2 Temporal Features

These features are computed directly from the audio signals. Autocorrelation Coefficients (AC), which address the general pattern of the spectrum, has been said to be beneficial in Zero Crossing Rates (ZCR), which help separate periodic signals (small ZCR values) from boisterous signals (high ZCR values) [22].

2.1.3 Cepstral features

Mel-frequency Cepstral Coefficients (MFCCs) are assumed about as well as their time first and second derivatives, which are estimated in many successive frames [22]. It has 13 coefficients, 13 delta coefficients, and 13 delta-delta coefficients features derived from the audio signal. It is mostly used in voice identification.

2.1.4 Amplitude and Frequency Modulation Features

Some musical instruments have amplitude modulation or frequency modulation, which will detect that modulation and use it to divide instruments with modulation. Therefore, we first want to recognize what amplitude modulation seems like in the time and frequency dominion. Amplitude modulation is utilized in a communique to hold facts via a signal, which has a fairly low frequency whilst in comparison to the carrier [23].

2.2 Harmonic and Overtone

A sound wave has discussed the fundamental frequency of a vibrating string and its scaling with length, which are also aware of the fact that there are many possible modes of vibration each with its vibrational frequency [24].

Many instruments have truly equal overtones present and regularly with very comparable relative amplitudes. What is different about them are the relative amplitudes. The amendment within the amplitude of an overtone over time is its time envelope.

The mixed time envelopes of the distinctive overtones (It can name as just "envelope" for simplicity) also are a crucial factor of instrument recognition.

Overtones that might be relatively stable for a specific period are referred to called overtone. Overtones that partials are best audible for a small duration, which are normally referred to as transients, and preliminary transients are crucial for musical instrument identification. The necessary aspects of pitched instrument identification are the envelope and also the complete overtones spectrum, collectively with partials and transients. Instruments typically have a few distinctive characteristics that can be described using their harmonic spectra and their temporal and spectral envelopes [25]. MPEG – 7 audio frameworks describe 7 features, which are harmonic centroid, harmonic deviation, harmonic spread, harmonic variation, spectral centroid, log attack time, and temporal centroid [26].

2.2.1 Harmonic Centroid

It is calculated as average of the instant harmonic spectral deviation in each frame across the sound section period [27]. The amplitude in the directly scale-weighted imply of the harmonic top of the spectrum is used to derive the spontaneous harmonic spectral centroid mass.

2.2.2 Harmonic Deviation

It is computed over the sound phase span at the spontaneous harmonic spectral deviation in every frame [27]. Due to the log-amplitude factors' spectral deviation from a spectral envelope, the spontaneous harmonic spectral deviation is seen.

2.2.3 Harmonic Spread

It is calculated as the instantaneous harmonic spectral spread of the frame averaged across the sound segment duration. [27]. The amplitude-weighted normal deviation of the harmonic peaks of the spectrum with respect to the instantaneous harmonic spectral centroid is used to measure the instantaneous harmonic spectral spread.

2.2.4 Harmonic Variation

It is defined as the instantaneous harmonic spectral difference's mean value calculated across the length of the sound section. [27]. Because of the standardized correlation between the harmonic peak amplitudes of two consecutive frames, the immediate harmonic spectral difference is highlighted.

2.2.5 Spectral Centroid

It is the root of the deviation of the log frequency power spectrum from the gravity center in a frame as measured by the mean square of the deviation. [27]. Similar to the spectrum centroid, it is related to the power spectrum's shape. It is also known as the sound's brightness.

2.2.6 Log Attack Time

The signal envelope is predicted by computing the local mean square value of the signal amplitude in each frame. It is defined as the logarithm of the period among the instances whilst the signal begins to the time it arrives at its steady part [27].

2.2.7 Temporal Centroid

It is calculated as the time average over the energy envelope.

2.3 Feature Extraction

There is a large number of audio characteristic extraction toolboxes available, introduced to the open use in differing formats, however generally as at least one of the following formats:

- Standalone applications
- Plug-ins for a host application
- Software function library

Input query to any records retrieval system includes relevant and beside-the-point information. Functions choice is the method of disposing of as plenty inappropriate facts as feasible and addressing realities facts in a compressed and significant shape. Kim *et al* [28] have explained some stages in the technique of offering a MIR machine, A selection of functions consisting of some statistical characteristics and a few primary signal properties such as zero-crossing rate and energy. Additionally, they used some of the methods for virtual signal processing to extract features like the FFT, LPCC, and MFCC from musical data [18].

The technique of audio classification includes extricating prejudicial characteristics from the audio records and taking care of them to a pattern classifier. Some strategies create pattern-shaped characteristics and use them for cataloging according to the degree of correlation. Several distinct tactics combine the statistical classification method with the mathematical values of the qualities. Pitch-identification methods can be broadly categorized into two classes: time-domain-based methods and frequency-domain methods [18].

ZCR is a well-known method in the time domain. The fundamental idea is that every unit of time, ZCR provides information on the spectral content waveform passing zero. Recently, ZCR appeared in a particular shape with VZCR or SZCR (smoothing ZCR) [18] files than ZCR; FFT is arguably the most well-known method. This method is based on the idea that any waveform can be broken down into basic sine waves as often as feasible. However, a low spectrum ratio for the longer windows can also help increase frequency resolution while simultaneously reducing time resolution [29]. Another drawback is that while music pitches are best designed on a logarithmic scale, the quality FFT frequency intervals are linearly distributed. The different methods used for feature extraction are MFCC (Mel-frequency cepstral coefficient), LPC (Linear Prediction Coefficient), and LDB (Local Discriminant Bases).

2.4 Classification

In mid of the 1990s, the musical instrument reorganization problem has been addressed by machine learning techniques. Bormane and Meenakshi [4] have created musical instrument classifiers using methods such as decision tree, multi-layered perceptron, support vector machines, k-nearest neighbors, Sequential Minimal Optimization algorithms, multi-class classifiers, and self-organizing maps among others. They use the number of features calculated from a choice of sound samples to educate and take a look at those classifiers.

2.5 Findings of Related Studies

For categorizing musical data, Arijit *et al.* [30] have presented a hierarchical classification. The intended scheme relies on MFCC rather than managing an excessive style of features. They employed a two-stage technique; particularly, string, woodwind, percussion, or keyboard signals are collectively referred to as instrumental signals. Features that are employed for classification serve as the foundation for wavelet and MFCC. The MFCC pattern, which appears in the high sub-band of wavelet-decomposed signals, is used to classify songs. Random sample consensus (RANSAC) is the best classifier, and it is employed at all stages of the classification process.

A framework for developing and comparing options for the content of content-based evaluation of musical signals has been provided by Bormane and Meenakshi [4] Wavelet disintegration is applied to the input signal. We select an optimum wavelet for disintegration. They used a Wavelet Packet transform to solve for deterioration. Spectral centroid, spectral Roll-off, Spectral Flux, energy, Zero Crossings, Linear Prediction coefficients, and Mel Frequency Cepstral Coefficients are examples of common timbre characteristic options (MFCCs). MFCCS is the most frequently used of these methods for music recognition.

Victimization with musical instruments is categorized by.Bhalke *et al.*[20] Higher Order Spectral Features and Mel Frequency Cepstral Coefficients (MFCC). In the task of instrument grouping, MFCC, cepstral, temporal, spectral, and timbre features are widely used. Counter Propagation Neural Network (CPNN) is given the retrieved features in order to identify the instruments and their family.

Essid *et al.* [22] have the notion approximately making use of statistical pattern reputation strategies to address and tackle the matter within the setting of solo musical expression. Ten instrument classes from absolutely one of a kind instrument families area unit thought-about. Over a hundred and fifty signal process options area units were studied together with new descriptors. Two characteristics choice methods, dormancy fraction expansion by characteristic area projection and genetic algorithms region unit thought-about in an exceedingly category pairwise manner. For the classification task, consequences are supplied by the use of Gaussian mixture models (GMMs) and support vector machines (SVMs).

A music signal was divided into 5 phases by Kharb and Hooda [31] using discrete wavelet treatment. The properties of the signal at each phase's Kurtosis, Skewness, and quantitative relation of power compaction are then identified. SVM and k-NN classifiers can be used as a classifier because they are ready for the attributes. Signals are subjected to wavelet transformation to extract complex information present in the signal.

For the purpose of identifying musical instruments, Kothe and Bhalke [32] presented feature analysis using wavelet coefficient histograms and compared it to conventional characteristics. An innovative proposed wavelet coefficient histogram attribute changed into discovered compact and powerful with predominant conventional attributes. The wavelet evaluation presents spectro-temporal facts of the music signal. By simultaneously passing the signal through a Low Decay Filter (LDF) and a High Decay Filter (HDF) in a successive tree-like arrangement, the wavelet packet assessment breaks down the evidence into packets. The Wavelet Transform received an accuracy of 76.83% contrasted with 73.82 % considering the use of MFCC and other features utilized by the system.

The identification of several North Indian musical instruments including the flute, sitar, dholak, bhapang, and mandar were worked on by Kumari *et al.* [33] The spectrum properties, which include spectral centroid, Spectral Spread, Spectral Skewness, Spectral Kurtosis, Spectral Slope, and Spectral Roll-off, were extracted from the enclosed music signal and classified according to each musical instrument. Additionally, Auto Correlation was given a higher ranking among the tools. In comparison to certain other features, it has been discovered that MFCC with thirteen coefficients provides a more reliable identification of monophonic instruments (Indian musical instruments).

Through the application of support vector machines, Ozbek *et al.* [34] evaluated the class's overall performance of a likelihood-frequency-time (LiFT) evaluation designed for partial monitoring and automatic recording of music. With a filter bank designed to filter 24 quarter-tone frequencies of an octave, the LiFT examination relies on constant-Q filtering of signals. For 19 musical instruments and 36 notes, the correct classification fractions were received.

Shelar and Bhalke [35] have developed a system and it is enforced in 2 stages; the primary stage could be a musical instrument recognized utilizing spectral attributes once recognizing the instrument musical notation is recognized victimization completely different frequency estimation strategies. As a classifier, a feedforward neural network has been employed. For Single Instrument Single Note (SISN), Single Instrument Multiple Note (SIMN), and Multiple Instrument Multiple Note, the framework is enforced (MIMN). The three instruments' average accuracy is assessed at 80%.

Bhalke *et al.* [19]have worked with 10 musical instruments with spectral, temporal, and modulation features are used and those are recognized using different Hidden Markov Model algorithms. Identification accuracy attained for monophonic musical notes is 91% and 87 % for polyphonic musical notes.

By extracting MFCC features and timbre-related audio descriptors from monophonic audio recordings, Jadhav [36] was able to recognize musical instruments. Additionally, using the feature vector generated by the component extraction method, three classifiers k-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and Binary Tree Classifier (BT)—were used to identify the musical instrument name. The evaluation is built from learning outcomes obtained using every possible combination of feature extraction techniques and classifiers. In order to test whether certain blends will produce more musical instrument recognized proof findings, percentage accuracies for each combination are calculated. For five, ten, and fifteen musical instruments, respectively, the structure offers improved percent accuracies of 90.00%, 77.00%, and 75.33% when MFCC is employed with a K-NN classifier. If BT classifier is utilized, better percent accuracies of 88.00%, 84.00%, and 73.33% are obtained for timbre ADs for five, ten, and fifteen musical instruments, respectively.

Philippe and Douglas [37] have developed a system consisting of a Deep Belief Network on the Discrete Fourier Transformation technique and Support Vector Machine classifier. It obtains 84.3% accuracy of classification.

The impact of four feature selection algorithms, including genetic algorithm, forward feature selection, information gain, and correlation, has been described by Kalapatapu *et al.* [38] based on four different techniques. The feature sets are extracted features from the MIR tool case in MATLAB, Neural network and SVM classifiers are the best-appropriate classifiers for the Indian Song dataset.

Bhalke *et al.* [39] have given the classification of musical instruments using MFCC and Higher-Order Spectral features of musical instruments. Conjugation of MFCC and

Higher-Order Spectral-based features are employed in the undertaking of musical instrument classification. To discover the instruments and their family, the collected features are fed into a Counter Propagation Neural Network (CPNN). For investigation, the remoted audio files of nineteen musical instruments are utilized from McGill University Master Sample (MUMS) audio database. That system achieved twelve features with the best accuracy.

The summary of the research findings of the literature studies is, that most of the researchers used spectral and temporal features of the sound in the classification process, Short Time Fourier Transformation, and MFCC feature exaction technique. The most identification processes is used to k-NN, BT and SVM classifiers and few processes is used to deep learning classification. The maximum musical instrument identification accuracy values are 91% for monophonic music and 87% for polyphonic music in the Bhalke *et al* research study.

2.6 Summary

This chapter has accentuated the background of the study. It has provided an overview of the scientific works in the field of musical instrument identification. The audio features have been elaborated from the literature review. The most important features are emphasized in the literature reviews. It has elaborated existing feature extraction techniques, feature selection methods, and classifications. Finally, structured the findings of the literature surveys in section 2.5.

CHAPTER 3

LOW-LEVEL FEATURES AND FEATURE LEARNING ALGORITHMS

This chapter explains the Low-Level features, feature extraction techniques, feature selection method, and classifiers. Audio features consist of a descriptor with condensed relevant information. The timbre is multi-dimensional perceptual property. Timbre features typically focus on the spectral shape. This study it has used 44 features for musical instrument identification, there are 11 spectral features, 13 MFCC features, 12 Spectral Pitch Chroma features, and 8 Time domain features. However, all the features are not relevant in all musical instrument identification processes. Which are described in this chapter.

3.1 Spectral Features

3.1.1 Spectral Centroid

It is a spectrum characteristic. It establishes the location of the spectrum's "center of mass," and the definition of the spectrum centroid of the frame is given in equation form (3.1) [40]. It can calculate the power spectrum and logarithm of the frequency scale also.

$$Spectral\ Centroid = \frac{\sum_{n=1}^{N} P(f_n) f_n}{\sum_{n=1}^{N} P(f_n)} \tag{3.1}$$

Where $P(f_n)$ represents the magnitude sample of the n^{th} frame. n = 1, 2,... N-1 and $f_n$is the frequency of the n^{th} frame.

3.1.2 Spectral Crest Factor

It compares a few shapes of the average level with the maximum peak level over an equal time window. It is defined as in equation (3.2).

$$Spectral\ Crest\ Factor = \frac{|x_{peak}|}{x_{rms}} \tag{3.2}$$

3.1.3 Spectral Decrease

The spectral decrease major, which is frequently used in music analysis, shows how much the spectrum has decreased while accentuating the slope of the lowest frequency. [41]. It is computed as in equation (3.3).

$$Spectral\ Decrease(n) = \frac{\sum_{k=1}^{K/2-1} \frac{1}{k}.(|X(k,n)| - |X(0,n)|)}{\sum_{k=1}^{K/2-1} |X(k,n)|} \tag{3.3}$$

It takes the value at the 0^{th} frequency and subtracts it from the k^{th} value, then weigh the subtracted value multiplied by 1/k. The high k-value will have less impact of the low value of the spectral decrease.

3.1.4 Spectral Flatness

It examines the geometric mean of the strength spectrum with the arithmetic mean. It is used in the MPEG-7 standard. It is defined as in equation (3.4).

$$Spectral\ Flatness = \frac{\exp(\frac{1}{N}\sum_{n-0}^{N-1} \ln x(n))}{\frac{1}{N}\sum_{n-0}^{N-1} x(n)} \tag{3.4}$$

Where *x(n)* represents the magnitude of the bin number n.

3.1.5 Spectral Flux

It is a percentage of the power spectrum's variation. By comparing the power spectrum for one frame to the power spectrum from the previous section, it is determined [40]. It is defined as in equation (3.5). The spectral flux is major changed along with the signal if the signal does not change the flux value is zero otherwise shows a specific value.

$$Spectral\ Flux = \sum_{n=2}^{N} |P(f_n) - P(f_{n-1})| \tag{3.5}$$

3.1.6 Spectral Kurtosis

It is a statistical tool used to identify the presence of several transients and where they are located in the frequency domain. Equation is used to express it (3.6).

$$Spectral\ Kurtosis = \frac{1}{N}\sum_{n=1}^{N} f(n)^4 - 3 \tag{3.6}$$

3.1.7 Spectral Roll-Off

This action shows the right-skewness of the power spectrum. It is stated in equation (3.7).

$$Spectral\ Rooll\_off = \sum_{0}^{f_{ny}} A^2(f_n) \tag{3.7}$$

Where A is amplitude and f_{ny} is Nyquist frequency.

3.1.8 Spectral Skewness

The difference between the shape of the spectrum below the usual center of gravity and the shape above the mean frequency is measured. The probability density function (PDF) of the amplitude in the time series is shown to be in equilibrium via skewness. A time series that includes a range of values with equal and opposite amplitudes has a skewness of zero. A time series with a lot of small values and a few large values is skewed (right tail), and the skewness index is positive. A period succession that is negatively skewed (has a left tail) has many enormously large values and few insignificant ones. It is stated in equation (3.8)

$$Spectral\ Skewness = \frac{\frac{1}{n}\sum_{i=0}^{n-1}(X_t(i) - \mu)^3}{\sigma^3} \tag{3.8}$$

Where n is the number of samples of the input time series Xt, m is the arithmetic mean of Xt, and s is the standard deviation of Xt. It is defined as in equation (3.9).

3.1.9 Spectral Slope

It is the assertion that reflectance is wavelength-dependent. Equation is used to express it (3.9).

$$Spectral\ Slope = \frac{R_{F_0} - R_{F_1}}{\lambda_0 - \lambda_1} \tag{3.9}$$

Where R_{F_0}, R_{F_1} {display style R0{F0{0}},R1{F1{1}}} are the reflectance measured with filters F_0, F_1 having the central wavelengths $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$, respectively.

3.1.10 Spectral Spread

It serves as a measure for the spectrum's mean diffuse with respect to its centroid [40]. Equation is used to express it (3.10). It computes how far the spectrum spread around the centroid. This feature can sense to use technical perfection.

$$Spectral\ Spread = \sqrt{\frac{\sum_{n=0}^{N/2}(P(f_n) - SC\ )^{\ 2}}{\sum_{n=0}^{N/2}(P(f_n))^{\ 2}}} \tag{3.10}$$

Where $P(f_n)$ represents the magnitude sample of the n^{th} frame. n = 1, 2,... N-1.

3.1.11 Spectral Tornal Power Ratio

The tonal power ratio is defined as major tonal energy divided by complete spectrum energy. Here not only do we look at the maximum of the spectrum but also we look at all local maxima. It is defined as in equation (3.11).

$$Spectral\ Tonal\ Power\ Ratio = \frac{E_t(n)}{\sum_{k=0}^{\frac{K}{2}-1}|X(k,n)|^2} \tag{3.11}$$

Where it is tonal energy and X is at n point.

3.1.12 Mel Frequency Cepstral Coefficients (MFCC)

MFCC features are mostly used in speech recognition. According to Kumari et al. [33] A signal's cepstrum is the Fourier transform of the decibel signal's logarithm and the Mel frequency cepstrum. The frequencies are scaled logarithmically using the Mel scale. Typically, five processing steps are involved for the MFCC which are computing the magnitude of the spectrum, linear frequency converts scale to the logarithm, group bin into the band, applying logarithm to all bands, and computing inverse cosine transform. It computes 39 coefficients but, the first 13 coefficients are sufficient for our classification task. It is defined as in equation (3.12).

$$mel(f) = 2595 log_{10}\left(1 + \frac{f}{700}\right) \tag{3.12}$$

The Mel scale has 40 filter channels. The primary pull-out filter bank result is the signal's measured strength, and the 12 lines space results address the spectral packet. The rest of the 27 log space channels are harmonic of the signal. The filter outputs are then transformed using a discrete cosine change to produce the MFCCs. For classification, the first thirteen obtained coefficients' average and variance were taken.

3.1.13 Spectral Pitch Chroma Features

The Chroma feature shows a high degree of invariance to variations in timbre. The kind time energy spreading of the original musical signal is encrypted using a 12-dimensional Chroma feature over the 12 Chroma bands, which correspond to the 12 pitch classes. It is defined as in equation (3.13).

$$C_f(b) = \sum_{n=0}^{Z-1} \left| X_{lf}(b + n\beta) \right| \tag{3.13}$$

Where X_{lf} is a log frequency spectrum, n is an integer octave index ∈ [0, Z-1], Z is the number of octaves, b is an integer pitch class (Chroma) index ∈ [0, β-1] and β is a bin per octave.

3.2 Feature Extraction Techniques

The selection of the optimized features for the musical instrument identification process does not directly use music files. At this time many various features have been handcrafted and handcrafted features are computed from the arbitrary definition, simple to compute, and mostly focus only on specific technical properties, which can use in machine learning still worthy. Therefore, many feature extraction techniques are available now. But all are not suitable for the particular situation.

3.2.1 Fast Fourier Transformation (FFT)

The basic idea of the FFT method is to divide and conquer. The problem dividing into smaller ones with similar structures. The problem successfully solved each of the smaller problems. In the signal processing function y(t) is measured over the time interval $0 \leq t \leq T$ is defined as equation (3.14) for all frequencies.

$$Y(\omega, T) = \int_0^T y(t) e^{-i2\pi\omega t} dt \tag{3.14}$$

Where Y is a function frequency. Here they are not thinking about frequency bandwidth anymore. Just looking at all of them jointly, easily computing a convex frame produces covariance matrices that are very large if it calculates covariance matrices for all frequency bins. It takes consists of time and resources, but it is an extremely powerful way to get all of the data to find the relevant frequency.

3.3 Feature Selection

A particular dataset has many features, sometimes the number of features increases the threshold value, and it is decreased the accuracy of the model. Whenever they are used to train data for the model, it may reduce the accuracy of the results. It is the process of identifying significant features and eliminating unrelated features. Vipin Kumar and Sonajharia Minz [42] defined four steps of the method in the feature selection process, which are subset generation, assessment of the subset, preventing criteria, and the result of validation. The subset generation is an investigative exam wherein every state requires a feature subset to determine the value in the search space. It has two fundamental sub-processes, one is successor generation to the choice beginning position and the pursuit direction, which are forward, backward, rank, and random, and the other is search organization, which are sequential, exponential, and random search [43]. This study selected a forward and sequential search technique to develop the best accuracy of the system. The valuation standards are utilized for getting an optimal feature subset by using the classifier and in terms of induced classifier accuracy [44]. The stopping criteria are used in the greedy algorithm to find the best feature subset [45].

3.3.1 Filter Method

A set of all features, choosing the best subset of features, and a machine-learning algorithm make up its three parts. The best sub-set features select from the set of all features and use some basic techniques are the ANOVA test, Chi-square test, and correlation coefficient these are basic use statistical tests [46]. This study uses the correlation coefficient technique to find some important features but these are selected without target output. That means the input features correlated with each other whether one is highly correlated to the other. The filtering approach assesses each feature's association and ranks them according to their coefficients. [47]. However, it is not suitable for this study because it is a selected feature without target output and does not reflect the highest accuracy value to the model. It can omit the features that are not useful but can be extremely valuable when joined with others [42].

3.3.2 Wrapper Method

In a dataset with the desired output, the wrapper technique selects the optimal way to recognize the strength of a small number of features or a lone feature. Table 3.1 is illustrated the general wrapper algorithm. This method has two basic mechanisms, which are sequential forward feature selection and backward elimination. The sequential forward feature selection mechanism is an iterative method in which it starts with having no feature in the model; we keep adding the feature, which gives the best improvement in the model until the expansion of another feature doesn't work on the performance of the model [48]. The backward elimination mechanism is started with all features and removes the least dominant feature from the set in each iteration, which improves the performance. It will iterate until no improvement is observed in the elimination of features. Because the Greedy search algorithm has the maximum accuracy value for the proposed system, sequential forward feature selection was used in this study to find the most suitable features [49].

Table 3.1: General Wrapper Algorithm.

INPUT:
D={X, L} //a training dataset with n number of features where   *// X={f1, f2, f3, ..., fn} and L labels*   *X' // predefined initial feature subset* ($X' \subset X$ *or* $X' = \{\phi\}$) θ *// a stopping criterion*   *OUTPUT:* X'_{opt} *// an optimal subset*
Begin: *Initialize:* $X_{opt} = X'$; $\varphi_{opt} = E(X', A)$ *//evaluate X' by using mining algorithm A*   *Do begin* X_g = *generate(X); // subset generation for evaluation* $\varphi = E(X_g, A)$*; //* X_g *current subset evaluation by A*   *If* ($\varphi > \varphi_{opt}$) $\varphi_{opt} = \varphi$ $X_{opt} = X_g$;   *repeat (until* θ *is not reached);*   *end*   *return* X'_{opt};   *end;*

The above algorithm has been used from the study in *[42]*.

3.3.3 Embedded Method

This method is combined with filter and wrapper methods. It performs the learning process and features selection altogether without separation. It can be saved computation costs larger than the wrapper method [50]. The embedded methods mostly used a learning algorithm SVM to select the feature and some feature selection algorithms are decision tree, weighted naive based and weighted vector of SVM [51].

3.4 Greedy Algorithm

The greedy algorithm all the time makes the selection that looks to be best at that instant. That means the algorithm achieves the optimal solution for a given problem. This approach is simple, easy to implement, and runs fast. Therefore, we used it in this study. The greedy algorithm finds the minimum number of features that give the highest accuracy value for the classification. It is mixed with the sequential forward selection method. The structure of the algorithm is shown in Table 3.2.

In the algorithm S is the subset of selected features. At first, the candidate subset is empty and in the first iteration, the feature that has the highest score is added to the candidate subset. Then a classifier is trained based on the candidate subset and the existing training data. The classification accuracy is maintained as the best result. The next step is done in two phases. In the first phase, a feature with a high score that has not been evaluated yet is replaced with each feature in the candidate subset. After each replacement, a new classifier is trained by using the obtained subset. Then the classification accuracy is calculated. If the addition of a new feature causes an increase in classification accuracy compared to the previous subset, the result is maintained as the best. In this way, the dependence of this feature on all previously selected features is measured and if it does not depend on any of the selected features, it will be added to the candidate subset.

In the second phase, the feature that is under review (the feature that was replaced with the features in the selected subset in the first phase) is added to the selected subset S (which was obtained in the previous stage), and a new classifier is trained based on the new subset and the classification accuracy is calculated. If the accuracy of the subset is higher than the accuracy of the candidate subset of the first phase, it is maintained as the best result. After the first and second phases, if we have achieved a better subset in each of these phases, the optimal subset is selected as the subset of this iteration and the feature is applied to the selected subset.

Table 3.2: Forward Feature Selection with Greedy Algorithm.

```
: D: Training Data  C: Classifier  F: Feature set
ut : S: Selected Feature subset
S= {F[1]}
Bestfeaures = evaluate(C, S, D)
//Phase I
    For i=2 to F.size( )
        Best = null  // Candidate subset is empty
        For j=1 to S.size( )
        S_sel = update(S, S[j], F[i])  // the feature that has the highest score is added to the
                                        // candidate subset
            Data = evaluate(C, S_sel, D)  // calculated accuracy
            If (Data>Bestfeaure) // classification accuracy compared to the previous subset
                best =(S[j], F[i]) // replaced with each feature in the candidate subset
                Bestfeature = Data // accuracy is maintained as the best result
                Saux = S U{F[i]} // replaced with each feature in the candidate subset
//Phase II
                Data = evaluate(C, S_sel, D) // calculated accuracy
            If (Data>Bestfeaure) // classification accuracy compared to the previous subset
                best =add( F[i] ) // added to the candidate subset
                Bestfeature = Data // assign new highest accuracy value
            If (best!=null)
                Update(S, best) //Optimal subset is selected
```

The above algorithm in Table 3.2 is used from the study in [52].

3.5 Feature Learning

The principle of feature learning takes a dataset, but the dataset has been huge, it has been a very significant number-learning sample. At that moment, it learns some way to reduce dimensionality while keeping as robust information as possible it’s relevant to the task. The specific features might contain more useful information, which is handcrafted to use feature learning. Feature learning is not required expert knowledge anymore, because we flow the data write an algorithm, and check if it is failed or not.

3.5.1 Feature Learning Methods

The inductive learning or predictive learning, it has given the examples of data, the examples are formed as a function of x and y where x for a particular instance. The output attribute is y, and the x contains the values of the instance's various features. In order to make the assumption that the instance's output is a function of the input feature vector, it may alternatively be thought of as being given x and f(x). Problem f(x) in categorization is discrete. This function f(x) in the regression issue is continuous. When attempting to determine the likelihood of a specific value of y, some difficulties involve probability estimation, where f(x) represents the probability of x. A new sample is predicted by the function f(x), this function is called the hypothesis space of the particular model. The Hypothesis space depends on features and language, once we have chosen the features and language or class function we have the hypothesis space.

Hypothesis space may represent by H and the learning algorithm outputs a hypothesis h belonging to H. Let us look at the classification problem, music has multiple features, it calls a feature vector, which has n features the feature vectors are n-dimensional space. The machine learning approach automatically learns features from the data set. Artificial intelligence has two sorts of realizing, which are supervised learning and unsupervised learning. The supervised learning split the difference of the information and the comparing result, every data occurrence has the input x and the corresponding output y, from the artificial intelligence structure will construct a model, that has offered another viewpoint x will attempt to expect output y. Unsupervised learning has to only given input data x without output label data, then machine learning clusters group data or finds some pattern of them. This study used supervised learning for predominant feature selection and monophonic music identification, then the polyphonic music identification process was used for semi-supervised learning.

3.5.2 Machine Learning Algorithm

There are numerous commonly used machine learning algorithms out there, but this study used four types of machine learning algorithms, which are DT, k-NN, SVM, and

CNN. These algorithms have different techniques used to learn the features and classify the data.

3.5.2.1 Decision Tree

The decision tree is a classifier of the tree structure and the tree has two kinds of hubs, which are decision hubs and leaf hubs. The decision hubs indicate a decision or test grounded on this; it can conclude which way can go. The leaf hubs show the classification of the example. It started with the root of the tree and depended on the worth of the test. It goes to the related branch and keeps doing this until it arrives at the leaf hub. It is the supervised learning algorithm.

3.5.2.2 k-Nearest Neighbor

Instance-based learning is called the k nearest neighbor algorithm. It is supervised learning, it gets training examples (x, y). A set of them (x_i, y_i) can say f(x_i). When we get training examples, we do not handle them and learn a model of all things. It simply stores examples, when it needs to classify cases that time it does the prediction. It gets the test case, it uses the stored case in the memory to find the conceivable y, and it finds a new x case for the corresponding possible predictive value of y, which is a very closed or neighboring case's value of x. In here, it consider distance function at the metric. The distance function is Euclidean distance it is defined in equation (3.15). The attribute weight is different; the Euclidean function also multiplies with the weighted value. In instanced space, we find what is the closest instance in this nearest case the distance is similar to the given sample averaging under circumstances it has noise but still closer than the case point all attributes at the same scale then secondly we make the assumption one attribute more attend the other attributes. We use weighted distance for our studies.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{N} (x_{ik} - x_{jk})^2} \tag{3.15}$$

Where x_i and x_j are in i^{th} sample features and j^{th} sample features.

3.5.2.3 Support Vector Machine Learning

One of the best classifiers among those that have some linearity is the support vector machine. It has excellent mathematical comprehension and can handle specific nonlinearity scenarios by applying kernel functions that deal with nonlinearity. [53]. Why now the SVM is so popular? It appears that offers a good method for avoiding overfitting and working with more features without requiring excessive processing.

3.5.2.4 Deep Learning Techniques

The deep learning neural network (DNN) is required to learn a huge amount of training data, because it is difficult to comprehend how the networks classify test instances into the appropriate categories through the deep network layers, DNNs are often referred to as "black boxes." [54]. CNN is a deep neural network that is specialized for image processing. It is mostly used in the image processing pattern recognition to the Two Dimensions (2D) application but wave-net approach analyzed One Dimension (1 D) audio raw data directly used in the waveform. The wave-net is used to analyze continuous or time-series data [54]. When extracting features from a fixed-length part of a larger dataset, it is less critical where the feature is placed inside the section and more beneficial to use a 1D CNN. The joint possibility of a waveform x = {x_1, …., x_{t-1}} is factorized as a made of conditional probabilities as follows in equation (3.16) [55].

$$P(X) = \prod_{t=1}^{T} P(x_t | x_1, \ldots\ldots, x_{t-1}) \tag{3.16}$$

The convolutional layer

However, the standard convolution considers future value in the computation. Solved by computing a casual formulation to convolution. That is a formulation in which the present value only depends on the past and present input values. The causality is easily obtained by padding asymmetrically. For a convolutional kernel of size, K adds a padding of K-1 in the past direction.

For K= 3 the pad as

0	0	X (0)	X (1)	X (2)	X (3)

Instead of

0	X (0)	X (1)	X (2)	X (3)	0

As the result, the convolutional kernel will only see the present and past input values. The output at position t for a convolutional kernel of size K can depend on input values up to K-1 steps in the past. The space it sees is called the receptive field. They dilate the convolutional filter of size K by a dilation factor d. The input values up to d(K-1) steps in the past can affect the output at point t.

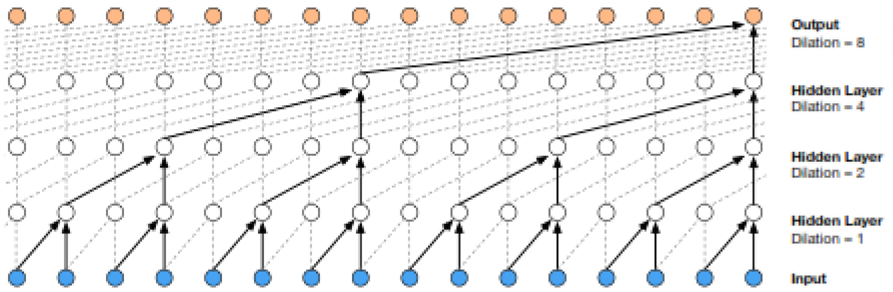

Figure 3.1: Visualization of a Stack of Dilated Casual Convolutional Layer.

Source: https://towardsdatascience.com/how-wavenet-works-12e2420ef386

The residual block is shown in Figure 3.2. The tanh branch is an activation filter or modifier of the dilated convolution that happened just below. It's the "squashing function" it has seen in CNNs before. The sigmoid branch service essentially as a binary gate, and can cancel everything up to it; it learns which data is significant, going back an arbitrary number of periods into the past [34].

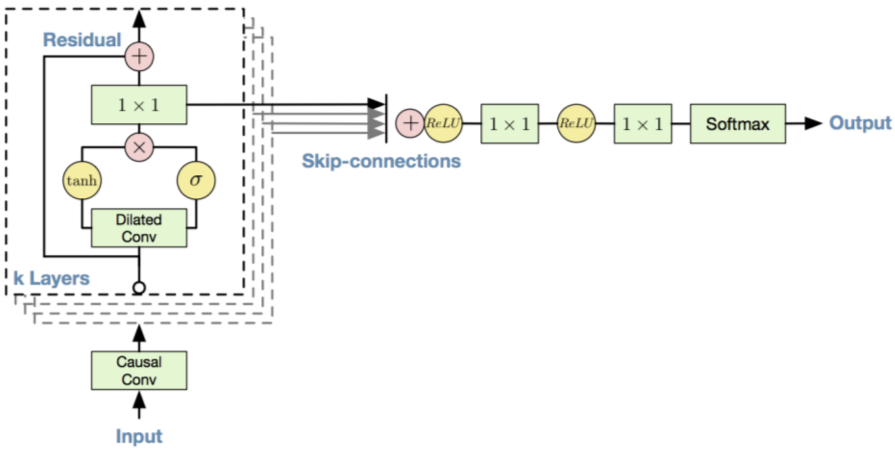

Figure 3.2: Overview of Residual Block and Entire Architecture.

Source: https://towardsdatascience.com/how-wavenet-works-12e2420ef386

3.6 Monophonic Music Analysis

The music player plays a musical instrument one note at a time, which means the single instrument plays at a time. The developed model and the accuracy value are high and easily identify it.

3.7 Polyphonic Music Analysis

Polyphonic music is defined as many musical instruments playing simultaneously at a time. The polyphonic music classification is difficult done, because of what type of facts wishes to be extracted to determine from it. Previous studies reduce the classification issue of polyphonic music to that of monophonic music by changing the information in a few ways [56].

3.7.1 Audio Source Separation

The division of the source signals from a given set of combined signals with no prior knowledge about the source. The separating sounds from musical instruments using Non-negative Matrix Factorization (NMF) [57]. Musical sound is much easier to handle in comparison to speech to analyze the spectral structure of musical signals that remain constant with time. The Vector V is nonnegative then every matrix elements are non-negative. Source separation is a strategy for isolating a specific source a particular source from a sound devoid of any information about the recording situation [58].

3.7.2 Non-Negative Matrix Factorization (NMF)

For a matrix V of dimension m x n where each element $V_{ij} \geq 0$. The V is matrix n number of features and m number of samples of the non-negative matrix. NMF decomposes it into two matrices W and H of dimensions m x r and r x n respectively.

$$V_{mxn} = \begin{bmatrix} V_{11} & \dots\dots\dots\dots\dots V_{1n} \\ \vdots & \dots\dots\dots\dots\dots\dots \\ \vdots & \dots\dots\dots\dots\dots\dots \\ V_{m1} & \dots\dots\dots\dots V_{mn} \end{bmatrix} \tag{3.17}$$

Where $\begin{array}{c} V_{ij} \geq 0 \\ 1 \leq i \leq m \\ 1 \leq j \leq n \end{array}$

Factorization of V is a product of two matrices W and H every matrix is non-negative. W is the fundamental matrix and H is the activation matrix. V represents the power spectrum. Each column and row vector of W and H represent frequency shape and temporal significance alternate for every factorized sound source, respectively. Most spectrograms of musical instrument sounds can be properly factorized since instrument sounds have stationary frequency structures denoted through the pitch and timbre. [59],

$$\left[\begin{bmatrix} V_{11} & \ldots\ldots\ldots\ldots\ldots V_{1n} \\ \vdots \ldots\ldots\ldots\ldots\ldots\ldots & \\ \vdots \ldots\ldots\ldots\ldots\ldots\ldots & \\ V_{m1} \ldots\ldots\ldots\ldots V_{mn} & \end{bmatrix}\right] = \begin{bmatrix} W_{11} & \ldots\ldots\ldots\ldots\ldots W_{1r} \\ \vdots & \ldots\ldots\ldots\ldots\ldots\ldots \\ \vdots & \ldots\ldots\ldots\ldots\ldots\ldots \\ W_{m1} & \ldots\ldots\ldots\ldots W_{mr} \end{bmatrix} \begin{bmatrix} H_{11} \ldots\ldots H_{1n} \\ \vdots \ \ldots\ldots\ldots\ldots \\ H_{r1} \ldots\ldots\ H_{rn} \end{bmatrix} \tag{3.18}$$

$$\begin{array}{ccc} V_{ij} \geq 0 & W_{ij} \geq 0 & H_{ij} \geq 0 \\ 1 \leq i \leq m & 1 \leq i \leq m & 1 \leq i \leq r \\ 1 \leq j \leq n & 1 \leq j \leq r & 1 \leq j \leq n \end{array}$$

3.7.2 Hierarchical Cluster

Hierarchical Clustering is an unsupervised learning process. It identifies groups in the dataset. The similar pattern relationship of the groups' dataset details was represented by the dendrogram tree views and the groups of objects was visualized in hierarchical form, which was a useful representation, and summarization of the data patterns [60].

3.8 Summary

This chapter has elaborated on the low-level features, the mathematical relationship to the audio fundamental audio feature, and the parameters. It expressed the mathematical formula of the low-level audio features. We introduced a sequential forward feature selection method with a Greedy algorithm to select optimum features for musical instrument identification. We described audio source separation, machine learning algorithms, and deep learning approach as 1D Wavenet Technique. We discussed monophonic and polyphonic music analysis.

CHAPTER 4

DATA AND FEATURE FORMATION

INTRODUCTION

The Artificial Intelligence process can be separated into three tasks: data task, training task, and estimate task. In data, tasks are data gathering, data cleaning, and component formation. This task is the most important because relevant data gives a good evaluation result. In the training, the task builds a machine-learning model that used the data feature, in the evaluation task assesses the model to validate the model for future use. This research study conducts on the Oriental Musical Instruments Identification (OMII) process. Figure 4.1 shows the flow diagram of the process and the details of the process are briefly explained below.

4.1 Experimental Framework of Oriental Musical Instruments Identification (OMII) Process

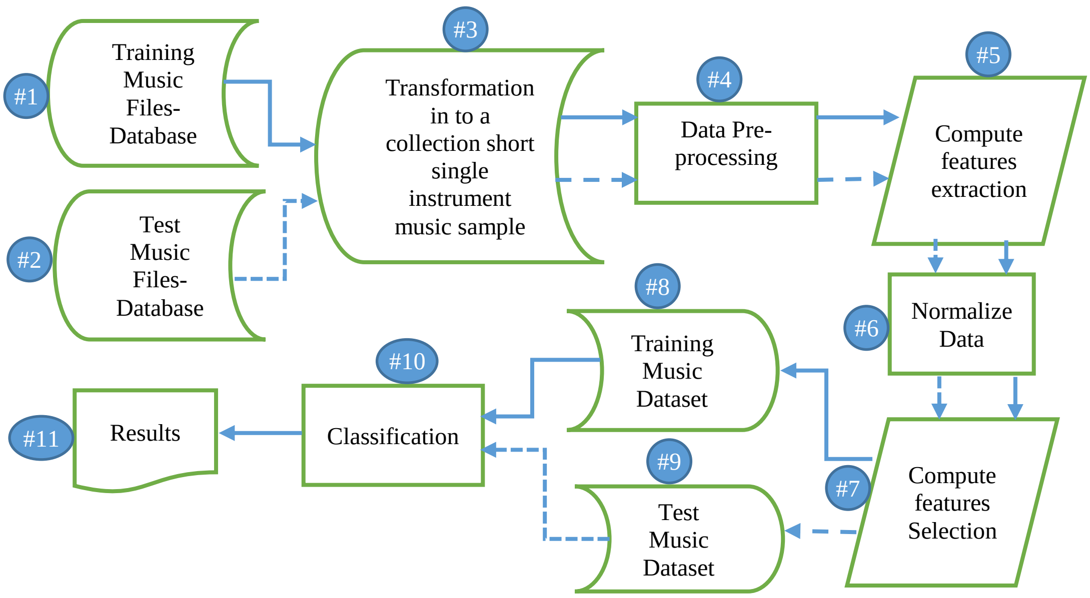

Figure 4.1: Flowchart of the OMII Process.

Figure 4.1 shows the flowchart of the OMII Process as performed in this study. The details of the process are described as follows.

Shape #1 Training Music Files –The database is a pool of musical instrument labeled music files. Which was used to carry out OMII of the unlabeled Test Music files

Database. The Training Music Files –Database contains 20 types of musical instrument monophonic files. The music files got from the website.

Shape #2 Test Music Files- The database consists of unlabeled monophonic and polyphonic music files.

Shape #3 Transformation into a collection of short single instrument music samples – The database consists of their conversion into short sound samples, each containing only sounds of a single musical instrument.

Shape #4 Data Pre-processing – Data Audio files have been treated for silent removal, noise removal, and monophonic conversion before the feature extraction.

Shape #5 Compute feature extraction -The most important characteristics provide from the set of raw data. The raw data transform to best meet the expectation of the learning algorithm.

Shape #6 Normalize Data - It has set up to normalize to range between 0 and 1 there are two normalization methods are available which are Z-score standardization and min-max scaling. In this study, the min-max scaling method has been used to normalize the extracted features.

Shape #7 – Compute the feature selection – Three types of techniques have been used here to select the best set of features. The techniques are random selection, ranking selection, and sequential forward selection.

Shape #8 – Training Music Data – 80% of the data used for training. It also randomly chooses the data to form the dataset.

Shape #9 – Testing Music Data – 20% of the data used for testing from the dataset.

Shape #10 – Classification – The feature vectors predict a set of related categories used to group data according to its similarities.

Shape #11 – Results are the identification of the musical instrument. The model has predicted the musical instrument how much percentage it has to influence the polyphonic musical file.

4.2. Dataset Preparation for Data Analysis:

Data is raw information that represents human and machine observation of the world. The process of gathering data from pertinent data sources, profiling, cleaning, enriching, and integrating it into a single data set for use in analytics is known as data preparation. Profiling understands the data. Cleansing is improving the quality of data. Enriching is adding more data from external data sources.

4.3 Data Task

The music audio files were collected from *"http://www.philharmonia.co.uk/ explore/sound_samples"* website, which includes all standard orchestral instruments. All music audio files have been recorded by a member of the Philharmonia and are free to use [61]. It has many hundreds of music files for each instrument and each instrument's music file has twelve octaves and dynamic phrases. All audio sampling frequencies are 44.1 kHz and more than 1s duration. Data cleaning is correcting or deleting inaccurate records from the database. At this point, some audio files were duplicated find them, and removed. The silent part of the audio should be removed because it is no meaning to the data and it consumes data training and testing time. Therefore, we removed the silent part of the audio sound, reformed it to the same music audio file without silence, and used it for feature extraction. Data formation as defined in the study the analogy signal has converted to a digital signal for the feature extraction.

4.3.1 Feature Extraction

Chapter 2 and Chapter 3 deeply explained the feature definition and details of the characteristics of the features related to the sound parameter in mathematical formation. The Audio Content Analysis Tool extracted 44 features. The signal is divided into 4096 block size and 2048-hop size. Audio Content Analysis - Applications in Signal Processing and Music Informatics by Alexander Lerch [62]. At the beginning of the study, 7 string musical instruments, and 640 music files were used in the experiments. Extracted feature details are shown in Table 4.1.

Table 4.1: The Feature Number Denoted to the Feature Name.

Feature Number	Feature Name
1	Spectral Centroid
2	Spectral Crest Factor
3	Spectral Decrease
4	Spectral Flatness
5	Spectral Flux
6	Spectral Kurtosis
7	Spectral Roll off
8	Spectral Skewness
9	Spectral Slope
10	Spectral Spread
11	Spectral Tonal Power Ratio
12	Time AcfCoeff
13	Time MaxAcf
14	Time Predictivity Ratio
15	Time Rms
16	Time Std
17	Time Zero Crossing Rate
18	Time Peak Envelope 1
19	Time Peak Envelope 2
20	Spectral Mfccs 1
21	Spectral Mfccs 2
22	Spectral Mfccs 3
23	Spectral Mfccs 4
24	Spectral Mfccs 5
25	Spectral Mfccs 6
26	Spectral Mfccs 7
27	Spectral Mfccs 8
28	Spectral Mfccs 9
29	Spectral Mfccs 10
30	Spectral Mfccs 11
31	Spectral Mfccs 12
32	Spectral Mfccs 13
33	Spectral Pitch Chroma 1
34	Spectral Pitch Chroma 2
35	Spectral Pitch Chroma 3
36	Spectral Pitch Chroma 4
37	Spectral Pitch Chroma 5
38	Spectral Pitch Chroma 6

39	Spectral Pitch Chroma 7
40	Spectral Pitch Chroma 8
41	Spectral Pitch Chroma 9
42	Spectral Pitch Chroma 10
43	Spectral Pitch Chroma 11
44	Spectral Pitch Chroma 12

4.4 Summary

This chapter has explained machine learning process tasks. It has presented the proposed Oriental Musical Instruments Identification Process for the classification. The details of the process have been explained here. It explained details of the feature extraction parameters and details of the extracted features.

CHAPTER 5

RESULTS AND DISCUSSION

5.1. Filter Method Used for Feature Selection

The extracted features mean values used for the training process and statistical method to find high correlation features, which were filter methods used for the correlation process and the significant value of the features are shown in Table 5.1 and the higher order of the features, which are shown in Table 5.2. The thirteen features have positively correlated values and the other features have negatively correlated values. The positive features considered for the classification process are shown in Figure 5.1 and the result is shown in Table 5.3.

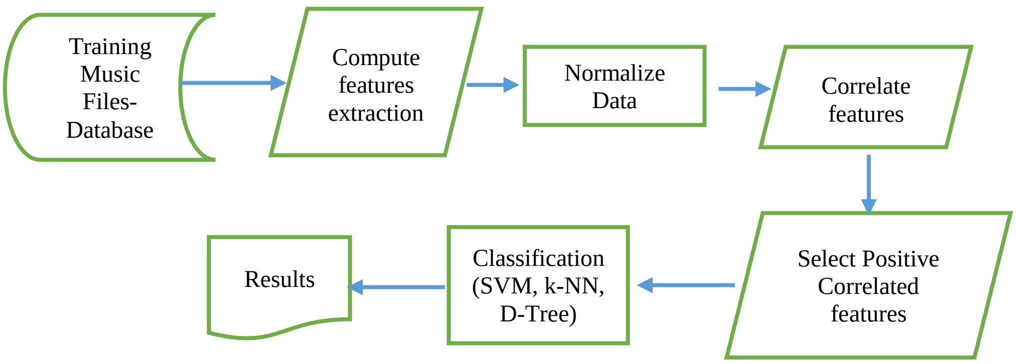

Figure 5.1: Flowchart of Correlated Features with the Classification Process

Table 5.1: Correlated significant value of the Features.

No.	Feature	Sig	Ranking
1	Spectral Centroid	0.435	38
2	Spectral Crest Factor	0.554	5
3	Spectral Decrease	0.494	19
4	Spectral Flatness	0.507	15
5	Spectral Flux	0.459	32
6	Spectral Kurtosis	0.494	19
7	Spectral Roll off	0.527	7
8	Spectral Skewness	0.507	15
9	Spectral Slope	0.465	29
10	Spectral Spread	0.431	40
11	Spectral Tonal Power Ratio	0.459	32
12	Time Acf Coeff	0.525	9
13	Time Max Acf	0.519	10

14	Time Predictivity Ratio	0.560	3
15	Time Rms	0.576	1
16	Time Std	0.559	4
17	Time Zero Crossing Rate	0.445	35
18	Time Peak Envelope1	0.476	25
19	TimePeakEnvelope2	0.570	2
20	SpectralMfccs1	0.516	11
21	SpectralMfccs2	0.462	30
22	SpectralMfccs3	0.493	21
23	SpectralMfccs4	0.506	17
24	SpectralMfccs5	0.473	27
25	SpectralMfccs6	0.527	7
26	SpectralMfccs7	0.445	35
27	SpectralMfccs8	0.451	34
28	SpectralMfccs9	0.493	21
29	SpectralMfccs10	0.485	23
30	SpectralMfccs11	0.461	31
31	SpectralMfccs12	0.514	12
32	SpectralMfccs13	0.509	13
33	SpectralPitchChroma1	0.500	18
34	SpectralPitchChroma2	0.386	44
35	SpectralPitchChroma3	0.431	40
36	SpectralPitchChroma4	0.473	27
37	SpectralPitchChroma5	0.508	14
38	SpectralPitchChroma6	0.474	25
39	SpectralPitchChroma7	0.534	6
40	SpectralPitchChroma8	0.484	24
41	SpectralPitchChroma9	0.421	42
42	SpectralPitchChroma10	0.433	39
43	SpectralPitchChroma11	0.441	37
44	SpectralPitchChroma12	0.404	43

Table 5.2: pValue of the Features in the Higher-Order.

No.	pValue	Feature	Feature Number
1	0.414232992	Spectral Flux	5
2	0.254245545	Time Rms	15
3	0.249135763	Spectral Tonal Power Ratio	11
4	0.230825402	Time Peak Envelope 1	18
5	0.177668196	Time MaxAcf	13
6	0.141349238	Spectral Pitch Chroma 9	41
7	0.128785997	Time Std	16
8	0.067299437	Spectral Mfccs 1	20
9	0.060049836	Spectral Pitch Chroma 10	42
10	0.026072027	Spectral Pitch Chroma 3	35
11	0.016492394	Spectral Skewness	8
12	0.010252254	Spectral Pitch Chroma 12	44
13	0.008490728	Spectral Pitch Chroma 8	40
14	-0.006839962	Spectral Pitch Chroma 11	43
15	-0.013552173	Spectral Decrease	3
16	-0.017510029	Time Peak Envelope 2	19
17	-0.028225595	Spectral Pitch Chroma 2	34
18	-0.035142147	Spectral Roll off	7
19	-0.037082472	Spectral Pitch Chroma 1	33
20	-0.040336948	Spectral Pitch Chroma 5	37
21	-0.050784065	Spectral Pitch Chroma 4	36
22	-0.06832113	Spectral Mfccs 13	32
23	-0.075081133	Spectral Pitch Chroma 7	39
24	-0.089315242	Spectral Mfccs 12	31
25	-0.105668028	Spectral Mfccs 11	30
26	-0.112058474	Spectral Mfccs 8	27
27	-0.119779219	Spectral Pitch Chroma 6	38
28	-0.128971377	Spectral Spread	10
29	-0.147106511	Spectral Crest Factor	2
30	-0.154197666	Time Zero Crossing Rate	17
31	-0.166318258	Spectral Mfccs 9	28
32	-0.167840363	Spectral Mfccs 10	29
33	-0.17281878	Spectral Slope	9
34	-0.174967178	Spectral Flatness	4
35	-0.188409405	Spectral Mfccs 6	25
36	-0.2015627	Spectral Mfccs 7	26
37	-0.240534912	Time AcfCoeff	12
38	-0.241845269	Time Predictivity Ratio	14
39	-0.251112947	Spectral Mfccs 4	23

40	-0.252586483	Spectral Centroid	1
41	-0.299447142	Spectral Kurtosis	6
42	-0.327920319	Spectral Mfccs 3	22
43	-0.394135439	Spectral Mfccs 2	21
44	-0.447278476	Spectral Mfccs 5	24

The number of sets of selected features increased by 5 from 5 to 40 in this process. The set of selected feature details is shown in Table 5.5. The set of selected features used for the three classification methods to identify the musical instruments and the results of the accuracy is shown in Table 5.3. According to the result, all features have the highest accuracy value of the SVM classifier but the 20 best sets of selected features have higher but 1.7% less than all features used in classification. In this experiment, the SVM classifier is the best classifier among others.

Table 5.3: The Classification Result of All Classifiers with Set of the Best Selected Features from Correlation Methods

Features Selection Type	SVM	k-NN	D-Tree
All Features (44)	**93.1**	**76.7**	**78.91**
Correlated Features (13)	73.0	67.7	64.2
5 Best Features	76.9	75.6	70.0
10 Best Features	85.6	80.0	72.8
15 Best Features	88.9	84.4	78.6
20 Best Features	91.4	85.3	77.7
25 Best Features	90.9	83.3	75.9
30 Best Features	90.5	78.0	76.7

Table 5.4: The Confusion Matric for String Musical Instruments.

	Banjo	Cello	Double Base	Guitar	Mandolin	Viola	Violin
Banjo	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Cello	0.0%	**88.2%**	0.0%	0.0%	0.0%	5.9%	5.9%

Double Base	0.0%	0.0%	**87.5%**	0.0%	0.0%	0.0%	12.5%
Guitar	0.0%	0.0%	0.0%	**83.3%**	0.0%	0.0%	16.7%
Mandolin	0.0%	0.0%	0.0%	0.0%	**94.7%**	0.0%	5.3%
Viola	0.0%	4.8%	0.0%	0.0%	0.0%	**71.4%**	23.8%
Violin	0.0%	0.0%	0.0%	9.5%	0.0%	4.8%	**85.7%**

Table 5.5: The Filter Method Find the Best Set of Selected Features

No of the Best Features	***Spectral Centroid***	***Spectral Crest Factor***	***Spectral Decrease***	***Spectral Flatness***	***Spectral Flux***	***Spectral Kurtosis***	***Spectral Rolloff***	***Spectral Skewness***	***Spectral Slope***	***Spectral Spread***	***Spectral Tonal Power***	***Time AcfCoeff***	***Time MaxAcf***	***Time Predictivity Ratio***	***Time Rms***	***Time Std***	***Time Zero Crossing Rate***	***Time Peak Envelope Min***	***Time Peak Envelope Max***	***Spectral Mfccs 1***	***Spectral Mfccs 2***	***Spectral Mfccs 3***	***Spectral Mfccs 4***	***Spectral Mfccs 5***	***Spectral Mfccs 6***	***Spectral Mfccs 7***	***Spectral Mfccs 8***	***Spectral Mfccs 9***	***Spectral Mfccs 10***	***Spectral Mfccs 11***	***Spectral Mfccs 12***	***Spectral Mfccs 13***	***Spectral Pitch Chroma 1***	***Spectral Pitch Chroma 2***	***Spectral Pitch Chroma 3***	***Spectral Pitch Chroma 4***	***Spectral Pitch Chroma 5***	***Spectral Pitch Chroma 6***	***Spectral Pitch Chroma 7***	***Spectral Pitch Chroma 8***	***Spectral Pitch Chroma 9***	***Spectral Pitch Chroma 10***	***Spectral Pitch Chroma 11***	***Spectral Pitch Chroma 12***
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
15	0	1	1	1	1	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
20	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
25	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	0	1	0
30	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	0
35	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Note: 0 indicated the absence of the feature and 1 indicated the presence of the feature.

5.2 Wrapper Method Used for Sequential Forward Features Selection (SFFS)

Using the SFFS methodology, the wrapper method identified the best combination of selected characteristics with the highest accuracy value. The experiment saw a 1 rise in the number of characteristics from 2 to 44. Each of the top sets of chosen features had the experiments carried out ten times, and for each of those times, the best set's accuracy value was recorded. The recorded features are shown in Table 5.6. When the number of the best feature was increased from two to eleven at the time the accuracies gradually increased from 63.57% to 87.5%.

Table 5.6: The SFFS Method Find the Set of Best Features and Identification Accuracy.

No of Features	Accuracy	Spectral Centroid	Spectral Crest Factor	Spectral Decrease	Spectral Flatness	Spectral Flux	Spectral Kurtosis	Spectral Roll off	Spectral Skewness	Spectral Slope	Spectral Spread	Spectral Tonal Power Ratio	Time AcfCoeff	Time MaxAcf	Time Predictivity Ratio	Time Rms	Time Std	Time Zero Crossing Rate	Time Peak Envelope Max	Time Peak Envelope Min	Spectral Mfccs1	Spectral Mfccs2	Spectral Mfccs3	Spectral Mfccs4	Spectral Mfccs5	Spectral Mfccs6	Spectral Mfccs7	Spect	Spectral Mfccs9	Spectral Mfccs10	Spectral Mfccs11	Spectral Mfccs12	Spectral Mfccs13	Spectral Pitch Chroma 1	Spectral Pitch Chroma 2	Spectral Pitch Chroma 3	Spectral Pitch Chroma 4	Spectral Pitch Chroma 5	Spectral Pitch Chroma 6	Spectral Pitch Chroma 7	Spectral Pitch Chroma 8	Spectral Pitch Chroma 9	Spectral Pitch Chroma 10	Spectral Pitch Chroma 11	Spectral Pitch Chroma 12
2	63.6%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	71.3%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	74.4%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	72.9%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	80.6%	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	81.4%	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	78.1%	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	83.6%	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	85.9%	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	87.5%	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	84.4%	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13	89.8%	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
14	84.4%	0	1	1	1	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15	90.6%	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
16	86.7%	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0
17	**91.4%**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

18	85.9%	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
19	89.1%	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
20	85.2%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0
21	78.1%	0	1	1	0	1	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
22	77.3%	0	1	1	0	1	0	0	0	1	0	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	1	1	1	0	0
23	82.0%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
24	77.3%	0	1	1	1	1	0	0	0	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0
25	83.6%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0
26	80.5%	0	1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	1	0
27	76.6%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	1	0	1	0	1	1	0	0
28	88.3%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1	0	1	1
29	82.1%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	1	1	1	0	1	0	1	1
30	88.3%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0	1	1
31	76.6%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	0	1
32	87.5%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1
33	75.0%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
34	74.2%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
35	80.5%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
36	74.2%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
37	86.7%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38	79.7%	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
39	80.5%	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	69.5%	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

41	22.7%	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
42	16.0%	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
43	15.6%	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
44	17.0%	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Note: 0 indicated the absence of the feature and 1 indicated the presence of the feature.

Table 5.7 is shown selected best features for filter method and wrapper method. The most of the time domain feature are selected by filter method and the most of the spectral features are selected by wrapper method.

Table 5.7: The Selected Features for Filter Method and Wrapper Method.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
Feature	Spectral Centroid	Spectral Crest Factor	Spectral Decrease	Spectral Flatness	Spectral Flux	Spectral Kurtosis	Spectral Roll off	Spectral Skewness	Spectral Slope	Spectral Spread	Spectral Tonal Power Ratio	Time Acf Coeff	Time Max Acf	Time Predictivity Ratio	Time Rms	Time Std	Time Zero Crossing Rate	Time Peak Envelope1	Time Peak Envelope2	Spectral Mfccs1	Spectral Mfccs2	Spectral Mfccs3	Spectral Mfccs4	Spectral Mfccs5	Spectral Mfccs6	Spectral Mfccs7	Spectral Mfccs8	Spectral Mfccs9	Spectral Mfccs10	Spectral Mfccs11	Spectral Mfccs12	Spectral Mfccs13	Spectral Pitch Chroma1	Spectral Pitch Chroma2	Spectral Pitch Chroma3	Spectral Pitch Chroma4	Spectral Pitch Chroma5	Spectral Pitch Chroma6	Spectral Pitch Chroma7	Spectral Pitch Chroma8	Spectral Pitch Chroma9	Spectral Pitch Chroma10	Spectral Pitch Chroma11	Spectral Pitch Chroma12
Filter Method	0	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0
Wrapper method	1	0	0	1	1	1	1	0	1	1	1	1	0	1	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note: 0 indicated the absence of the feature and 1 indicated the presence of the feature.

Figure 5.2 displays the experiment outcome of the number of the best set of selected features against accuracy. The accuracy value started from 63.6% at the number of the best set of selected features was 2. The accuracy values gradually increased by 91.4% at the number of the best set of selected features was 17. Then it fluctuated a little below the number of the best set of selected features was 39. After it suddenly decreased from 86.7% to 17.0% the number of the best set of selected features was 37 and 44. The maximum accuracy value is 91.43% when the number of the best-selected features has been 17 and that point is denoted by the shaded circle in Figure 5.1. The best-selected feature particulars are displayed in Table 5.5 and bold text in row 17 are indicated the features, which are selected as one and not selected as zero.

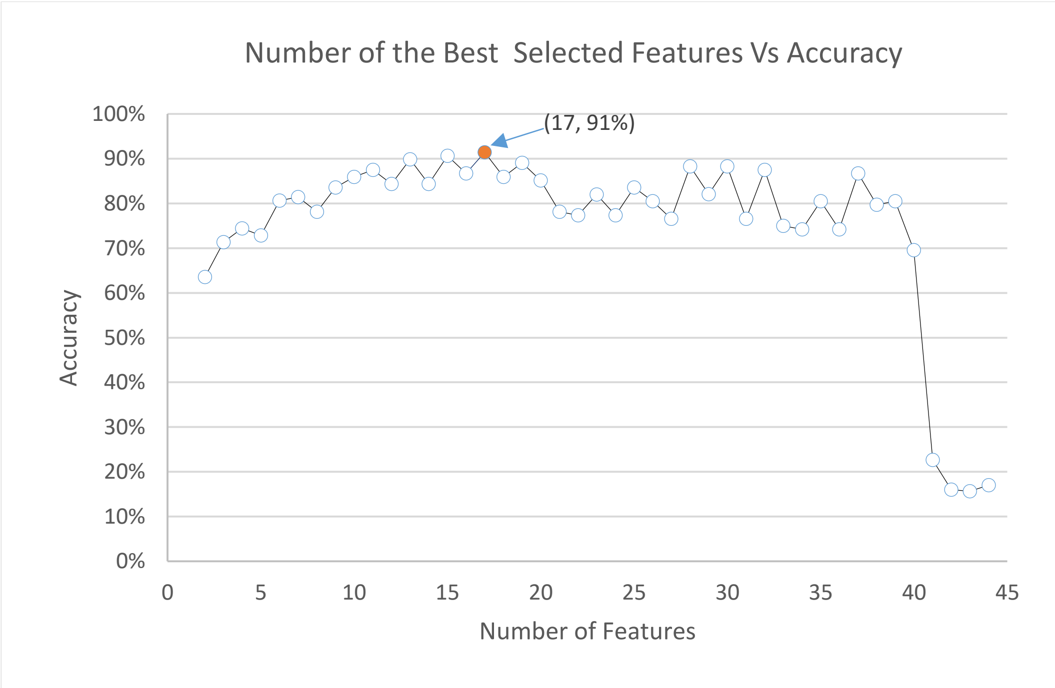

Figure 5.2: Accuracy of the Number of the Best-Selected Features

The selected best features are used to the three classifiers. The accuracy value are shown in Table 5.8. SVM classifier has the highest accuracy value in both selection methods but wrapper method has more than filter method. Therefore wrapper method is selected for the feature selection method for this study.

Table 5.8: Accuracy Comparison of the Feature Selection Methods

Classification Type	Filter Method	Wrapper Method
SVM	91.7%	93.37%
KNN	86.4%	91.87%
DT	86.5%	90.57%

5.3 Musical Instrument Classification

The three classifiers used in this task, were Decision Tree, kNN, and SVM. The 17 best-selected features have utilized the classification. The confusion matrix got from each classification for evaluation to the performance of the classifiers.

A confusion matrix, one of the natural matrices for assessing the accuracy and correctness of the classifier, was used to offer a comprehensive picture of the results. The values in this matrix represented the proportions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [63].

Using the aforementioned parameters yields the following equations.

Accuracy: The accuracy evaluates whether the model is trained correctly and it may be performed generally. It is defined as the proportion of the quantity of accurately classified results to the aggregate number of classified results. Equation 5.1 is derived as the accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5.1)$$

Precision: It is defined as the ratio of the number of true positives to all the positives. Equation 5.2 is derived as the precision.

$$Precision = \frac{TP}{TP+FP} \qquad (5.2)$$

Sensitivity (Recall): It measures our model to correctly identify true positives. It is outlined because the number of correct positive predictions is divided by the whole number of positives.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (5.3)$$

Specificity: It is outlined because the number of correct negative predictions is divided by the whole number of negatives.

$$Specificity = \frac{TN}{TN+FP} \qquad (5.4)$$

Error Rate: It is used to validate the performance of the model. It is defined as one minus accuracy.

$$Error\ Rate = 1 - Accuracy \qquad (5.5)$$

Table 5.9: Recognition Measures for Classifier

Musical Instrument Family	Musical Instrument	Classifier														
		SVM					kNN					DT				
		Accuracy	Precision	Sensitivity	Specificity	F-Score	Accuracy	Precision	Sensitivity	Specificity	F-Score	Accuracy	Precision	Sensitivity	Specificity	F-Score
String	Banjo	**0.99**	1.00	0.97	1.00	0.99	0.99	0.99	1.00	0.99	0.99	0.99	1.00	0.97	1.00	0.99
	Cello	**0.93**	0.93	0.93	0.93	0.93	0.88	0.90	0.86	0.90	0.88	0.87	0.88	0.87	0.88	0.87
	Double Bass	**0.97**	0.97	0.97	0.97	0.97	0.94	0.93	0.96	0.93	0.94	0.95	0.94	0.95	0.94	0.95
	Guitar	**0.87**	0.98	0.76	0.99	0.85	0.79	0.89	0.66	0.92	0.76	0.78	0.97	0.58	0.98	0.73
	Mandolin	0.95	0.93	0.96	0.93	0.95	0.96	0.97	0.95	0.97	0.96	0.97	0.96	0.98	0.96	0.97
	Tambura	0.57	0.59	0.44	0.70	0.50	0.78	1.00	0.56	1.00	0.72	0.75	0.86	0.60	0.90	0.71
	Viola	**0.94**	0.93	0.94	0.93	0.94	0.88	0.86	0.90	0.85	0.88	0.87	0.86	0.88	0.86	0.87
	Violin	**0.95**	0.95	0.96	0.94	0.95	0.91	0.91	0.91	0.91	0.91	0.92	0.91	0.93	0.91	0.92
Brass	Bass Clarinet	**0.97**	0.97	0.98	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.95
	French Horn	**0.96**	0.96	0.95	0.96	0.96	0.93	0.94	0.92	0.94	0.93	0.90	0.91	0.89	0.91	0.90
	Trombone	**0.96**	0.96	0.97	0.96	0.96	0.95	0.94	0.96	0.93	0.95	0.95	0.93	0.96	0.93	0.95
	Trumpet	**0.93**	0.93	0.93	0.93	0.93	0.88	0.89	0.87	0.89	0.88	0.87	0.93	0.80	0.94	0.86
	Tuba	**1.00**	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.99	0.99	0.99	0.99
Wood Wind	Bassoon	**0.96**	0.96	0.96	0.96	0.96	0.95	0.95	0.94	0.95	0.95	0.91	0.91	0.90	0.92	0.91
	Clarinet	**0.93**	0.93	0.92	0.93	0.93	0.91	0.90	0.91	0.90	0.91	0.88	0.88	0.88	0.88	0.88
	Contra Bassoon	**0.97**	0.97	0.97	0.97	0.97	0.94	0.92	0.96	0.92	0.94	0.91	0.91	0.92	0.91	0.91
	English Horn	**0.96**	0.96	0.97	0.96	0.96	0.94	0.94	0.95	0.93	0.94	0.91	0.90	0.92	0.90	0.91
	Flute	**0.99**	0.99	0.99	0.99	0.99	0.98	0.97	0.98	0.97	0.98	0.97	0.95	0.98	0.95	0.97
	Oboe	**0.96**	0.96	0.96	0.96	0.96	0.95	0.94	0.95	0.94	0.95	0.93	0.91	0.94	0.91	0.93

	Saxophone	**0.92**	0.92	0.92	0.92	0.92	0.87	0.88	0.85	0.89	0.87	0.84	0.86	0.82	0.86	0.84
	Overall	**0.93**	**0.94**	**0.92**	**0.95**	**0.93**	0.92	0.93	0.90	0.94	0.91	0.91	0.92	0.89	0.92	0.90

5.4 Validation of Model Performance

There are many statistical validation techniques that can be found in older works. The performance of the model is validated in this work using the error rating and area under the receiver operating characteristic curve (AUROC). One of the most crucial assessment matrices for assessing the effectiveness of any classification model is the AUC and ROC (receiver operating characteristics) curve, which may be measured up to multiclass classification problems. The AUC value ranges from 0 to 1. Equation (5.5) was used to calculate the error rating value. The outcomes are displayed in Figure 5.3. The SVM classifier has a minimal error rate excluding the musical instrument is tambura.

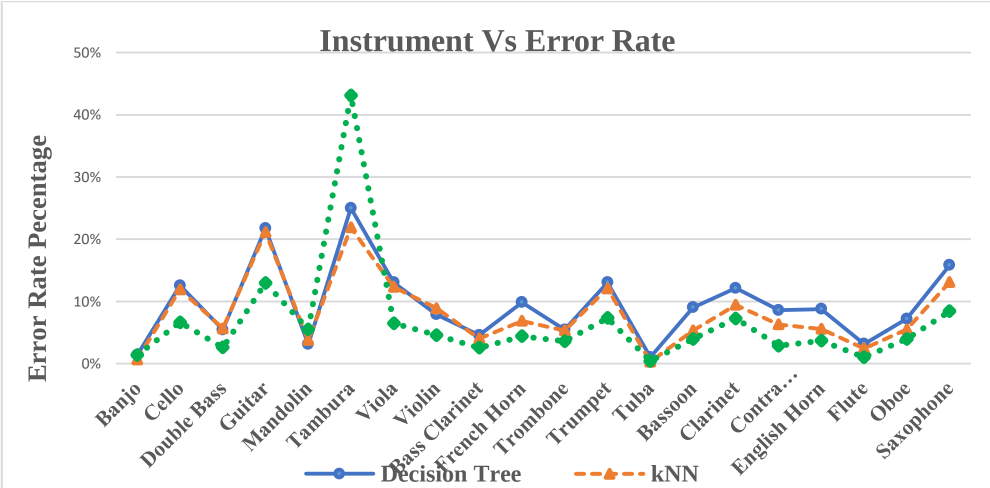

Figure 5.3: The Error Rate of the Musical Instruments

The values of the AUC matrix are derived from the Receiver Operating Characteristics (ROC) curve in the classification. The AUC rating of the musical instruments is displayed in Figure 5.4. The maximum AUC scores for the SVM classifier in this calculation are also 1, with a few values being 0.99 or 0.98. In the end, the validation findings are decided. Of the three, the SVM classifier performs the best.

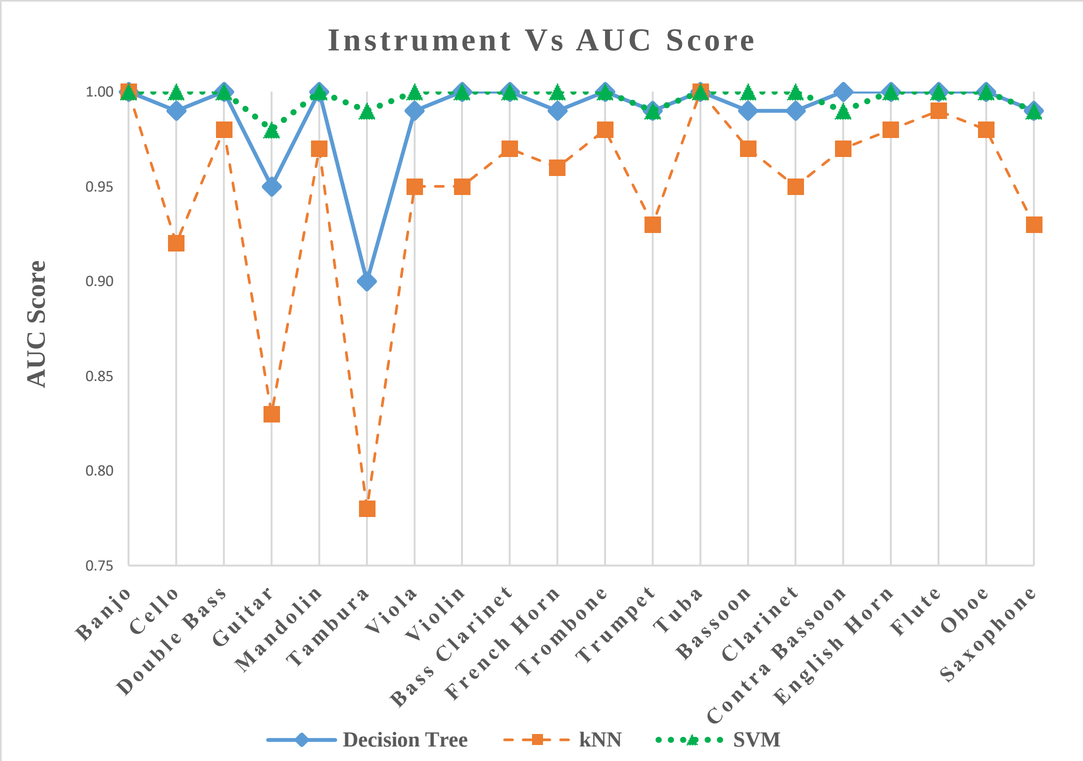

Figure 5.4: The AUC Score of the Musical Instruments

5.5 Effect of Analysis Window Size

The research study analyzed for window size is independent of the feature extraction process. The signal was divided into different window sizes exactly 2048, 4096, 6144, 8192, 10240, and 12228 used those widow sizes to extract the 44 features. The best-selected features were used for SVM classification to get accuracy. The result of the accuracy values are not significantly different from window size 4096 to window size 12288, therefore the feature extraction process does not dependent on the window size of the music audio file. As the result of the classification, 4096 widow size is selected to use in the feature extraction process because it has the highest accuracy value among others. The accuracy of the classification results shows in Table 5.10 and more details of the experiment results are shown in Appendix C.

Table 5.10: The Accuracy of the Overall Musical Instrument Identification with Window Size

Window Size	Accuracy%
2048	94.52
4096	**95.46**
6144	95.15
8192	95.32
10240	95.45
12288	95.33

5.6 Musical Instrument and Family Identification

In this experiment, the musical instrument family was identified first, and after that found individual musical instruments. Four SVM classifiers used for experiment. One for finding the musical instrument family (family model) at these points 17 features used to identify musical instrument family. The details of the features are shown in Figure 5.4. The model has identified musical instrument families' accuracy values were 96.09%, 95.13%, and 95.43% respectively string, brass, and woodwind. After that, the model has been decided to use another suitable classifier for individual musical instrument identification, which the musical instrument family was the string family when the system chooses the string SVM model for identifying any one string instrument that is in the training process. Figure 5.5 shows the details of the musical instrument identification hierarchy. The same process has been done for another two instruments families and identified individual instruments. Table 5.11, Table 5.12, Table 5.13, and Table 5.14 are shown the accuracy of the musical instrument family, brass musical instrument, string musical instrument, and woodwind musical instrument respectively.

The accuracy value of individual musical instrument identification in the Brass family is higher than that of other musical instruments family.

Table 5.11: The Musical Instrument Family Recognition Measure the SVM Classifier

Family	Accuracy	Precision	Sensitivity	Specificity
String	0.9609	0.9641	0.9574	0.9644
Brass	0.9513	0.9523	0.9501	0.9524
Woodwind	0.9545	0.9513	0.9580	0.9509
Average	**0.9555**			

Table 5.12: The Brass Musical Instrument Recognition Measure the SVM Classifier

Musical Instrument	Accuracy	Precision	Sensitivity	Specificity
Bass Clarinet	0.9870	0.9842	0.9900	0.9841
French Horn	0.9910	0.9920	0.9900	0.9920
Trombone	0.9672	0.9627	0.9720	0.9624
Trumpet	0.9762	0.9791	0.9732	0.9793
Tuba	0.9960	1.0000	0.9920	1.0000
Average	**0.9835**			

Table 5.13: The String Musical Instrument Recognition Measure the SVM Classifier

Musical Instrument	Accuracy	Precision	Sensitivity	Specificity
Banjo	0.9291	0.9045	0.9595	0.8987
Cello	0.9558	0.9611	0.9500	0.9615
Double Bass	0.9710	0.9701	0.9720	0.9701
Guitar	0.8980	0.9690	0.8222	0.9737
Mandolin	0.9383	0.9383	0.9383	0.9383
Tambura	0.9789	0.9837	0.9740	0.9838
Viola	0.9440	0.9320	0.9580	0.9301
Violin	0.9585	0.9551	0.9623	0.9547

Table 5.14: The Woodwind Musical Instrument Recognition Measure the SVM Classifier

Musical Instrument	Accuracy	Precision	Sensitivity	Specificity
Bassoon	0.9729	0.9794	0.9660	0.9797
Clarinet	0.9236	0.8994	0.9540	0.8933
Contra Bassoon	0.9960	0.9960	0.9960	0.9960
English Horn	0.9741	0.9705	0.9780	0.9702
Flute	0.9891	0.9843	0.9940	0.9842
Oboe	0.9769	0.9798	0.9740	0.9799
Saxophone	0.9373	0.9655	0.9070	0.9676
Average	**0.9671**			

5.6.1 Musical Instrument and Family Identification Hierarchy

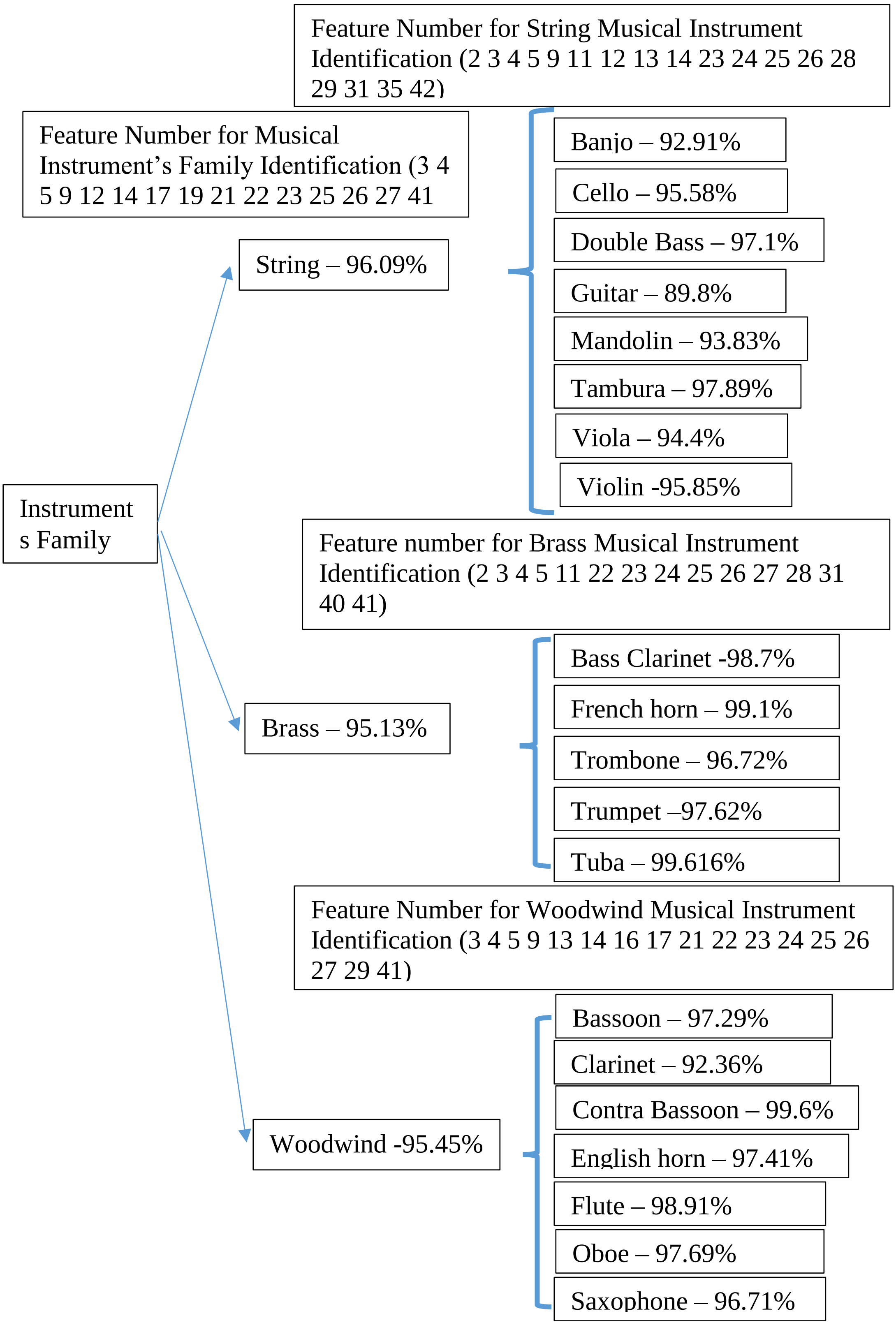

Figure 5.5: Hierarchy View of the Musical Instrument Identification

5.7 Deep Learning Approach to Find the Instrument Identification

WaveNet is a powerful new predictive technique that uses multiple Deep Learning (DL) strategies from Computer Vision (CV) and Audio Signal Processing models and applies them to longitudinal (time-series) data. This same process can be applied to one-dimensional sequences of data. The model extracts features from sequences data and maps the internal features of the sequence. A 1D CNN is very effective for deriving features from a fixed-length segment of the overall dataset, where it is not so important where the feature is located in the segment.

5.7.1 Fine-Tuning to Wavenet model

Set to the hyper parameters are kernel size, number of filters, dilation depth, and batch size. In this experiment set, the kernel size is 2, and the number of filters is 45 at the time accuracy was the highest value. The dilation depth is 9 and the batch size is 50. The accuracy values are shown in Table 5.15

Table 5.15: Details of Fine Tuning to Wavenet Model

No Filters	Accuracy	Validate Accuracy	Loss	Validate Loss	Epoch
5	0.8132	0.7643	0.5405	0.7474	92
10	0.8989	0.8489	0.2829	0.5081	57
15	0.9314	0.8674	0.2121	0.4527	33
20	0.9021	0.8496	0.2852	0.511	23
25	0.9012	0.8352	0.3021	0.5463	19
30	0.9253	0.8559	0.2193	0.8559	23
35	0.9102	0.8534	0.2914	0.7485	24
40	0.9085	0.8822	0.2822	0.4004	18
45	**0.9676**	**0.9003**	**0.099**	**0.4572**	**30**
50	0.9471	0.877	0.1571	0.5218	24

The figure 5.6 is shown error for music instrument identification process with epoch. The lowest value of validation loss is 0.48% and train loss is 0.15% at 29 epoch.

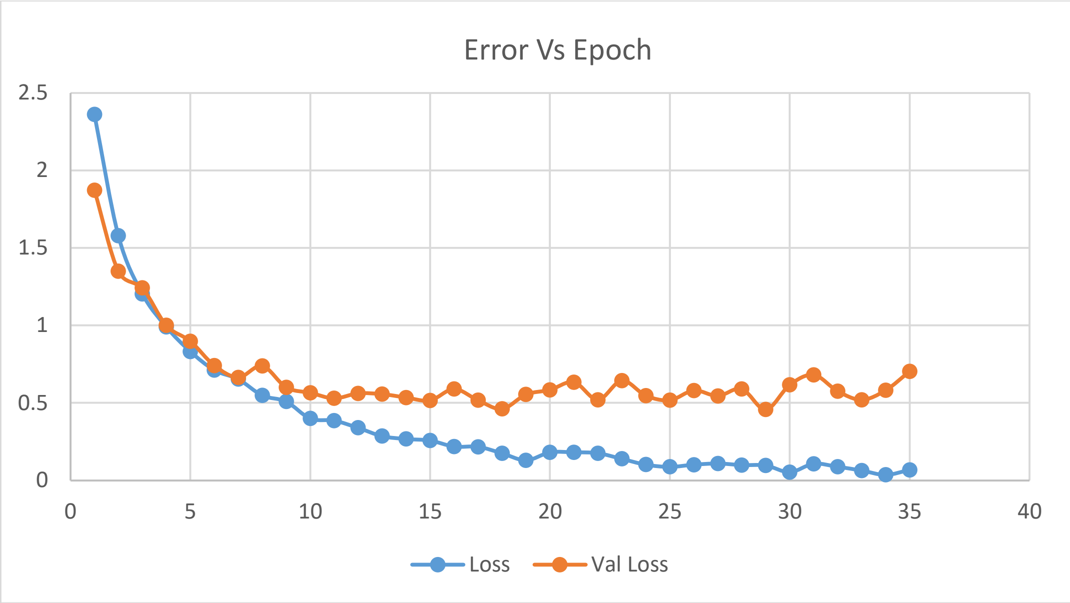

Figure 5.6: Wavenet Model Used to Find the Error Value for Musical Instrument Identification

Figure 5.7 is shown accuracy verses epoch details. The highest validated accuracy value is 90.24% and normal train accuracy value is 98.25% at 29 epoch.

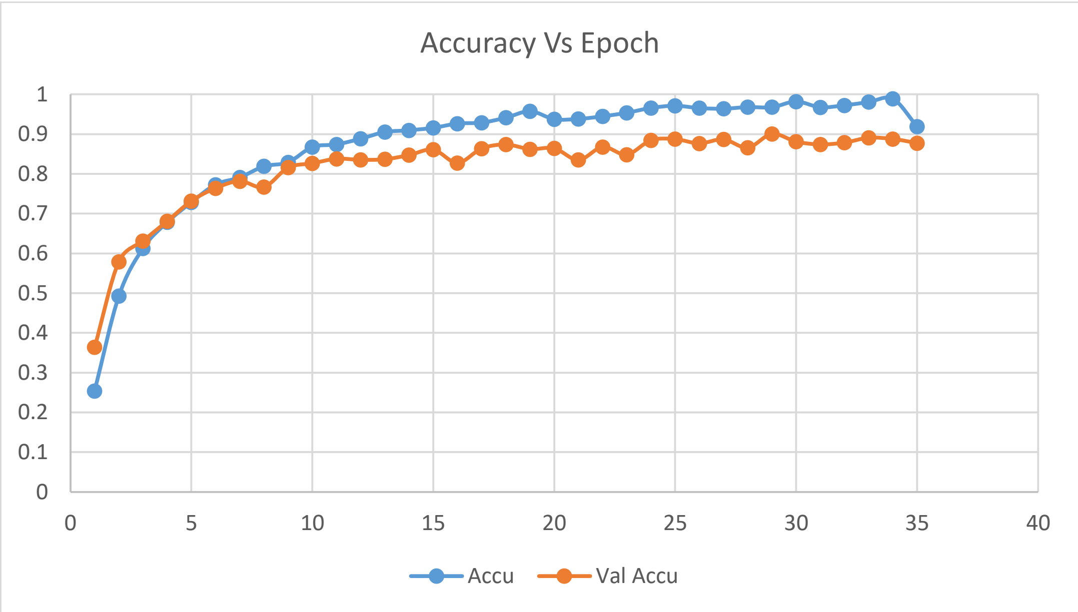

Figure 5.7: Wavenet Model Used to Find the Accuracy Value for Musical Instrument Identification

5.8 Individual Musical Instrument Identification

5.8.1 Find Predominant Features for Musical Instrument Identification

The previous experiments were done for the selection of the best set of features for the musical instrument identification among the group of musical instruments. However, the predominant feature selection for a specific set of features belongs to each musical instrument. The predominant features were selected by three different feature selection techniques, which were ranking selection, random selection, and sequential forward feature selection. The predominant features selection and classification process is shown in Figure 5.8. In this order, the higher-rank features were selected by the ranking selection technique. The random feature selection technique carries out a randomized subset feature search supported by classification. It randomly generates subsets of features utilized to classify samples. Each subset is appraised with the apparent error. Only the best subsets are retained and, they are amalgamated into a single pool. The most important for every feature in the pool gives the measurement of the significance. The random features were selected by the random selection technique. However, the best optimized features were chosen using the sequential forward feature selection method. Those details are shown in Table 5.16.

The predominant features have selected by the SFFS (Sequential Forward Feature Selection) method and the numbers of predominant features were dependent on the musical instrument. Despite that, the number of predominant features detected by the other two selection techniques is taken to be equal to the number of predominant features to be detected in the SFFS method. Using them in the SVM classifier, the musical instruments were identified and the accuracy values were tabulated respectively. The result of the classification is shown in Table 5.17. The accuracy of the SFFS technique from the response received is higher than those other two techniques. Therefore, the SVM model classified the individual musical instruments with the features selected in the SFFS method. The automated system has trained SVM models for musical instrument identification. Therefore, the selected trained SVM models have been used in the automated system for the classification of the individual musical instrument identification in polyphonic music.

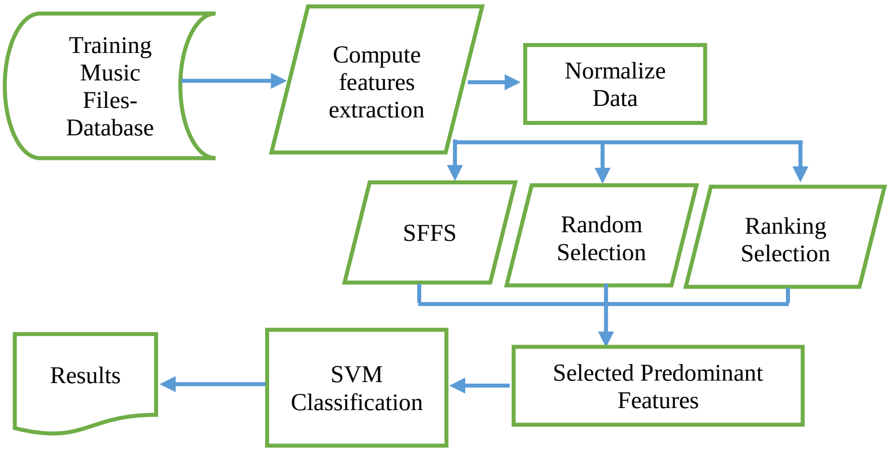

Figure 5.8: Flowchart of Predominant Feature Selection with the Classification Process

Table 5.16: Details of Predominant Features Selection.

Class No.	Musical Instrument	Method	Feature Number																								
1	Banjo	Ranking	9	13	16	14	5	4	19	20	18	11	15	7	1	10	17	6	8	22	21	2	12	3	29	40	41
		Random	7	36	41	29	4	40	10	43	6	17	18	37	38	5	12	20	23	34	25	30	1	13	14	22	26
		SFFS	15	23	22	3	6	21	26	44																	
2	Bass Clarinet	Ranking	19	13	3	16	11	2	8	22	24	20	7	17	34	4	39	10	21	35	26	32	14	9	12	29	30
		Random	15	2	28	35	17	8	41	16	23	24	32	40	10	11	27	30	34	3	13	18	25	26	29	36	43
		SFFS	39	14	19	10	13	7	21	2	3	16															
3	Bassoon	Ranking	10	7	21	14	19	17	1	4	28	29	22	26	23	18	9	24	16	13	15	11	20	25	12	3	32
		Random	27	42	6	28	31	1	12	20	26	36	37	43	7	8	9	11	17	34	35	2	4	10	19	25	30
		SFFS	1	2	3	4	5	7	15	22	29	24	20														
4	Cello	Ranking	23	24	1	10	17	6	13	41	22	14	27	21	26	7	3	39	12	20	11	40	15	38	18	25	29
		Random	2	30	4	24	25	31	37	3	11	14	16	21	23	6	12	15	20	29	40	41	10	27	28	36	39
		SFFS	1	2	3	4	5	6	21	22	23	24	27	44													
5	Clarinet	Ranking	13	4	19	11	3	16	22	14	21	15	40	8	2	18	12	7	6	39	26	10	23	9	41	38	24
		Random	7	9	10	1	14	28	35	4	8	19	27	38	11	13	17	22	24	26	32	34	42	43	2	6	12
		SFFS	24	29	16	4	12	26	21	44	33	13	20	35													
6	Contra Bassoon	Ranking	22	10	17	12	1	21	24	7	14	23	29	44	5	25	28	4	41	13	3	20	26	19	31	9	2
		Random	5	21	22	9	32	40	4	6	12	17	26	30	39	1	11	13	23	25	28	29	43	2	7	15	16
		SFFS	1	2	29	22	26	11	44	12	21	4	10														
7	Double Bass	Ranking	24	12	17	1	10	25	26	7	22	14	4	29	44	21	23	28	8	27	3	30	2	11	15	43	37
		Random	3	8	37	39	40	43	7	10	18	23	26	27	29	13	14	19	20	25	28	34	42	1	2	6	9
		SFFS	17	21	15	40	23	2	4	26	42	20	1	9													
8	English Horn	Ranking	12	24	23	5	19	6	22	40	20	15	21	16	14	26	27	13	3	18	10	1	11	8	25	33	29
		Random	39	5	11	37	3	10	31	33	16	36	40	1	4	13	14	18	19	20	29	2	7	9	12	22	25

		SFFS	1	2	3	4	5	19	13	14	22	15															
9	Flute	Ranking	6	22	21	4	5	8	7	39	27	15	41	18	9	2	24	23	40	13	38	11	28	16	30	19	10
		Random	20	24	16	27	29	42	44	5	18	22	31	32	43	7	12	14	28	30	35	40	1	2	8	10	11
		SFFS	1	7	5	27	6	22	31	21	39	36	37	15													
10	French Horn	Ranking	17	1	21	10	14	25	23	22	7	18	27	12	28	15	19	11	35	26	5	34	4	44	41	29	40
		Random	15	22	28	43	3	20	12	14	25	31	35	8	10	11	19	24	30	34	39	40	1	13	16	18	21
		SFFS	1	2	3	25	23	27	17	24	9	20	7														
11	Guitar	Ranking	18	15	4	20	9	3	6	7	23	22	14	5	1	19	44	42	11	16	43	25	12	24	35	10	2
		Random	13	34	24	6	21	23	31	20	41	42	2	4	7	14	16	17	26	28	30	32	33	43	8	9	15
		SFFS	3	4	9	13	19	22	6	39	2	29															
12	Mandolin	Ranking	16	14	13	6	19	9	7	4	15	18	10	11	5	21	1	17	22	23	20	8	2	29	31	24	32
		Random	5	24	30	17	31	1	8	10	16	19	20	35	41	7	11	12	13	14	15	22	26	44	4	9	27
		SFFS	19	13	10	2	12	9	21	5	25	6	36														
13	Oboe	Ranking	22	12	13	5	14	21	18	15	4	23	17	1	7	10	6	16	2	11	33	24	9	8	25	28	41
		Random	15	40	1	2	5	13	3	9	10	24	30	34	6	7	8	17	19	26	27	29	33	37	41	11	14
		SFFS	40	26	21	32	4	41	24	22	6	7	27	25													
14	Saxophone	Ranking	6	13	9	21	22	16	24	19	12	3	5	15	18	41	40	39	10	26	32	2	8	17	1	4	29
		Random	34	2	7	8	11	14	24	29	36	3	4	15	19	27	6	17	28	42	44	1	12	21	22	23	30
		SFFS	1	2	3	4	16	11	32	19	26	21															
15	Tambura	Ranking	20	9	18	2	6	15	25	8	7	24	11	10	42	4	36	17	23	40	3	14	26	13	29	39	27
		Random	29	19	1	10	25	5	8	15	20	24	35	39	44	3	4	13	14	17	21	23	32	42	6	9	11
		SFFS	1	2	15	23	32	27	19	5	21	7	12														
16	Trombone	Ranking	15	16	18	3	11	27	23	4	5	9	2	7	21	1	26	8	12	17	22	24	28	32	39	6	29
		Random	3	28	39	4	14	26	31	40	11	5	13	21	23	25	29	32	35	41	43	7	9	12	24	34	38
		SFFS	1	2	4	6	19	23	7	32	33	39	13	27													
17	Trumpet	Ranking	22	19	24	3	12	15	18	11	14	25	16	27	21	9	6	30	32	7	23	29	31	41	37	20	42

		Random	42	38	41	19	25	1	12	17	18	21	6	11	16	28	33	35	44	3	9	10	13	14	22	31	39
		SFFS	1	2	3	4	19	22	21	13	43	5															
18	Tuba	Ranking	21	12	7	10	17	22	24	1	6	23	26	25	14	29	44	20	4	43	19	28	42	2	32	31	9
		Random	14	28	32	44	6	10	30	12	17	20	26	33	43	2	4	5	11	27	1	7	8	19	22	23	25
		SFFS	6	12	5	25	11	26	1	3	4	32	7	10													
19	Viola	Ranking	38	5	20	37	12	39	15	18	4	23	6	21	11	36	27	40	16	1	13	28	33	3	26	25	34
		Random	38	2	31	3	30	39	40	9	11	13	16	17	33	44	4	8	10	14	15	19	36	37	42	5	6
		SFFS	1	2	3	13	20	8	4	23	21	12	25														
20	Violin	Ranking	6	21	34	16	2	9	41	24	7	25	15	18	5	4	40	31	20	11	29	19	32	10	14	33	22
		Random	9	28	11	15	29	2	5	14	17	19	34	37	42	8	10	12	16	20	22	26	27	33	35	38	41
		SFFS	20	22	42	7	19	29	25	27	14	11	28	3													
21	Other Instrument	Ranking	**16**	11	13	2	19	8	10	7	15	17	1	14	4	21	29	3	9	6	35	22	32	18	28	41	40
		Random	21	40	41	6	19	26	33	8	24	38	10	11	17	18	29	36	44	7	13	25	27	31	32	1	2
		SFFS	10	21	11	2	13	1	29	36	4	9															

Table 5.17: Details of Accuracy Value of the SVM Classifier with Predominant Feature Selection Methods.

Class No.	Musical Instrument	Selection Method	Accuracy	SVM Classifier
1	Banjo	Ranking	99.8	Medium Gaussian SVM
		Random	99.9	Medium Gaussian SVM
		SFFS	99.9	Medium Gaussian SVM
2	Bass Clarinet	Ranking	99.1	Cubic SVM
		Random	98.1	Cubic SVM
		SFFS	98.6	Fine Gaussian SVM
3	Bassoon	Ranking	97	Fine Gaussian SVM
		Random	95.4	Fine Gaussian SVM
		SFFS	98	Fine Gaussian SVM
4	Cello	Ranking	98.1	Fine Gaussian SVM
		Random	96.7	Fine Gaussian SVM
		SFFS	97.8	Fine Gaussian SVM
5	Clarinet	Ranking	97.7	Fine Gaussian SVM
		Random	97.2	Fine Gaussian SVM
		SFFS	98.1	Medium Gaussian SVM
6	Contra Bassoon	Ranking	98.3	Fine Gaussian SVM
		Random	97.7	Fine Gaussian SVM
		SFFS	98.5	Fine Gaussian SVM
7	Double Bass	Ranking	98.3	Fine Gaussian SVM
		Random	98.5	Fine Gaussian SVM
		SFFS	98.6	Cubic SVM

8	English Horn	Ranking	97.3	Fine Gaussian SVM
		Random	97.6	Fine Gaussian SVM
		SFFS	97.8	Fine Gaussian SVM
9	Flute	Ranking	99	Cubic SVM
		Random	97.4	Cubic SVM
		SFFS	97.7	Cubic SVM
10	French Horn	Ranking	97.3	Fine Gaussian SVM
		Random	97.6	Fine Gaussian SVM
		SFFS	97.8	Fine Gaussian SVM
11	Guitar	Ranking	99.5	Medium Gaussian SVM
		Random	99.4	Fine Gaussian SVM
		SFFS	99.5	Quadratic SVM
12	Mandolin	Ranking	99.8	Cubic SVM
		Random	99.1	Fine Gaussian SVM
		SFFS	99.7	Cubic SVM
13	Oboe	Ranking	99.4	Cubic SVM
		Random	98.2	Cubic SVM
		SFFS	98.2	Cubic SVM
14	Saxophone	Ranking	97.3	Fine Gaussian SVM
		Random	96.3	Fine Gaussian SVM
		SFFS	97.6	Fine Gaussian SVM
15	Tambura	Ranking	99.3	Cubic SVM
		Random	99.1	Fine Gaussian SVM
		SFFS	99.5	Cubic SVM

16	Trombone	Ranking	97.4	Fine Gaussian SVM
		Random	97.8	Fine Gaussian SVM
		SFFS	98.1	Fine Gaussian SVM
17	Trumpet	Ranking	97	Fine Gaussian SVM
		Random	97.4	Cubic SVM
		SFFS	97.2	Cubic SVM
18	Tuba	Ranking	99.7	Fine Gaussian SVM
		Random	97.7	Cubic SVM
		SFFS	99.8	Cubic SVM
19	Viola	Ranking	96.5	Cubic SVM
		Random	97.1	Fine Gaussian SVM
		SFFS	97.4	Cubic SVM
20	Violin	Ranking	97.3	Cubic SVM
		Random	98.1	Cubic SVM
		SFFS	98.6	Cubic SVM
21	Other Instrument	Ranking	98.9	Cubic SVM
		Random	98.4	Cubic SVM
		SFFS	99.2	Cubic SVM

5.9 Binary Classification used for Individual Musical Instrument Identification

The trained 21 SVM models set up the experimental model. Each SVM classifier was used to find their respective musical instrument. By the way, it helps to identify one more musical instrument in polyphonic music that has many musical instrument sounds.

5.9.1 Combine Two Musical Instruments

Combining the music files of two instruments. It was separated into three components by NMF and each component was divided into a 100-millisecond frame extracted the properties from each identified and identify the musical instruments from them. The first combination of the musical file is cello and double bass. Their results were shown in Table 5.18. The model identified cello musical instruments in the first, second, and third components but double bass was identified in the second component only. Mandolin and violin were also identified in the third component. The second combination of the musical files is the bass clarinet and double bass. Their results are shown in Table 5.18. The model identified bass clarinet and double bass in the first and third components.

Table 5.18: Musical Instrument Identification of the Combination of Cello and Double Bass Musical File.

Musical Instrument	1st Component		2nd Component		3rd Component	
	1st Frame	2nd Frame	1st Frame	2nd Frame	1st Frame	2nd Frame
Banjo	0	0	0	0	0	0
Bass Clarinet	0	0	0	0	0	0
Bassoon	0	0	0	0	0	0
Cello	1	1	1	1	0	1
Clarinet	0	0	0	0	0	0
Contra Bassoon	0	0	0	0	0	0
Double Bass	0	0	1	0	0	0
English Horn	0	0	0	0	0	0
Flute	0	0	0	0	0	0

French Horn	0	0	0	0	0	0
Guitar	0	0	0	0	0	0
Mandolin	0	0	0	0	1	0
Oboe	0	0	0	0	0	0
Saxophone	0	0	0	0	0	0
Tambura	0	0	0	0	0	0
Trombone	0	0	0	0	0	0
Trumpet	0	0	0	0	0	0
Tuba	0	0	0	0	0	0
Viola	0	0	0	0	0	0
Violin	0	0	0	0	0	0
Other Instrument	0	0	0	0	0	0

Note: 0 indicated the absence of the musical instrument and 1 indicated the presence of the musical instrument

Table 5.19: Musical Instrument Identification of the Combination of Bass Clarinet and Double Bass Musical File.

Musical Instrument	**1st Component**		**2nd Component**		**3rd Component**	
	1st Frame	**2nd Frame**	**1st Frame**	**2nd Frame**	**1st Frame**	**2nd Frame**
Banjo	0	0	0	0	0	0
Bass Clarinet	1	0	0	0	1	0
Bassoon	0	0	0	0	0	0
Cello	1	0	0	0	1	0

Clarinet	0	0	0	0	0	0
Contra Bassoon	0	0	0	0	0	0
Double Bass	1	0	0	0	1	0
English Horn	0	0	0	0	0	0
Flute	0	0	0	0	0	0
French Horn	0	0	0	0	0	0
Guitar	0	0	0	0	0	0
Mandolin	0	0	0	0	0	0
Oboe	0	0	0	0	0	0
Saxophone	0	0	0	0	0	0
Tambura	0	0	0	0	0	0
Trombone	0	0	0	0	0	0
Trumpet	0	0	0	0	0	0
Tuba	0	0	0	0	0	0
Viola	0	0	0	0	0	0
Violin	0	0	0	0	0	0
Other Instrument	1	0	1	1	1	0

Note: 0 indicated the absence of the musical instrument and 1 indicated the presence of the musical instrument

5.9.2 Combine Three Musical Instruments

Combining three musical instruments and music files. It also did the same procedure for combining two musical instruments' musical files. The first combination of the musical file is the cello, clarinet, and contrabassoon. Their results are shown in Table 5.20. The model identified the cello instrument in the first component. Contrabassoon was identified in the second component. However, the clarinet was nowhere to be identified but another instrument type was identified in the third component. The second combination of the musical file is the cello, the contrabassoon, and the English horn. The model has identified cello musical instruments in the third component.

Table 5.20: Musical Instrument Identification of the Combination of Cello, Clarinet, and Contrabassoon Musical File.

Musical Instrument	**1st Component**			**2nd Component**			**3rd Component**		
	1st Frame	**2nd Frame**	**3rd Frame**	**1st Frame**	**2nd Frame**	**3rd Frame**	**1st Frame**	**2nd Frame**	**3rd Frame**
Banjo	0	0	0	0	0	0	0	0	0
Bass Clarinet	0	0	0	0	0	0	0	0	0
Bassoon	0	0	0	0	0	0	0	0	0
Cello	1	1	1	0	0	0	0	0	0
Clarinet	0	0	0	0	0	0	0	0	0
Contra Bassoon	0	0	0	1	0	1	0	0	0
Double Bass	0	0	0	0	0	0	0	0	0
English Horn	0	0	0	0	0	0	0	0	0
Flute	0	0	0	0	0	0	0	0	0
French Horn	0	0	0	0	0	0	0	0	0
Guitar	0	0	0	0	0	0	0	0	0
Mandolin	0	0	0	0	0	0	0	0	0
Oboe	0	0	0	0	0	0	0	0	0
Saxophone	0	0	0	0	0	0	0	0	0
Tambura	0	0	0	0	0	0	0	0	0
Trombone	0	0	0	0	0	0	0	0	0

Trumpet	0	0	0	0	0	0	0	0	0
Tuba	0	0	0	0	0	0	0	0	0
Viola	0	0	0	0	0	0	0	0	0
Violin	0	0	0	0	0	0	0	0	0
Other Instrument	1	0	0	0	0	0	1	1	1

Note: 0 indicated the absence of the musical instrument and 1 indicated the presence of the musical instrument

Table 5.21: Musical Instrument Identification of the Combination of Cello, Contrabassoon and English horn Musical File.

Musical Instrument	**1st Component**		**2nd Component**		**3rd Component**	
	1st Frame	**2nd Frame**	**1st Frame**	**2nd Frame**	**1st Frame**	**2nd Frame**
Banjo	0	0	0	0	0	0
Bass Clarinet	0	0	0	0	0	0
Bassoon	0	0	0	0	0	0
Cello	0	0	0	0	1	0
Clarinet	0	0	0	0	0	0
Contra Bassoon	0	1	0	0	0	1
Double Bass	0	0	0	0	0	0
English Horn	0	0	0	1	0	0
Flute	0	0	0	0	0	0
French Horn	0	0	0	0	0	0
Guitar	0	0	0	0	0	0

Mandolin	0	0	0	0	0	0
Oboe	0	0	0	0	0	0
Saxophone	0	0	0	0	0	0
Tambura	0	0	0	0	0	0
Trombone	0	0	0	0	0	0
Trumpet	0	0	0	0	0	0
Tuba	0	0	0	0	0	0
Viola	0	0	0	0	0	0
Violin	0	0	0	0	0	0
Other Instrument	0	0	0	0	0	0

Note: 0 indicated the absence of the musical instrument and 1 indicated the presence of the musical instrument

5.9.3 Unknown Music File Analysis

The design model has identified polyphonic music. Table 5.22 shows the details of the identification of the musical instruments. 75% of cello musical instruments and 25% of the guitar musical instrument were identified in the first component of the polyphonic music file. The bass clarinet, saxophone, and tambura are 25% of each in the second component of the polyphonic music file. The guitar and trombone were identified in the third component of the polyphonic music file respectively 100% and 25%. The individual instrument is identified in each frame is shown in Table 5.23. Four musical instruments are identified in the first frame, five musical instruments are identified in the second frame and the identified instruments details are shown in Table 5.23. Table 5.23 is providing details of the identified musical instruments in the entire musical audio file. The second unknown musical file was processed in the automated system the result of the identified musical instruments is shown in Table 5.24. The Guitar was identified as 22.73% and 4.55% in the second and third components. Table 5.25 is providing details of the identified musical instruments in the entire musical audio file.

Table 5.22: Percentage of Musical Instrument Identification of the Polyphonic Music Audio File 1.

Musical Instrument	**1st Component %**	**2nd Component %**	**3rd Component %**
Banjo	0	0	0
Bass Clarinet	4	4	79
Bassoon	0	0	0
Cello	88	0	29
Clarinet	0	4	0
Contra Bassoon	0	0	0
Double Bass	0	0	0
English Horn	0	0	4
Flute	0	4	0
French Horn	0	0	0
Guitar	0	63	54
Mandolin	0	0	0
Oboe	33	38	29
Saxophone	0	0	0
Tambura	13	17	8
Trombone	29	0	67
Trumpet	0	0	0
Tuba	0	0	17
Viola	0	0	0
Violin	0	0	0
Other Instrument	0	29	0

Table 5.23: Automated System Identified Musical Instrument in 100 s Segmented Frame in the Polyphonic Music Audio File 1.

Musical Instrument	1st Frame	2nd Frame	3rd Frame	4th Frame	5th Frame	6th Frame	7th Frame	8th Frame	9th Frame	10th Frame	11th Frame	12th Frame	13th Frame	14th Frame	15th Frame	16th Frame	17 th Frame	18 th Frame	19 th Frame	20 th Frame
Banjo																				
Bass Clarinet																				
Bassoon																				
Cello																				
Clarinet																				
Contra Bassoon																				
Double Bass																				
English Horn																				
Flute																				
French Horn																				
Guitar																				
Mandolin																				
Oboe																				
Saxophone																				
Tambura																				
Trombone																				
Trumpet																				
Tuba																				
Viola																				
Violin																				
Other Instrument																				

Table 5.24: Percentage of Musical Instrument Identification of the Polyphonic Music Audio File 2.

Musical Instrument	**1st Component %**	**2nd Component %**	**3rd Component %**
Banjo	0.00	0.00	0.00
Bass Clarinet	9.09	0.00	20.45
Bassoon	0.00	0.00	0.00
Cello	31.82	38.64	56.82
Clarinet	27.27	0.00	0.00
Contra Bassoon	0.00	0.00	4.55
Double Bass	0.00	0.00	0.00
English Horn	0.00	0.00	0.00
Flute	0.00	0.00	0.00
French Horn	4.55	0.00	0.00
Guitar	0.00	22.73	4.55
Mandolin	0.00	0.00	0.00
Oboe	0.00	13.64	6.82
Saxophone	0.00	0.00	0.00
Tambura	56.82	63.64	27.27
Trombone	0.00	4.55	4.55
Trumpet	0.00	0.00	0.00
Tuba	0.00	0.00	0.00
Viola	0.00	0.00	0.00
Violin	0.00	6.82	0.00
Other Instrument	0.00	6.82	0.00

Table 5.25: Automated System Identified Musical Instrument in 100 s Segmented Frame in the Polyphonic Music Audio File 2.

Instrument	1st Frame	2nd Frame	3rd Frame	4th Frame	5th Frame	6th Frame	7th Frame	8th Frame	9th Frame	10th Frame	11th Frame	12th Frame	13th Frame	14th Frame	15th Frame	16th Frame	17th Frame	18th Frame	19th Frame	20th Frame
Banjo																				
Bass Clarinet																				
Bassoon																				
Cello																				
Clarinet																				
Contra Bassoon																				
Double Bass																				
English Horn																				
Flute																				
French Horn																				
Guitar																				
Mandolin																				
Oboe																				
Saxophone																				
Tambura																				
Trombone																				
Trumpet																				
Tuba																				
Viola																				
Violin																				
Other Instrument																				

Table 5.26: Percentage of Musical Instrument Identification of the Polyphonic Music Audio File 3.

Musical Instrument	**1st Component %**	**2nd Component %**	**3rd Component %**
Banjo	0.00	0.00	0.00
Bass Clarinet	0.00	0.00	0.00
Bassoon	0.00	0.00	0.00
Cello	0.00	14.29	0.00
Clarinet	35.71	0.00	0.00
Contra Bassoon	0.00	0.00	0.00
Double Bass	0.00	0.00	0.00
English Horn	14.29	0.00	7.14
Flute	35.71	0.00	14.29
French Horn	0.00	0.00	0.00
Guitar	7.14	42.86	0.00
Mandolin	0.00	0.00	0.00
Oboe	14.29	0.00	0.00
Saxophone	28.57	7.14	7.14
Tambura	0.00	21.43	0.00
Trombone	0.00	21.43	0.00
Trumpet	0.00	0.00	0.00
Tuba	0.00	0.00	0.00
Viola	0.00	0.00	0.00
Violin	57.14	0.00	7.14
Other Instrument	35.71	21.43	7.14

Table 5.27: Automated System Identified Musical Instrument in 100 s Segmented Frame in the Polyphonic Music Audio File 3.

	1st Frame	2nd Frame	3rd Frame	4th Frame	5th Frame	6th Frame	7th Frame	8th Frame	9th Frame	10th Frame	11th Frame	12th Frame	13th Frame	14th Frame	15th Frame	16th Frame	17th Frame	18th Frame	19th Frame	20th Frame
Banjo																				
Bass Clarinet							■													
Bassoon																				
Cello	■	■	■	■	■	■	■	■	■			■	■	■		■	■			■
Clarinet				■	■															
Contra Bassoon										■					■			■		
Double Bass																				
English Horn	■																■			
Flute																				
French Horn																				
Guitar		■				■							■				■	■		
Mandolin																				
Oboe			■									■		■			■			
Saxophone			■			■														
Tambura		■					■	■	■	■		■	■	■		■		■	■	■
Trombone																				
Trumpet	■																			
Tuba																				
Viola																				
Violin						■		■	■	■					■	■	■			
Other Instrument				■																

Table 5.28: Percentage of Musical Instrument Identification of the Polyphonic Music Audio File 4.

Musical Instrument	**1st Component %**	**2nd Component %**	**3rd Component %**
Banjo	0.00	0.00	0.00
Bass Clarinet	7.14	7.14	0.00
Bassoon	0.00	0.00	0.00
Cello	0.00	0.00	0.00
Clarinet	28.57	14.29	0.00
Contra Bassoon	0.00	0.00	0.00
Double Bass	0.00	0.00	0.00
English Horn	21.43	28.57	7.14
Flute	35.71	35.71	0.00
French Horn	0.00	0.00	0.00
Guitar	7.14	0.00	28.57
Mandolin	0.00	0.00	0.00
Oboe	14.29	14.29	0.00
Saxophone	28.57	14.29	0.00
Tambura	7.14	0.00	21.43
Trombone	0.00	0.00	28.57
Trumpet	7.14	14.29	0.00
Tuba	0.00	0.00	0.00
Viola	0.00	0.00	0.00
Violin	50.00	28.57	0.00
Other Instrument	7.14	21.43	21.43

Table 5.29: Automated System Identified Musical Instrument in 100 s Segmented Frame in the Polyphonic Music Audio File 4.

	1st Frame	2nd Frame	3rd Frame	4th Frame	5th Frame	6th Frame	7th Frame	8th Frame	9th Frame	10th Frame	11th Frame	12th Frame	13th Frame	14th Frame	15th Frame	16th Frame	17th Frame	18th Frame	19th Frame	20th Frame
Banjo																				
Bass Clarinet	■	■	■					■							■			■	■	
Bassoon																				
Cello	■		■		■			■	■	■	■	■	■	■	■	■	■		■	■
Clarinet	■								■	■										
Contra Bassoon																				
Double Bass																				
English Horn									■										■	■
Flute				■			■		■			■	■							
French Horn																				
Guitar				■					■	■										
Mandolin																				
Oboe	■						■	■		■	■	■	■	■	■	■	■	■	■	
Saxophone		■	■	■										■	■					
Tambura	■							■									■		■	
Trombone							■													
Trumpet																	■	■		
Tuba																				
Viola																				
Violin		■				■				■	■			■	■					
Other Instrument		■	■	■	■	■		■												

5.10 Hierarchical Cluster (HC) Method for Monophonic and Polyphonic Music File

The monophonic music files were examined by the HC technique. Firstly, we make 105 subsets of raw data to five data for each musical instrument. Each instrument's five musical files' features and known musical instrument's features were taken to the HC method. Table 5.30 was shown the details of the musical instruments belong to the sample number. The examined dendrogram and cluster results were shown in Figure 5.9, Figure 5.10, Figure 5.9, Table 5.31, Table 5.32, and Table 5.33 respectively.

Table 5.30: Sample Number belongs to the Musical Instrument Used for HC.

Sample Number Denoted for Musical Instrument	Musical Instrument
1 to 5	Belong to **Banjo**
6 to 10	Belong to **Bass Clarinet**
11 to 15	Belong to **Bassoon**
16 to 20	Belong to **Cello**
21 to 25	Belong to **Clarinet**
26 to 30	Belong to **Contra Bassoon**
31 to 35	Belong to **Double Bass**
36 to 40	Belong to **English Horn**
41 to 45	Belong to **Flute**
46 to 50	Belong to **French Horn**
51 to 55	Belong to **Guitar**
56 to 60	Belong to **Mandolin**
61 to 65	Belong to **Oboe**
66 to 70	Belong to **Saxophone**
71 to 75	Belong to **Tambura**
76 to 80	Belong to **Trombone**
81 to 85	Belong to **Trumpet**
86 to 90	Belong to **Tuba**
91 to 95	Belong to **Viola**
96 to 100	Belong to **Violin**
101 to 105	Belong to **Other Instrument**
106	belongs to **Known/Unknown Music File**

5.10.1 Known Music File Analysis

The new known musical file was learned from the hierarchical cluster method. It identifies groups in the given dataset. The first new known file is a banjo and it joined the group of banjo the musical instrument file's number range is 2 to 5. The dendrogram was visualized the output was shown in Figure 5.9. The grouped files are denoted by the red rectangular object. The full cluster groups' result is shown in Table 5.34 and were highlighted in green color. It was grouped correctly in the same banjo category.

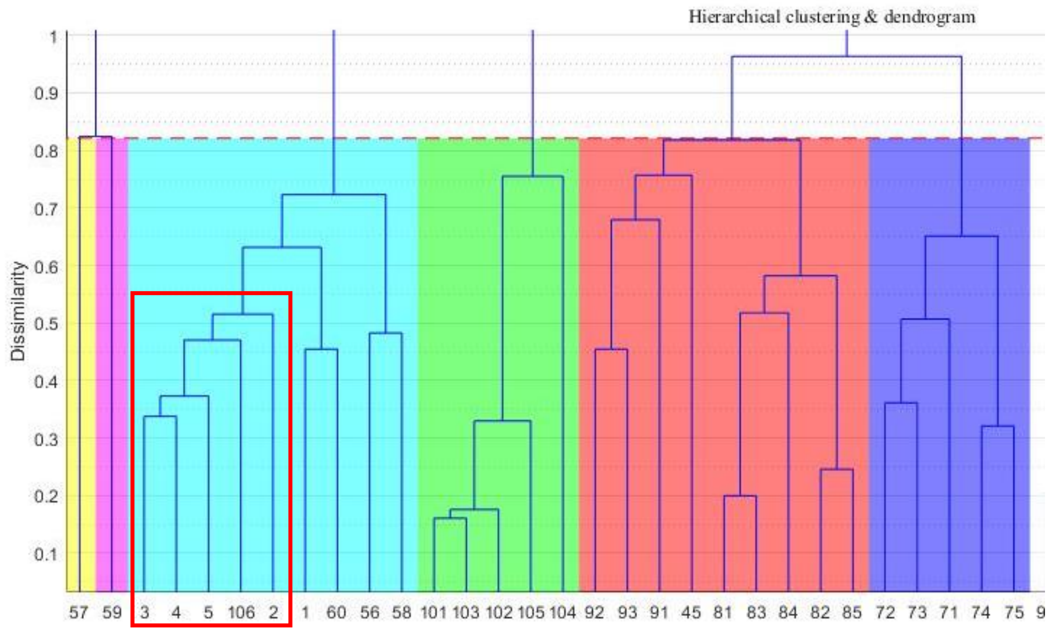

Figure 5.9: Dendrogram for New Known Banjo Musical Instrument Audio File.

Table 5.31: Cluster Details of New Known Banjo Musical Instrument Audio File.

Clusters
57
59
[3,4,5,106,2,1,60,56,58]
[101,103,102,105,104]
[92,93,91,45,81,83,84,82,85]
[72,73,71,74,75]
[98,99,96,100,97]
[20,34,16,17,94,95]

[32,33,35,31,13,14,11,12,15,67,68,66]
[86,88,89,90]
[46,50,47,49,48,28,29,26,27,30]
[7,8,9,10,6]
[69,70]
[76,80,18,78,79,77,19,43,37,38,39,40,36]
[61,62,63,64,65]
[41,42]
44
[21,24,25,22]
23
87
[51,52,54,55,53]

The next new known musical file is the bass clarinet and it joined the group banjo the musical instrument file number range is 6 to 10. The dendrogram was visualized the output was shown in Figure 5.10. The grouped files are denoted by the red rectangular object. The full clusters' results were shown in Table 5.32 and were highlighted in green color. It was grouped correctly in the same bass clarinet category.

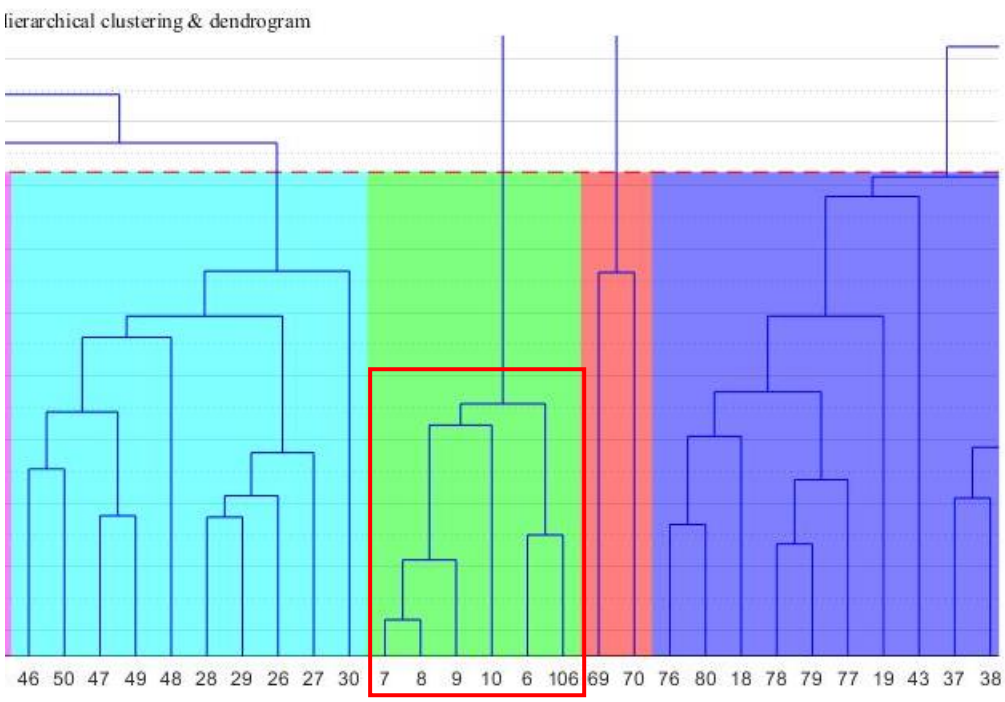

Figure 5.10: Dendrogram for Bass Clarinet Musical Instrument Audio File.

Table 5.32: Cluster Details of New Known Bass Clarinet Musical Instrument Audio File

Clusters
57
59
[3,4,5,2,1,60,56,58]
[101,103,102,105,104]
[92,93,91,45,81,83,84,82,85]
[72,73,71,74,75]
[98,99,96,100,97]
[20,34,16,17,94,95]
[32,33,35,31,13,14,11,12,15,67,68,66]
[86,88,89,90]
[46,50,47,49,48,28,29,26,27,30]
[7,8,9,10,6,106]
[69,70]
[76,80,18,78,79,77,19,43,37,38,39,40,36]
[61,62,63,64,65]
[41,42]
44
[21,24,25,22]
23
87
[51,52,54,55,53]

5.10.2 Unknown Music File Analysis

The unknown musical file was clustered with the known sample files. It was visualized dendrogram in Figure 5.11 and denoted by the red rectangular object. The full clusters' results were shown in Table 5.33 and were highlighted in green color it was joined by a trumpet or viola, a musical instrument group.

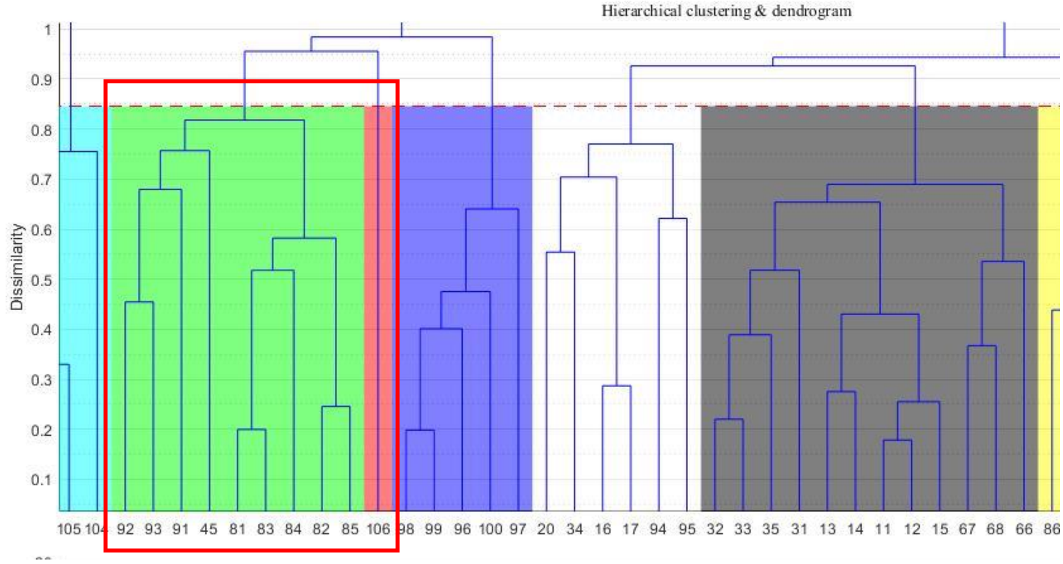

Figure 5.11: Dendrogram for the First Unknown Polyphonic Music Audio File

Table 5.33: Cluster Details of the First Unknown Polyphonic Music Audio File.

Clusters
[57,59]
[3,4,5,2,1,60,56,58]
[101,103,102,105,104]
[92,93,91,45,81,83,84,82,85]
106
[98,99,96,100,97]
[20,34,16,17,94,95]
[32,33,35,31,13,14,11,12,15,67,68,66]
[86,88,89,90]
[46,50,47,49,48,28,29,26,27,30]
[69,70]
[72,73,71,74,75]
[7,8,9,10,6]
[76,80,18,78,79,77,19,43,37,38,39,40,36]
[61,62,63,64,65]
[41,42]
44
[21,24,25,22]
23
87
[51,52,54,55,53]

HC method was applied for another unknown music file and cluster details are shown in Figure 5.12 and denoted by the red rectangular object. The full clusters' results were shown in Table 5.34 and were highlighted in green color. It was grouped into saxophone musical instruments.

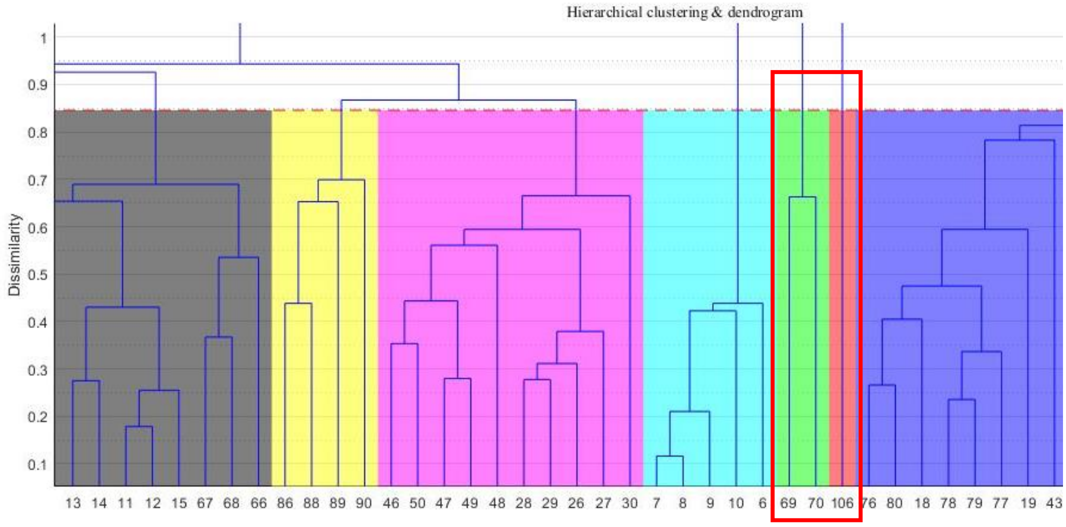

Figure 5.12: Dendrogram for the Second Unknown Polyphonic Music Audio File.

Table 5.34: Cluster Details of the Second Unknown Polyphonic Music Audio File.

Clusters
[92,93,91,45,81,83,84,82,85]
[72,73,71,74,75]
[98,99,96,100,97]
[20,34,16,17,94,95]
[32,33,35,31,13,14,11,12,15,67,68,66]
[86,88,89,90,46,50,47,49,48,28,29,26,27,30]
[7,8,9,10,6]
[69,70]
106
[76,80,18,78,79,77,19,43,37,38,39,40,36]
[61,62,63,64,65]
[41,42]
44
107

[21,24,25,22]
23
87
[51,52,54,55,53]
[57,59]
[3,4,5,2,1,60,56,58]
[101,103,102,105,104]

HC method was applied for another unknown music file and cluster details are shown in Figure 5.13 and denoted by the red rectangular object. The full clusters' results were shown in Table 5.35 and were highlighted in green color. It was grouped to viola musical instrument.

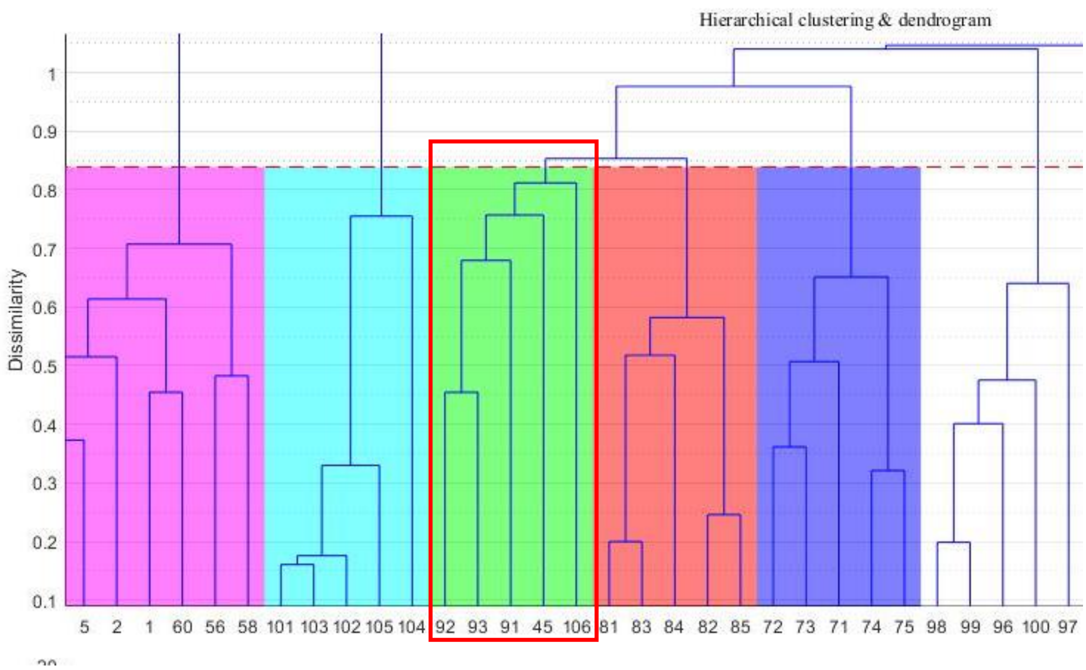

Figure 5.13: Dendrogram for the Third Unknown Polyphonic Music Audio File.

Table 5.35: Cluster Details of the Third Unknown Polyphonic Music Audio File.

Clusters
[57,59]
[3,4,5,2,1,60,56,58]
[101,103,102,105,104]
[92,93,91,45,106]
[81,83,84,82,85]
[72,73,71,74,75]
[98,99,96,100,97]
[20,34,16,17,94,95]
[32,33,35,31,13,14,11,12,15,67,68,66]
[86,88,89,90]
[46,50,47,49,48,28,29,26,27,30]
[7,8,9,10,6]
[69,70]
[76,80,18,78,79,77,19,43,37,38,39,40,36]
[61,62,63,64,65]
[41,42]
44
[21,24,25,22]
23
87
[51,52,54,55,53]

The hierarchical cluster method was applied for monophonic and polyphonic music identification. It was recognized only monophonic instruments which were similar features values were the same as the samples. However, the polyphonic music file also recognized only one instrument. Therefore, the hierarchical cluster method was suitable for only monophonic music identification

5.11 Summary

This chapter has discussed the result of the set of best features selection that has been selected by the different feature selection methods. The selected set of best features used for classification of the results is analyzed and selected the suitable feature selection method. The classification performance was compared with the selected best set of features and the classifiers were evaluate.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

This study has introduced the SFFS method with the greedy algorithm selecting features; it has selected the best set of features as predominant features for each musical instrument. The selected significant features were used to identify the musical instruments in polyphonic music.

The research study has focused on three major parts, which are feature extraction, feature selection, and classification. The features extracted from the audio signal. This objective was achieved through the literature reviews. It is a comprehensive investigation of the view of the recognized existing technique and the result of the research findings. The audio signal has the time domain and frequency domain. The Audio signal segment to 100 milliseconds frame. The "Audio Content Analysis" tool extracted forty-four features. It has extracted 11 spectral features, 8-time domain features, 13 MFFC features and, 12 Pitch Chroma features, but not all features do necessary for musical instrument identification because some features are irrelevant. The extracted features have been discussed in chapter 4. The feature selection used three methods, namely ranking, random and sequential forward feature selection. Three feature selection methods are examined to study the impact of finding the best features. A set of 44 features are used with Rank Feature selection. Random Feature Selection and Sequential Forward Feature Selection algorithms. This objective determines the predominant features of each musical instrument. The details of the techniques are in chapter 3 and the results and suitable techniques are illustrated in chapter 5.

The theoretical framework is formed based on the initial step to the final step of the identification process. Data acquisition, data format, data pre-processing, feature extraction, feature selection, and classification are the six phases that make up this process. These phases are constructed during the study period. The details of the theoretical framework are described in chapter 4.

The ranking feature selection method selects in order from the highest rank of feature to low-rank feature. The random feature selection method carries out a randomized subset feature pursuit supported by classification. It randomly generates a subset of features utilized to classify samples. Each subset was appraised with the apparent error. Only the best subsets were retained and, they were amalgamated into a single pool. The sequential forward feature selection starts with one predictor and adds more iteratively. At each iteration, the best of the remained original predictors were added based on the performance. The most suitable feature as S1 selects firstly using some selection criterion. Then the pair of features were formed with S1 and the best pair selects as S2. This process was repeated until the optimized number of best features is selected. The three feature selection methods results were tabled in Table 5.13 and the optimized number of best features amount to be taken for the other two methods. Four classifiers were used to classify musical instrument identification, which are Decision Tree (DT), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). Classifiers functions and details are described in chapter 3 and experimental results are discussed in chapter 5.

The selected features were used in the SVM classifier and the highest accuracy value has been got from the SFFS feature selection method, where the result for the each musical instrument, is shown in Table 5.14 and CNN (Wavenet) classifier has the highest validated accuracy with 90.25%, but it is less than SVM's accuracy and CNN classifier did not given selected feature details and it was in hidden layer. Therefore, the automated model has been developed by 21 SVM classifiers. Each SVM classifier consists of the related SFFS optimized predominant features for each musical instrument. The combined music audio file is separated into three components by NMF. Each component is segmented into a 100 ms frame. After that, the automated model examines each frame and identifies the musical instrument. The results are shown in Table 5.18, Table 5.19, Table 5.20, and Table 5.21.

The same procedure was also applied for polyphonic music audio files (unknown audio files), but experiment results were computed component-wise. The component contains the percentage of each musical instrument, which musical instrument has higher percentage values, so those musical instruments are in the polyphonic music file. The results are shown in Table 5.22, Table 5.24, Table 5.26 and Table 5.28. If the amplitude of the combined audio signals is different, which audio has a high amplitude that dominated the identification results. Therefore amplitude is influenced in the musical instrument identification.

SFFS approach-based classification mechanism introduced in this study has provided a new direction for the oriental music instrument identification with respective to the results obtained. This methodology has shown an improved way to handle the complexity of feature dimensions by reducing the unwanted noises due to the ambiguous features.

CHAPTER 7

FUTURE DIRECTIONS AND LIMITATIONS

7.1 Limitations

- This study has explored up to 20 musical instruments, current state, it has identified only 20 musical instruments. If the number of musical instruments increases in this study, the accuracy and precision may vary where it is to be experimented.
- There are many feature selection techniques are used in past works but three feature selection techniques were used in this study.
- The study has not employed any spatial segmentation of spectrum of the musical instruments where some set of features could be positive from unnecessary instruments because of using binary multiclass classifiers.

7.2 Future Directions

- This research work done with the low-level features and limited numbers of features, but future work could show that music classification based on more low-level descriptors. It could show that music classification based on low-level descriptors is comparable or even preferable to classification based on high-level audio features.
- Future work could focus on developing models for added instruments like percussion family, electronics, string, woodwind, and brass instruments which were not used in the current studies.

REFERENCES

[1] H. Editors, "World Wide Web (WWW) launches in the public domain," March 2020. [Online].

[2] N. Alias, S. H. Abd Razak, G. elHadad, N. R. Mat Noh, K. Kunjambu and P. Muniandy, "A content analysis in the studies of YouTube in selected journals," in *13th International Educational Technology Conference*, 2013.

[3] L. Kalapuri and M. Davy, Signal Processing Methods for Music Transcription, New York: Springer Science + Business Media LLC, 2000.

[4] D. S. Bormane and D. Meenakshi, "A Novel Techniques for Classification of Musical Instruments," *Information and Knowledge Management*, vol. 3, no. 10, 2013.

[5] M. Muller, D. P. Ellis, A. Klapuri and G. Richard, "Signal Processing for Music Analysis," *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*, 2011.

[6] Rick, "Music Theory Lessons," [Online]. Available: https://studionotesonline.com/category/music-theory-lessons/. [Accessed 20 Feb 2021].

[7] A. Piacsek, "Sound," *AccessScience*, July 2020.

[8] H. A. Arjimand, J. Hogajen, B. Paton and N. S. Rickard, "Emotional Responses to Music: Shifts in Frontal Brain Asymmetry Mark Periods of Musical Change," *Front. Psychol*, 2017.

[9] V. Brandes, "Music as medicine: incorporating scalable music-based interventions into standard medical practice," in *Music that works*, Austria, SpringerWienNewYork, 2009, pp. 83-103.

[10] J. London, M. Thompson, B. Burger, M. Hildreth and P. Toiviainen, "Tapping doesn’t help: Synchronized self-motion and judgments of musical tempo," *Atten Percept Psychophys*, vol. 81, p. 2461–2472, 2019.

[11] Hans-Eckhardt Schaefer, "Music-Evoked Emotions—Current Studies," *Frontiers in Neuroscience*, vol. 11, Nov 2017.

[12] M. A. Casey, C. Rhodes, R. Veltkamp, M. Goto, M. Leman and M. Slaney, "Content-Based Music Information Retrieval: Current Directions and Future Challenges," in *Proceedings of the IEEE*, 2008.

[13] I. Fujinaga, M. Goto and G. Tzanetakis, "Music Information Retrieval Based on Signal Processing," *EURASIP Journal on Advances in Signal Processing*, 2007.

[14] J. D. Fernandez and F. Vico, "AI Methods in Algorithmic Composition: A Comprehensive Survey," *Journal of Artificial Intelligence Research*, vol. 48, pp. 513-582, 2013.

[15] R. L. Harrison, S. Bilbao, J. Perry and T. Wishart, "An Environment for Physical Modeling of Articulated Brass Instruments," *Computer Music Journal*, vol. 39, no. 5, pp. 80-95, 2016.

[16] S. H. Chon, D. Huron and D. DeVlieger, "An Exploratory Study of Western Orchestration: Patterns through History," *Empirical Musicology Review,* vol. 7, 2012.

[17] F. Alias, J. C. Socoro and X. Sevillano, "A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds," *applied sciences,* 2016.

[18] P. Poonam and C. Soubhik, "Signal Processing for Music Analysis," *American International Journal of Research in Science, Technology, Engineering & Mathematics,* 2015.

[19] D. G. Bhalke, C. B. Ramo Rao, D. S. Bormane and M. Vibhute, "Spectrogram Base Musical Instrument Identification Using Hidden Markov Model for Monophonic and Polyphonic Musical Signals," *ACT TECHNICA NAPOCENSIS Electronics and Telecommunications,* vol. 52, 2 November 2011.

[20] D. G. Bhalke, C. B. Ramo Rao and D. S. Bormane, "Hybridization of Fractional Fourier Transform and Acoustic Features for Musical Instrument Recognition," *International Journal of Signal Processing, Image Processing and Pattern Recognition,* vol. 7, pp. 275-282, 2014.

[21] D. Chandwadkar and M. Sutaone, "Selecting Proper Features and Classifiers for Accurate Identification of Musical Instruments," *International Journal of Machine Learning and Computing,* vol. 3, pp. 172-175, April 2013.

[22] S. Essid, D. Richard and D. Bertrand, "Musical Instrument Recognition by Pairwise Classification Strategies," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 14, pp. 1401-1411, July 2006.

[23] P. X. JORIS, E. C. SCHREINER and A. REES, "Neural Processing of Amplitude-Modulated Sounds," *Physiol Rev,* vol. 84, pp. 541-577, 2004.

[24] M. Raghu, "A Study to Explore the Effects of Sound Vibrations on Consciousness," *International Journal of Social Work and Human Services Practice,* vol. 6, no. 3, pp. 75-88, 2018.

[25] G. Peeters, S. McAdams and P. Herrera, "Instrument Sound Description in the Context of MPEG-7," in *Proceeding of ICMC2000 (International Computer Music Conference),* Berlin, 2000.

[26] J. Deng, C. Simmermacher and S. Cranefield, "A Study on Feature Analysis for Musical Instrument Classification," *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics,* vol. 38, no. 2, April 2008.

[27] M. KarpakaMurthy, S. Seetha and L. P. Flavio, "Generating MPEG 7 Audio Descriptor for Content-Based Retrieval," *978-1-4244-9477-4/11/$26.00 ©2011 IEEE.*

[28] T. Kim, J. Kim and K. Lee, "Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques," *BioMedical Engineering Online,* vol. 17, no. 16, 2018.

[29] M. Mandel and D. Ellis, "Song-level Features and SVMs for Music Classification," *Proc. Int. Conf. Music Information Retrieval,* 2005.

[30] G. Arijit, B. Chan, S. Saha and C. Rudrasis, "Music Classification Based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach," *International Journal of Signal Processing, Image Processing and Pattern Recognition,* vol. 5, no. 1, March 2012.

[31] A. Kharb and N. Hooda, "Performance Analysis of Daubechies Wavelets for Music Signals," *International Journal of Recent Research Aspects,* vol. 2, no. 2, pp. 117-124, June 2015.

[32] R. S. Kothe and D. G. Bhalke, "Musical Instrument Recognition using Wavelet Coefficient Histograms," *Wavelet Coefficient Histograms. Proceedings on Emerging Trends in Electronics and Telecommunication Engineering,* 2013.

[33] M. Kumari, P. Kumar and S. Solanki, "Classification of North Indian Musical Instruments Using Spectral Features," *GESJ: Computer Science and Telecommunications,* vol. 6, no. 29, 2010.

[34] M. E. O zbek, C. Delpha and P. Duhamel, "Musical Note and Instrument Classification with Likelihood-Frequency-Time Analysis and Support Vector Machines.," *15th European Signal Processing Conference (EUSIPCO 2007),* September 2007.

[35] V. S. Shelar and D. G. Bhalke, "Musical Instrument Recognition and Transcription Using Neural Network," *International Journal of Computer Applications,* 2013.

[36] P. S. Jadhav, "Classification of Musical Instruments Sounds by Using MFCC and Timbral Audio Descriptors," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 3, no. 7, July 2015.

[37] P. Hamel and D. EcK, "Learning Features from Music Audio with Deep Belief Networks," *11th International Society for Music Information Retrieval Conference,* 2010.

[38] P. Kalapatapu, S. Goli, P. Arthum and A. Malapati, "A Study on Feature Selection and Classification Techniques of Indian Music," *ELSEVIER,* 2016.

[39] D. G. Bhalke, C. B. Ramo Rao and D. S. Bormane, "Hybridisation of Mel Frequency Cepstral Coefficient and Higher Order Spectral Features for Musical Instruments Classification," *PAN, ARCHIVES OF ACOUSTICS,* vol. 41, no. 3, pp. 427- 436, 2016.

[40] R. S. Kothe and D. G. Bhalke, "Musical instrument recognition using wavelet coefficient histograms," *International Journal of Computer Application,* pp. 37-41, 2013.

[41] S. Esside, G. Richard and B. David, "Instrument Recognition in Polyphonic Music Based on Automatic Taxonomies," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 14, no. 1, pp. 68-80, 2006.

[42] V. Kumar and S. Minz, "Feature Selection: A Literature Review," *Smart Computing Review*, vol. 4, no. 3, pp. 211-229, 2014.

[43] P. Narendra and K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection," *IEEE Transactions on Computer*, vol. 26, no. 9, pp. 917-922, 1977.

[44] G. H. John, R. Kohavi and K. Pfleger, "Irrelevant feature and the subset selection problem," in *Proc. of the Eleventh International Conference on Machine Learning*, 1994.

[45] H. Liu and H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Springer, Boston, MA, 1998.

[46] Osman Gokalp, Erdal Tasci and Aybars Ugur, "A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification," *Expert Systems With Applications Elsevier*, 2020.

[47] A. C. Haury, P. Gestraud and J. P. Vert, "The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures," *PLoS ONE*, vol. 6, no. 12, 2011.

[48] A. K. Farahat, A. Ghodsi and M. S. Kamel, "An Efficient Greedy Method for Unsupervised Feature Selection," in *IEEE 11th International Conference on Data Mining*, 2011.

[49] S. S. Hameed, O. O. Petinrin, A. O. Hashi and F. Saeed, "Filter-Wrapper Combination and Embedded Feature Selection for Gener Expression Data," *Int. J. Advance Soft Compu. Appl*, vol. 10, no. 1, pp. 90-105, Mar. 2008.

[50] Meihong Zhu and Jie Song, "An Embedded Backward Feature Selection Method for MCLP Classification Algorithm," in *Information Technology and Quantitative Management (ITQM2013)*, 2013.

[51] S. Vanaja and K. Ramesh Kumar, "Analysis of Feature Selection Algorithms on Classification: A Survey," *International Journal of Computer Applications (0975 – 8887)*, vol. 96, no. 17, pp. 28-35, 2014.

[52] J. Pirgazi, M. Almoradi, T. E. Afharian and M. H. Olyaee, "An efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets," *SCIENTIFIC REPORTS*, 2019.

[53] Prafulla Kalapatapu, Srihita Goli, Prasanna Arthum and Aruna Malapati, "A Study On Feature Selection And Classification Techniques Of Indian Music," in *The 7th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2016)*.

[54] Hyungsik Shin and Jeongyeup Paek, "Automatic Task Classification via Support Vector Machine and Crowdsourcing," *Mobile Information Systems*, 2018.

[55] A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior and K. Kavukcuoglu, "Wavenet: A Generative Model for Raw Audio," in *arXiv:1609.03499v2 [cs.SD] 19 Sep 2016*.

[56] Yoko Anan, Kohei Hatano, Hideo Bannai, Masayuki Takeda and Ken Satoh, "Polyphonic Music Classification on Symbolic Data Using Dissimilarity

Functions," in *13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, 2012.

[57] B. Wang and M. D. Plumbley, "Musical audio stream separation by non-negative matrix factorization," in *Conference: Proceedings of the DMRN Summer Conference*, 2005.

[58] S. Vembu and S. Baumann, "Separation of Vocals from Polyphonic Audio Recordings," in *Conference: ISMIR 2005, 6th International Conference on Music Information Retrieval*, London, 2005.

[59] Paul Magron, Roland Badeau and Bertrand David, "Phase recovery in NMF for audio source separation: an insightful benchmark," in *40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, 2015.

[60] K. Uma Maheswari and A. Valarmathi, "A Novel Feature Selection Algorithm for Coronary Artery Disease Prediction," *Applied Mathematics & Information Sciences An International Journal*, vol. 12, no. 4, pp. 735-744, 2018.

[61] P. musicians., "Sound samples".

[62] A. Lerch, "audio content analysis music information retrieval & machine listening," 2016. [Online]. Available: https://www.audiocontentanalysis.org/.

[63] M. Mauch, H. Fujihara, K. Yoshii and M. Goto, "Timber and melody features for the recognition of vocal activity and instrument solos in polyphonic music," in *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 2011.

[64] L. E. Gomez, H. Sossa, R. Barron and J. F. Jimenez, "Dynamic Neural Networks Applied to Melody Retrieval," in *Advances in Soft Computing - 9th Mexican International Conference on Artificial Intelligence*, Pachuca, 2010.

APPENDIX A: THE DETAILS OF THE PERFORMANCE OF CLASSIFIERS

The following TableA.1, Table A.2, and Table A.3 show the results obtained for the performance of the SVM, kNN, and DT classifiers calculation from the confusion matrices. Table A.4 shows the validated details of the 20 Musical Instruments.

Table A.1: The Details of SVM Classifier Identified for 20 Musical Instruments.

Musical Instrument	True Positive	False Negative	True Negative	False Positive	Accuracy	Precision	Sensitivity	Specificity
Banjo	0.97	0.03	1.00	0.00	0.99	1.00	0.97	1.00
Cello	0.93	0.07	0.93	0.07	0.93	0.93	0.93	0.93
Double Bass	0.97	0.03	0.97	0.03	0.97	0.97	0.97	0.97
Guitar	0.76	0.24	0.99	0.01	0.87	0.98	0.76	0.99
Mandolin	0.96	0.04	0.93	0.07	0.95	0.93	0.96	0.93
Tambura	0.44	0.56	0.70	0.30	0.57	0.59	0.44	0.70
Viola	0.94	0.06	0.93	0.07	0.94	0.93	0.94	0.93
Violin	0.96	0.04	0.94	0.06	0.95	0.95	0.96	0.94
Bass Clarinet	0.98	0.02	0.97	0.03	0.97	0.97	0.98	0.97
French Horn	0.95	0.05	0.96	0.04	0.96	0.96	0.95	0.96
Trombone	0.97	0.03	0.96	0.04	0.96	0.96	0.97	0.96
Trumpet	0.93	0.07	0.93	0.07	0.93	0.93	0.93	0.93
Tuba	0.99	0.01	1.00	0.00	1.00	1.00	0.99	1.00
Bassoon	0.96	0.04	0.96	0.04	0.96	0.96	0.96	0.96
Clarinet	0.92	0.08	0.93	0.07	0.93	0.93	0.92	0.93
Contra Bassoon	0.97	0.03	0.97	0.03	0.97	0.97	0.97	0.97
English Horn	0.97	0.03	0.96	0.04	0.96	0.96	0.97	0.96
Flute	0.99	0.01	0.99	0.01	0.99	0.99	0.99	0.99

Oboe	0.96	0.04	0.96	0.04	0.96	0.96	0.96	0.96
Saxophone	0.92	0.08	0.92	0.08	0.92	0.92	0.92	0.92
Overall					**0.93**			

Table A.2: The Details of kNN Classifier Identified for 20 Musical Instruments.

Musical Instrument	**True Positive**	**False Negative**	**True Negative**	**False Positive**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	1.00	0.00	0.99	0.01	0.99	0.99	1.00	0.99
Cello	0.86	0.14	0.90	0.10	0.88	0.90	0.86	0.90
Double Bass	0.96	0.04	0.93	0.07	0.94	0.93	0.96	0.93
Guitar	0.66	0.34	0.92	0.08	0.79	0.89	0.66	0.92
Mandolin	0.95	0.05	0.97	0.03	0.96	0.97	0.95	0.97
Tambura	0.56	0.44	1.00	0.00	0.78	1.00	0.56	1.00
Viola	0.90	0.10	0.85	0.15	0.88	0.86	0.90	0.85
Violin	0.91	0.09	0.91	0.09	0.91	0.91	0.91	0.91
Bass Clarinet	0.96	0.04	0.96	0.04	0.96	0.96	0.96	0.96
French Horn	0.92	0.08	0.94	0.06	0.93	0.94	0.92	0.94
Trombone	0.96	0.04	0.93	0.07	0.95	0.94	0.96	0.93
Trumpet	0.87	0.13	0.89	0.11	0.88	0.89	0.87	0.89
Tuba	0.99	0.01	1.00	0.00	1.00	1.00	0.99	1.00
Bassoon	0.94	0.06	0.95	0.05	0.95	0.95	0.94	0.95
Clarinet	0.91	0.09	0.90	0.10	0.91	0.90	0.91	0.90
Contra Bassoon	0.96	0.04	0.92	0.08	0.94	0.92	0.96	0.92
English Horn	0.95	0.05	0.93	0.07	0.94	0.94	0.95	0.93
Flute	0.98	0.02	0.97	0.03	0.98	0.97	0.98	0.97

Oboe	0.95	0.05	0.94	0.06	0.95	0.94	0.95	0.94
Saxophone	0.85	0.15	0.89	0.11	0.87	0.88	0.85	0.89
Overall					**0.92**			

Table A.3: The Details of DT Classifier Identified for 20 Musical Instruments.

Musical Instrument	**True Positive**	**False Negative**	**True Negative**	**False Positive**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	0.97	0.03	1.00	0.00	0.99	1.00	0.97	1.00
Cello	0.87	0.13	0.88	0.12	0.87	0.88	0.87	0.88
Double Bass	0.95	0.05	0.94	0.06	0.95	0.94	0.95	0.95
Guitar	0.58	0.42	0.98	0.02	0.78	0.97	0.58	0.98
Mandolin	0.98	0.03	0.96	0.04	0.97	0.96	0.96	0.95
Tambura	0.60	0.40	0.90	0.10	0.75	0.86	0.53	1.00
Viola	0.88	0.12	0.86	0.14	0.87	0.86	0.90	0.88
Violin	0.93	0.07	0.91	0.09	0.92	0.91	0.93	0.92
Bass Clarinet	0.95	0.05	0.96	0.04	0.95	0.96	0.96	0.95
French Horn	0.89	0.11	0.91	0.09	0.90	0.91	0.88	0.91
Trombone	0.96	0.04	0.93	0.07	0.95	0.93	0.96	0.94
Trumpet	0.80	0.20	0.94	0.06	0.87	0.93	0.80	0.94
Tuba	0.99	0.01	0.99	0.01	0.99	0.99	0.99	0.99
Bassoon	0.90	0.10	0.92	0.08	0.91	0.91	0.89	0.90
Clarinet	0.88	0.12	0.88	0.12	0.88	0.88	0.87	0.88
Contra Bassoon	0.92	0.08	0.91	0.09	0.91	0.91	0.92	0.88
English Horn	0.92	0.08	0.90	0.10	0.91	0.90	0.91	0.89
Flute	0.98	0.02	0.95	0.05	0.97	0.95	0.99	0.94
Oboe	0.94	0.06	0.91	0.09	0.93	0.91	0.95	0.90

Saxophone	0.82	0.18	0.86	0.14	0.84	0.86	0.85	0.87
Overall					**0.91**			

Table A.4 The Validation Details of 20 Musical Instruments.

DT				kNN				SVM			
Musical Instrument	**Accuracy**	**Error Rate**	**AUC**	**Musical Instrument**	**Accuracy**	**Error Rate**	**AUC**	**Musical Instrument**	**Accuracy**	**Error Rate**	**AUC**
Banjo	0.99	0.01	1	Banjo	0.99	0.01	1	Banjo	0.99	0.01	1.00
Cello	0.87	0.13	0.99	Cello	0.88	0.12	0.92	Cello	0.93	0.07	1.00
Double Bass	0.95	0.05	1	Double Bass	0.94	0.06	0.98	Double Bass	0.97	0.03	1.00
Guitar	0.78	0.22	0.95	Guitar	0.79	0.21	0.83	Guitar	0.87	0.13	0.98
Mandolin	0.97	0.03	1	Mandolin	0.96	0.04	0.97	Mandolin	0.95	0.05	1.00
Tambura	0.75	0.25	0.9	Tambura	0.78	0.22	0.78	Tambura	0.57	0.43	0.99
Viola	0.87	0.13	0.99	Viola	0.88	0.12	0.95	Viola	0.94	0.06	1.00
Violin	0.92	0.08	1	Violin	0.91	0.09	0.95	Violin	0.95	0.05	1.00
Bass Clarinet	0.95	0.05	1	Bass Clarinet	0.96	0.04	0.97	Bass Clarinet	0.97	0.03	1.00
French Horn	0.90	0.10	0.99	French Horn	0.93	0.07	0.96	French Horn	0.96	0.04	1.00
Trombone	0.95	0.05	1	Trombone	0.95	0.05	0.98	Trombone	0.96	0.04	1.00
Trumpet	0.87	0.13	0.99	Trumpet	0.88	0.12	0.93	Trumpet	0.93	0.07	0.99
Tuba	0.99	0.01	1	Tuba	1.00	0.00	1	Tuba	1.00	0.00	1.00
Bassoon	0.91	0.09	0.99	Bassoon	0.95	0.05	0.97	Bassoon	0.96	0.04	1.00
Clarinet	0.88	0.12	0.99	Clarinet	0.91	0.09	0.95	Clarinet	0.93	0.07	1.00
Contra Bassoon	0.91	0.09	1	Contra Bassoon	0.94	0.06	0.97	Contra Bassoon	0.97	0.03	0.99
English Horn	0.91	0.09	1	English Horn	0.94	0.06	0.98	English Horn	0.96	0.04	1.00
Flute	0.97	0.03	1	Flute	0.98	0.02	0.99	Flute	0.99	0.01	1.00
Oboe	0.93	0.07	1	Oboe	0.95	0.05	0.98	Oboe	0.96	0.04	1.00
Saxophone	0.84	0.16	0.99	Saxophone	0.87	0.13	0.93	Saxophone	0.92	0.08	0.99
Overall	**0.91**	**0.09**		**Overall**	**0.92**	**0.08**		**Overall**	**0.93**	**0.07**	

APPENDIX B: THE DETAILS OF THE CONFUSION MATRIX OF CLASSIFIERS

The following Figure A.1, Figure A.2 and Figure A.3 show the confusion matrices for SVM, kNN, and DT Classifiers.

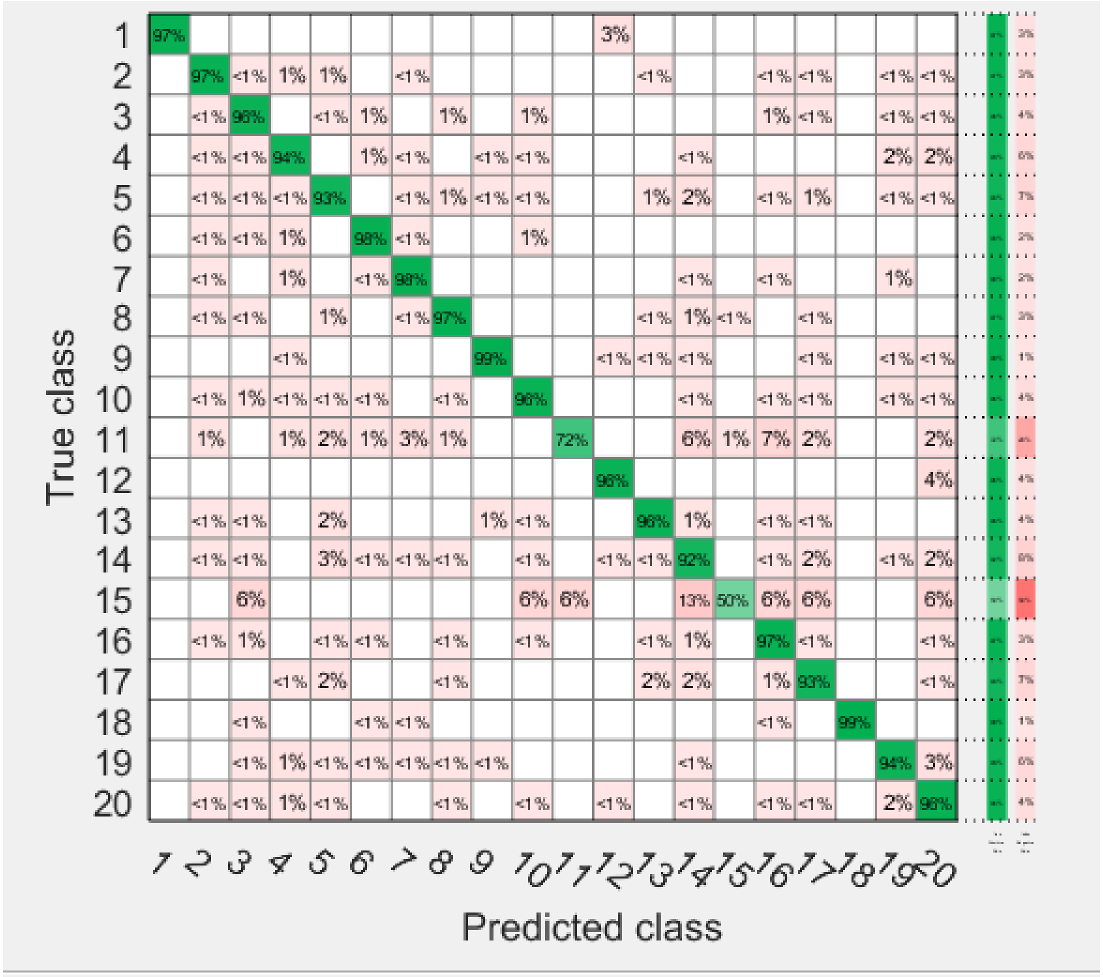

Figure A.1: Confusion Matrix of SVM Classifier.

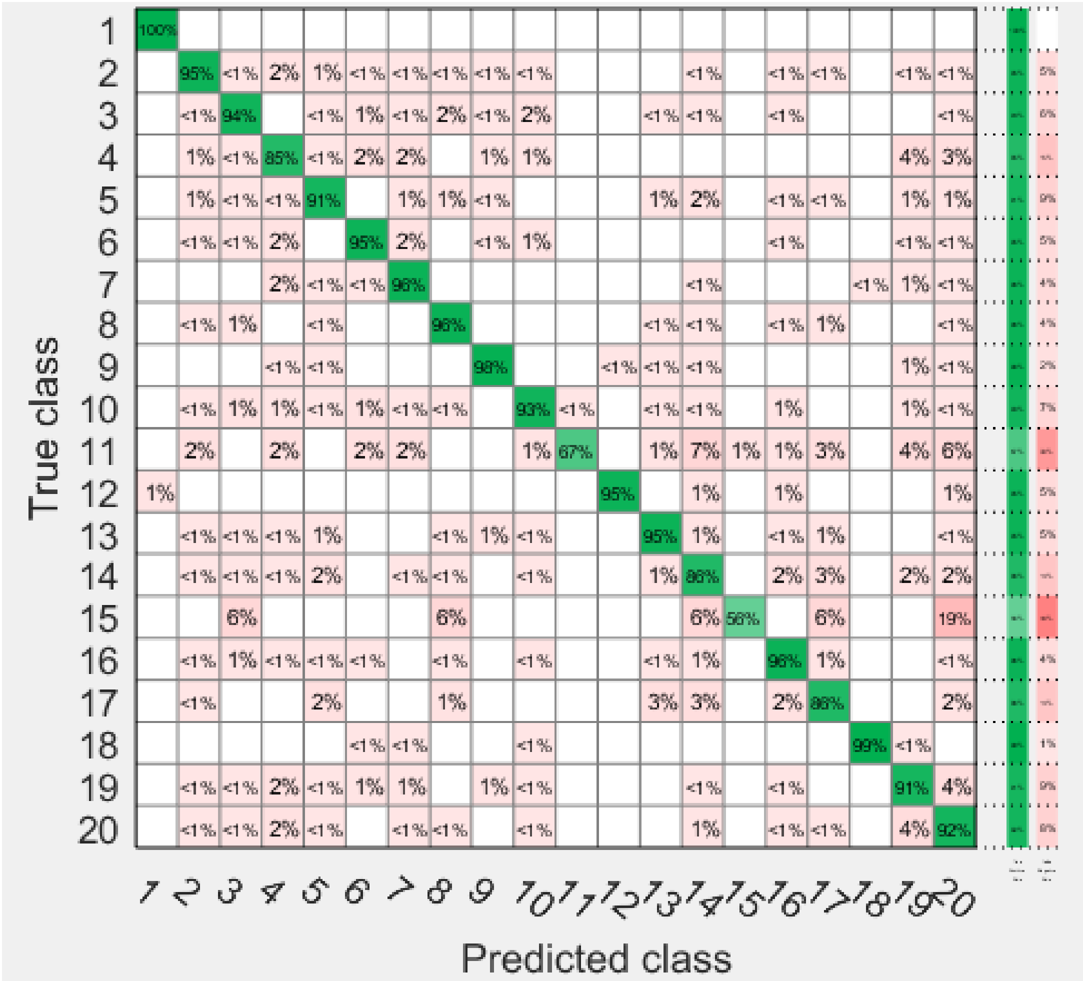

Figure A.2: Confusion Matrix of kNN Classifier

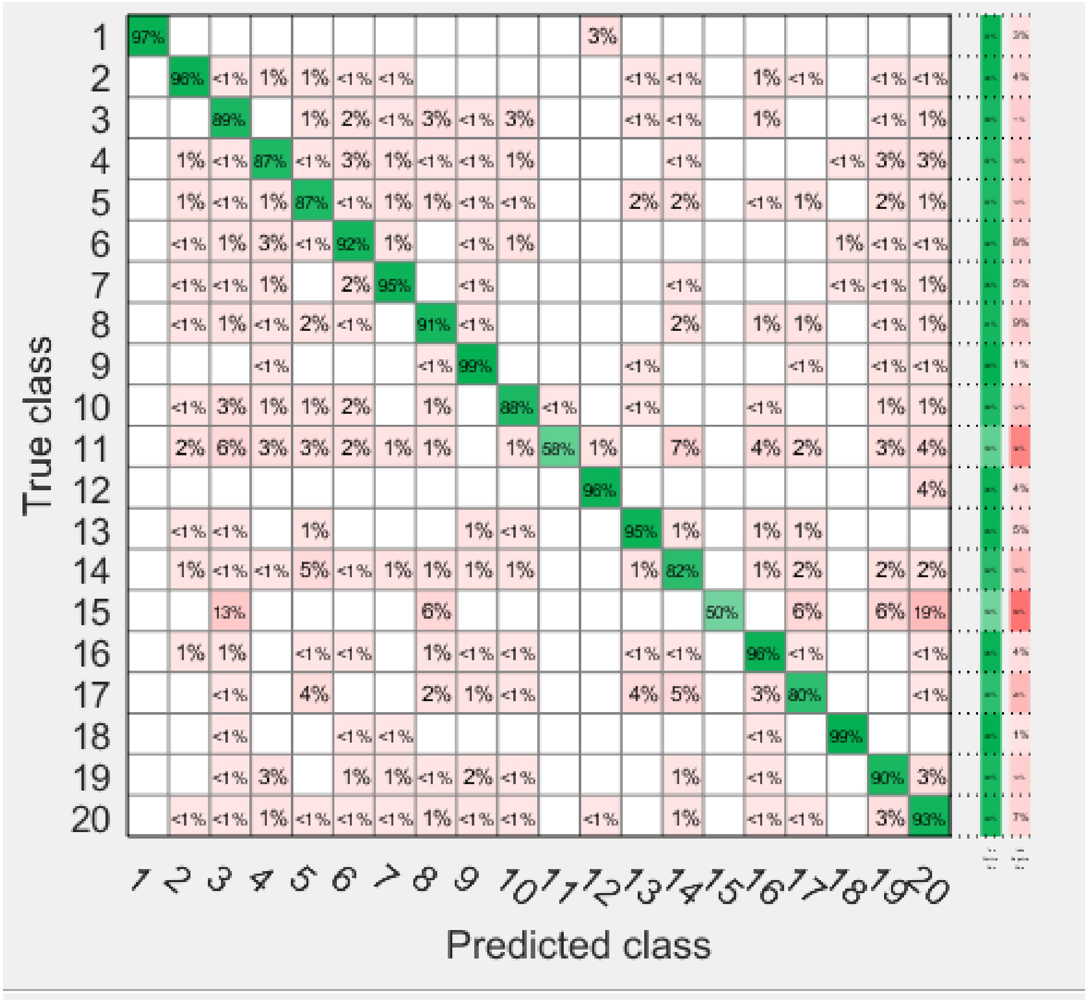

Figure A.3: Confusion Matrix of DT Classifier.

APPENDIX C: MEASURE DETAILS OF DIFFERENT WINDOW SIZES FOR CLASSIFIERS

The following Table A.5, Table A.6, Table A.7, Table A.8, Table A9 and Table A.10 show the accuracy, precision, sensitivity, and specificity of the SVM classifier for different window sizes used for the feature extraction.

Table A.5: Measure Details of Window Size 2048 for SVM Classifier.

Musical Instrument	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	0.97	0.99	0.96	0.99
Cello	0.92	0.92	0.93	0.92
Double Bass	0.98	0.98	0.98	0.98
Guitar	0.85	0.98	0.72	0.98
Mandolin	0.92	0.94	0.90	0.95
Tambura	0.98	0.98	0.98	0.98
Viola	0.93	0.93	0.92	0.93
Violin	0.94	0.93	0.96	0.93
Bass Clarinet	0.96	0.97	0.96	0.97
French Horn	0.94	0.95	0.93	0.95
Trombone	0.97	0.96	0.97	0.96
Trumpet	0.92	0.94	0.91	0.94
Tuba	1.00	1.00	0.99	1.00
Bassoon	0.95	0.95	0.95	0.95
Clarinet	0.91	0.90	0.91	0.90
Contra Bassoon	0.97	0.96	0.97	0.96
English Horn	0.95	0.95	0.95	0.95
Flute	0.98	0.99	0.98	0.99
Oboe	0.96	0.96	0.96	0.96
Saxophone	0.91	0.91	0.90	0.91
Average	**0.9452**			

Table A.6: Measure Details of Window Size 4096 for SVM Classifier.

Musical Instrument	Accuracy	Precision	Sensitivity	Specificity
Banjo	0.99	1.00	0.97	1.00
Cello	0.93	0.93	0.94	0.93
Double Bass	0.98	0.98	0.97	0.98
Guitar	0.85	0.98	0.71	0.98
Mandolin	0.92	0.89	0.96	0.89
Tambura	0.98	0.99	0.98	0.99
Viola	0.95	0.94	0.95	0.94
Violin	0.95	0.95	0.96	0.94
Bass Clarinet	0.97	0.97	0.97	0.97
French Horn	0.96	0.97	0.96	0.97
Trombone	0.97	0.97	0.96	0.97
Trumpet	0.93	0.94	0.93	0.94
Tuba	1.00	1.00	1.00	1.00
Bassoon	0.96	0.96	0.96	0.96
Clarinet	0.93	0.93	0.93	0.93
Contra Bassoon	0.97	0.97	0.98	0.97
English Horn	0.97	0.96	0.97	0.96
Flute	0.99	0.99	0.99	0.99
Oboe	0.97	0.97	0.98	0.97
Saxophone	0.92	0.92	0.92	0.92
Average	**0.9546**			

Table A.7: Measure Details of Window Size 6144 for SVM Classifier.

Musical Instrument	Accuracy	Precision	Sensitivity	Specificity
Banjo	0.97	0.96	0.97	0.96
Cello	0.94	0.95	0.93	0.95
Double Bass	0.98	0.97	0.98	0.97
Guitar	0.85	0.95	0.74	0.96
Mandolin	0.90	0.93	0.88	0.93
Tambura	0.98	0.99	0.97	0.99
Viola	0.94	0.93	0.95	0.92
Violin	0.95	0.94	0.96	0.94
Bass Clarinet	0.98	0.98	0.97	0.98
French Horn	0.96	0.96	0.95	0.96
Trombone	0.96	0.97	0.96	0.97

Trumpet	0.93	0.93	0.93	0.92
Tuba	1.00	1.00	0.99	1.00
Bassoon	0.96	0.96	0.96	0.96
Clarinet	0.93	0.93	0.93	0.93
Contra Bassoon	0.97	0.96	0.98	0.96
English Horn	0.96	0.96	0.96	0.96
Flute	0.99	0.99	0.98	0.99
Oboe	0.97	0.97	0.97	0.97
Saxophone	0.92	0.92	0.93	0.92
Average	**0.9515**			

Table A.8: Measure Details of Window Size 8192 for SVM Classifier.

Musical Instrument	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	0.99	1.00	0.97	1.00
Cello	0.94	0.95	0.93	0.95
Double Bass	0.97	0.97	0.98	0.97
Guitar	0.88	0.93	0.81	0.94
Mandolin	0.91	0.92	0.90	0.92
Tambura	0.98	0.99	0.97	0.99
Viola	0.94	0.95	0.94	0.95
Violin	0.94	0.93	0.96	0.92
Bass Clarinet	0.97	0.97	0.98	0.97
French Horn	0.95	0.96	0.94	0.96
Trombone	0.96	0.96	0.97	0.96
Trumpet	0.92	0.93	0.91	0.93
Tuba	1.00	1.00	0.99	1.00
Bassoon	0.96	0.96	0.96	0.96
Clarinet	0.93	0.93	0.93	0.93
Contra Bassoon	0.96	0.96	0.96	0.96
English Horn	0.96	0.95	0.97	0.95
Flute	0.99	0.99	0.98	0.99
Oboe	0.97	0.97	0.97	0.97
Saxophone	0.93	0.94	0.92	0.94
Average	**0.9532**			

Table A.9: Measure Details of Window Size 10240 for SVM Classifier.

Musical Instrument	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	0.99	0.99	0.99	0.99
Cello	0.94	0.95	0.93	0.95
Double Bass	0.97	0.97	0.98	0.97
Guitar	0.86	0.98	0.74	0.99
Mandolin	0.94	0.96	0.91	0.96
Tambura	0.98	0.99	0.97	0.99
Viola	0.95	0.94	0.95	0.94
Violin	0.95	0.94	0.96	0.93
Bass Clarinet	0.98	0.98	0.97	0.98
French Horn	0.95	0.96	0.95	0.96
Trombone	0.97	0.97	0.97	0.97
Trumpet	0.93	0.93	0.92	0.93
Tuba	1.00	1.00	0.99	1.00
Bassoon	0.96	0.96	0.95	0.96
Clarinet	0.93	0.93	0.93	0.93
Contra Bassoon	0.96	0.95	0.97	0.95
English Horn	0.96	0.95	0.97	0.95
Flute	0.98	0.98	0.98	0.99
Oboe	0.96	0.96	0.97	0.96
Saxophone	0.93	0.93	0.93	0.93
Average	**0.9545**			

Table A.10: Measure Details of Window Size 12288 for SVM Classifier.

Musical Instrument	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
Banjo	0.97	0.96	0.99	0.96
Cello	0.93	0.94	0.93	0.94
Double Bass	0.97	0.96	0.98	0.96
Guitar	0.88	1.00	0.76	1.00
Mandolin	0.92	0.97	0.86	0.97
Tambura	0.98	0.99	0.98	0.99
Viola	0.94	0.94	0.94	0.94
Violin	0.94	0.93	0.96	0.93
Bass Clarinet	0.98	0.98	0.97	0.98
French Horn	0.95	0.96	0.94	0.97
Trombone	0.97	0.97	0.97	0.97
Trumpet	0.93	0.93	0.92	0.94
Tuba	1.00	1.00	0.99	1.00

Bassoon	0.96	0.96	0.95	0.96
Clarinet	0.93	0.93	0.94	0.93
Contra Bassoon	0.96	0.96	0.97	0.96
English Horn	0.96	0.95	0.97	0.95
Flute	0.98	0.98	0.98	0.98
Oboe	0.97	0.96	0.97	0.96
Saxophone	0.94	0.95	0.93	0.95
Average	**0.9533**			