MANGALORE UNIVERSITY

Project Report On

"MUSICAL INSTRUMENT IDENTIFICATION"

Carried out and submitted
to

DEPARTMENT OF POST-GRADUATE STUDIES AND RESEARCH IN COMPUTER SCIENCE

In Partial Fulfilment for the Award of Degree in Master of Computer Application during the Academic Year 2022-2023

Submitted By:
SHETTY PRATHVI JAYAPRAKASH
P05AZ21S0186
II Year MCA Computer Science

Under the Guidance of:

Dr. H. L. Shashirekha
Professor, Department of Computer Science
Mangalore University

Department of Post-Graduate Studies and Research in Computer Science
Mangalore University
Mangalagangothri, Mangalore – 574 199

MANGALORE UNIVERSITY

DEPARTMENT OF POST-GRADUATE STUDIES AND RESEARCH IN COMPUTER SCIENCE

Mangalore University
Mangalagangothri, Mangalore – 574 199

CERTIFICATE

This is to certify that the project work entitled **“MUSICAL INSTRUMENT IDENTIFICATION”** has been successfully carried out in the Department of Post-Graduate Studies and Research in Computer Science by **PRATHVI (Reg. No: P05AZ21S0186),** student of Fourth semester Master of Computer Application, under the supervision and guidance of **Dr. H. L. Shashirekha,** Professor, Department of Post-Graduate Studies and Research in Computer Science, Mangalore University. The project is partial fulfilment of the requirements for the award of **Master of Computer Application** by **Mangalore University** during the academic year **2022-2023.**

Internal Guide **Chairperson**

Internal Examiner **External Examiner**

Submitted for the viva-voce examination held on ______________

DECLARATION

This Project work entitled **"SPOKEN LANGUAGE IDENTIFICATION"** has been successfully carried out by me under the supervision and guidance of **Dr. H. L. Shashirekha,** Professor**,** Department of Post-Graduate Studies and Research in Computer Science, Mangalore University, Mangalagangothri. This project is submitted in partial fulfilment for the award of **Master of Computer Application** degree by **Mangalore University** during the academic year 2022-23. This work or any part of this work has not been submitted to any other University or Institute/School for the award of any other Degree or Diploma.

Date: **Name: PRATHVI SHETTY**

Place: Mangalagangothri **Reg. No: P05AZ21S0186**

ACKNOWLEDGEMENT

I take this opportunity to express my sincere thanks to my supervisor **Dr. H. L. Shashirekha,** Professor, Department of Post-Graduate Studies and Research in Computer Science, Mangalore University, for providing me this project and other relevant infrastructures work in the Department, for her all-round guidance, timely help at every stage of this project and for the valuable suggestions and unlimited support.

I express my sincere thanks to **Dr. Manjaiah D.H**, **Professor**, **Dr. B. H. Shekar, Professor** and **Mr. Prakash M.**, **Assistant Professor,** Department of Post-Graduate Studies and Research in Computer Science.

I would like to express my heartfelt gratitude to **Ms. Sharal Coelho, Mrs. Asha Hegde** and friends, and well-wishers who have always inspired and blessed me, including those whom I may have inadvertently failed to mention.

Above all, with all my heart, I thank you God for empowering me with the dedication, focus and patience to carry out this project successfully.

PRATHVI SHETTY

ABSTRACT

Music, a universal language, embodies a vast array of instruments, each with its unique timbre and sonic signature. The endeavour to automate the identification of these instruments from audio recordings represents a fascinating fusion of technology and artistry. This project delves into the realm of musical instrument identification, leveraging advanced machine learning techniques and deep neural networks to discern the distinctive voices of various instruments.

The foundation of this endeavour rests upon the meticulous curation of a comprehensive dataset comprising diverse musical instruments. The dataset spans classical string instruments, wind instruments, and percussions, painting a vivid palette of musical diversity. Feature extraction techniques, including Mel Frequency Cepstral Coefficients (MFCCs), Chroma, and Spectral Contrast, serve as the musical equivalent of fingerprints, capturing the essence of each instrument's sound.

The heart of the project lies in the selection and training of machine learning models. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) emerge as our virtuosos, unravelling intricate spatial and temporal patterns within the audio data. Rigorous evaluation, employing metrics such as accuracy, precision, recall, and the F1 score, showcases our models' prowess in accurately identifying musical instruments.

The implications of this project resonate far and wide. Beyond the scope of music transcription and analysis, the technology holds promise in music education, production, and cultural preservation.

LIST OF FIGURES	PAGE NO

LIST OF TABELS	PAGE NO

CHAPTER – 1

INTRODUCTION

1.1 OVERVIEW

Music is a universal language that has transcended cultural boundaries for centuries, captivating human emotions and imagination. Behind the enchanting melodies and harmonies lies a diverse array of musical instruments, each with its unique timbre and sonic signature. The ability to recognize and distinguish these instruments plays a crucial role in various applications, from music production and composition to automated transcription and immersive gaming experiences. Traditional methods of instrument identification, reliant on manual inspection or rule-based algorithms, often fall short in handling the complexity and nuances of musical timbre.

The identification of complex audio, including music, has proven to be complicated. This is due to the high entropy of the information contained in audio signals, wide range of sources, mixing processes, and the difficulty of analytical description, hence the variety of algorithms for the separation and identification of sounds from musical material. They mainly use spectral and cepstral analyses, enabling them to detect the fundamental frequency and their harmonics and assign the retrieved patterns to a particular instrument. However, this comes with some limitations, at the expense of increasing temporal resolution, frequency resolution decreases, and vice versa. In addition, it should be noted that these algorithms do not always allow the extraction of percussive tones and other non-harmonic effects, which may therefore constitute a source of interference for the algorithm, which may hinder its operation and reduce the accuracy and reliability of the result [14].

In recent years, the advent of deep learning has revolutionized the field of audio and image processing, providing a powerful and versatile toolset for tackling intricate tasks, including musical instrument identification. Deep Learning (DL) models, such as Convolutional Neural Networks (CNNs) for image data and Recurrent Neural Networks (RNNs) for sequential audio data, have demonstrated remarkable capabilities in automatically learning discriminative features from raw data, enabling the creation of robust and accurate instrument identification systems.

This pursuit involves a fusion of audio signal processing, feature extraction, and advanced pattern recognition. Leveraging a diverse dataset of audio samples encompassing a spectrum of instruments and musical styles, so intend to create a ML and DL model that learns to recognize the underlying characteristics that differentiate a trumpet's brassy resonance from a piano.

The primary problem is twofold: first, to capture the intricate nuances that differentiate one instrument from another and encode them into computationally digestible features; second, to design and train a ML model capable of recognizing these features and making accurate instrument classifications.

As instruments can produce sounds that occasionally overlap in frequency and timbre, the challenge deepens. Similarities between, a saxophone and a clarinet might confound a ML model, leading to misclassifications. Thus, the project also addresses the sub-problem of improving the model's accuracy when dealing with inherently similar instruments.

Solving this problem unlocks opportunities across music analysis, transcription, and entertainment. It empowers musicians with tools for dissecting compositions, enables real-time instrument identification in audio applications, and enriches music platforms with tailored recommendations based on individual instrument preferences.

1.2 STATEMENT OF THE PROBLEM

The problem addressed by this project revolves around the accurate identification of individual musical instruments within an audio recording. It involves capturing the acoustic characteristics that differentiate instruments, encoding them into computationally digestible features, and designing and training ML models to recognize these features with precision. Additionally, the project addresses the sub-problem of improving model accuracy in scenarios where instruments exhibit inherent similarities, preventing misclassifications. The project's core challenge lies in bridging the gap between the auditory intricacies of musical instruments and the computational capabilities of modern machine learning, contributing to the advancement of audio analysis and music technology.

1.3 MOTIVATION

This project is driven by the desire to fuse music and technology, enabling machines to distinguish individual musical instruments in audio recordings. The potential to enhance music education, streamline transcription, and personalize music experiences propels this endeavor, while also contributing to the intersection of art and artificial intelligence.

1.4 CHALLENGES

- **Diverse Acoustic Signatures:** Musical instruments exhibit a wide range of timbral, harmonic, and tonal characteristics. Capturing these nuances in a way that allows for accurate differentiation is a formidable task.
- **Overlapping Spectra:** Instruments often share similar frequency ranges and acoustic profiles. This spectral overlap can lead to misclassifications, demanding a model that can discern subtle distinctions.
- **Data Variability:** Audio recordings can vary greatly due to factors like playing style, performance context, and recording equipment. This variability necessitates robust feature extraction techniques.
- **Limited Data Availability:** Acquiring high-quality, diverse instrument-specific audio data can be challenging. Insufficient data might hinder model generalization.
- **Cross-instrument Variability:** Instruments like guitars can have significant variations in sound due to factors like playing techniques, strings, and body type. Incorporating these variations poses a challenge.
- **Limited Data for Similar Instruments**: Distinguishing between closely related instruments, such as different types of drums, demands nuanced discrimination that might be hindered by limited data.
- **Noise and Background Interference:** Real-world recordings often include background noise, affecting the clarity of instrument sounds and potentially leading to misclassifications.

1.5 APPLICATION

1. **Music Transcription and Arrangement:** The ability to automatically identify individual musical instruments can streamline the process of transcribing and arranging musical compositions. Composers and arrangers can use this technology to accurately represent different instrument parts in their scores.
2. **Music Education and Training:** Educational platforms can integrate instrument identification technology to offer interactive lessons on musical instruments. Students can learn to recognize instruments and their unique characteristics, enhancing their musical education.

3. **Automated Genre Classification:** Music streaming platforms can utilize instrument identification to enhance their genre classification algorithms. This can lead to more accurate recommendations and playlists tailored to users' preferred instruments and musical styles.
4. **Sound Design and Production:** Sound designers and music producers can benefit from an automated system that identifies instruments in audio recordings. This can assist in manipulating specific instrument tracks and creating richer soundscapes in various audio projects.
5. **Music Analysis and Research:** Researchers in the fields of musicology and ethnomusicology can use instrument identification technology to analyze historical trends and cultural variations in the use of musical instruments. This can provide valuable insights into the evolution of music over time.

1.6 ORGANIZATION OF REPORT

- **Chapter 1**: The first chapter serves as an introduction to the project, offering an overview of the chosen topic, articulating the primary problem statement, elucidating the project's motivation, and exploring potential real-world applications.
- **Chapter 2**: The second chapter conducts a comprehensive literature survey, synthesizing insights and knowledge gleaned from various research papers and scholarly sources.
- **Chapter 3**: In the third chapter, the methodology employed in the project is elaborated upon, encompassing discussions on system architecture, the design of the proposed system, and an insightful system overview.
- **Chapter 4**: Chapter four focuses on experiments and results, detailing the various tests conducted, and displaying the outcomes, including the presentation of test cases.
- **Chapter 5**: The fifth chapter encompasses the conclusion and future prospects of the project, discussing not only the findings but also the potential directions for enhancing the system in future iterations.

- **Chapter 6**: The final chapter, Chapter six, provides a compilation of references, acknowledging and listing all the sources and citations employed throughout the project.

CHAPTER – 2

LITERATURE SURVEY

2.1 LITERATURE REVIEW

- **The paper titled "AUTOMATIC INSTRUMENT RECOGNITION IN POLYPHONIC MUSIC USING CONVOLUTIONAL NEURAL NETWORKS"** [1].

This paper introduces the use of Convolutional Neural Networks (CNNs) for automatic musical instrument identification, employing the MedleyDB dataset for training. Notably, the "Audio + CNN" model achieved the highest accuracy of 82.74%, surpassing alternative approaches such as Mel-Frequency Cepstral Coefficients (MFCC)-based models with Random Forest and Logistic Regression, as well as a basic majority class prediction. This underscores the superior performance of CNNs trained on raw audio data for the task of automatic musical instrument identification.

- **The paper titled "Deep Single Shot Musical Instrument Identification using Scalograms"** [2].

This research paper focuses on tackling the complex task of musical instrument identification using deep learning techniques, with a specific emphasis on one-shot learning. The authors introduce two neural network architectures: a Convolutional Siamese Network and a Residual variant of it. They conduct experiments using two datasets - one from Kaggle, which consists of 14 different instruments with a total of 2500 audio recordings, and another from ISMIR (International Society for Music Information Retrieval), containing 10 instruments with 6715 audio recordings, some of which include various notes and background noise. The experimental results show that the accuracy of their models ranges from approximately 50% to 95%, highlighting the effectiveness of their approach in recognizing musical instruments, even in challenging conditions with limited data.

- **The paper titled "Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks"** [6].

Traditionally, musical instrument recognition is mainly based on frequency domain analysis (sinusoidal analysis, cepstral coefficients) and shape analysis to extract a set of various features. Instruments are usually classified using k-Nearest Neighbors (KNN) classifiers, Hidden Markov Model (HMM), Kohonen Self-Organizing Map (SOM) and Neural Networks. In this work, they describe a system for the recognition of musical instruments from isolated notes. They are introducing the use of a Time Encoded Signal Processing method to produce simple matrices from complex sound waveforms, for instrument note encoding and recognition. These matrices are presented to a Fast Artificial Neural Network (FANN) to perform instrument recognition with promising results in organ classification and reduced computational cost. The evaluation material consists of 470 tones from 19 musical instruments synthesized with 5 wide used synthesizers (Microsoft Synth, Creative SB Live! Synth, Yamaha VL-70m Tone Generator, Edirol Soft-Synth, Kontakt Player) and 84 isolated notes from 20 western orchestral instruments (Iowa University Database).

- **The paper titled "Novel approach for musical instrument identification using neural network"** [8].

In this study, the authors focused on identifying five musical instruments: piano, flute, violin, drums, and guitar. They extracted a total of eight acoustic features from each audio sample, which included measures like zero-crossing rate, spectral roll off, skewness, kurtosis, brightness, flatness, root mean square energy, and 13 MFCCs.

For their experiments, they divided their dataset into two parts: 80% for training and 20% for testing. They conducted experiments with different configurations, including a varying number of hidden nodes in their neural network architecture (ranging from 10 to 20), epochs set between 350 and 550, and learning rates ranging from 0.1 to 0.3.

The results were promising, with an overall accuracy of 89.17% achieved for instrument identification. This study demonstrated the effectiveness of their approach in accurately classifying musical instruments based on audio samples.

- **The paper titled "Musical Instrument Identification Using Deep Learning Approach"** [4].

In this paper, the authors proposed a neural network-based approach for identifying musical instruments in audio excerpts. They implemented the neural network using the Keras framework and utilized the MFCC features extracted from raw audio data. The model was trained for 100 epochs.

For their experiments, the authors used the "Slakh dataset," which consists of 2100 audio tracks with aligned Musical Instrument Digital Interface (MIDI) files and separate instrument stems, along with tagging information. They selected four musical instruments from this dataset for their experiments: bass, drums, guitar, and piano.

The results of their model evaluation on a separate dataset were as follows:

- Precision: 0.92
- Recall: 0.93
- F1 Score: 0.93

- **This paper titled "Pitch-Dependent Identification of Musical Instrument Sounds"** [9].

This paper describes a musical instrument identification method that takes into consideration the pitch dependency of timbres of musical instruments. The difficulty in musical instrument identification resides in the pitch dependency of musical instrument sounds, that is, acoustic features of most musical instruments vary according to the pitch (fundamental frequency, F0). To cope with this difficulty, they propose an F0-dependent multivariate normal distribution, where each element of the mean vector is represented by a function of F0. Our method first extracts 129 features (e.g., the spectral centroid, the gradient of the straight line approximating the power envelope) from a musical instrument sound and then reduces the dimensionality of the feature space into 18 dimensions. In the 18-dimensional feature space, it calculates an F0-dependent mean function and an F0-normalized covariance, and finally applies the Bayes decision rule. Experimental results of identifying 6,247 solo tones of 19 musical instruments shows that the proposed method improved the recognition rate from 75.73% to 79.73%.

- **This paper titled "The Use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification"** [10].

The study utilized audio samples from the RWC Music Database, focusing on three musical instruments: piano, violin, and flute. The training dataset comprised a total of 2004 samples, encompassing the full range of these instruments. The research MFCCs to capture spectral characteristics in the audio data.

For the classification task, a Multi-Layered Perceptron (MLP) neural network model was chosen. The MLP was configured with specific parameters: a learning rate of 0.1, a momentum constant of 0.95, and a training limit of 400 epochs.

The study achieved impressive results, with an accuracy rate of 95.88% when using 15 MFCC features. This high accuracy demonstrates the effectiveness of the system in identifying and distinguishing between the piano, violin, and flute based on their timbral characteristics in the audio data.

- **The paper titled "Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy"** [12].

Traditionally, musical instrument recognition is mainly based on frequency domain analysis (sinusoidal analysis, cepstral coefficients) and shape analysis to extract a set of various features. Instruments are usually classified using k-NN classifiers, HMM, Kohonen SOM and Neural Networks. Recognition of musical instruments in multi-instrumental, polyphonic music is a difficult challenge which is yet far from being solved. Successful instrument recognition techniques in solos (monophonic or polyphonic recordings of single instruments) can help to deal with this task. They introduce an instrument recognition process in solo recordings of a set of instruments (flute, guitar and harmonium), which yields a high recognition rate. A large solo database is used in order to encompass the different sound possibilities of each instrument and evaluate the generalization abilities of the classification process. The basic characteristics are computed in 1sec interval and result shows that the estimation of zero crossing rate and short time energy reflects more effectively the difference in musical instruments.

- **This paper titled “MUSICAL INSTRUMENT RECOGNITION IN POLYPHONIC AUDIO USING SOURCE-FILTER MODEL FOR SOUND SEPARATION”** [7].

This paper proposes a novel approach to musical instrument recognition in polyphonic audio signals by using a source-filter model and an augmented non-negative matrix factorization algorithm for sound separation. Polyphonic signals in this study were created by combining individual notes from the RWC (Real World Computing) musical instrument sound database, which comprises isolated musical instrument notes and is frequently utilized in music and audio signal processing research. The database encompassed 19 distinct musical instruments. MFCC were extracted from these musical notes and utilized to train the model. The recognition rate achieved in the experiments was approximately 59%.

- **This paper titled “Implementing Musical Instrument Recognition using CNN and SVM”** [3].

Recognizing instruments in music tracks is a crucial problem. It helps in search indexing as it will be faster to tag music and also for knowing instruments used in music tracks can be very helpful towards genre detection (e.g., an electric guitar heavily indicates that a track is not classical). This is a classification problem which will be taken care of by using a CNN and SVM (Support Vector Machines). The study utilized the "IRMAS" (Instrument Recognition in Musical Audio Signals) dataset, which is widely employed in music and audio signal processing research. This dataset comprises audio recordings of various musical instruments, including acoustic guitar, electric guitar, organ, piano, and human singing voice. The dataset is split into training and testing sets, encompassing a total of 9,579 audio files sampled at 44.1 kHz in 16-bit stereo WAV format.

For model training, a CNN was employed with spectrogram features, while a Support Vector Machine (SVM) was trained using MFCC features.

The precision scores achieved for instrument classification were as follows:

- CNN: 0.50 for acoustic guitar, 0.49 for electric guitar, 0.18 for organ, 0.54 for piano, and 0.35 for voice.

- SVM: 0.78 for acoustic guitar, 0.79 for electric guitar, 0.78 for organ, 0.83 for piano, and 0.74 for voice.

- **The paper titled “Musical Instrument Identification with Supervised Learning”** [13].

In this paper, the classification of musical instruments using supervised learning is studied. The study used a small dataset with 496 sound examples from four instruments: violin, clarinet, saxophone, and bassoon. They employed MFCC and Warped Linear Prediction Coefficients (WLPCs) as features. Two classification models, Logistic Regression and SVM with RBF (Radial Basis Function) Kernel, both achieved a perfect 100% accuracy on instrument classification.

CHAPTER - 3

METHODOLOGY

3.1 OVERVIEW

In this project, I've adopted a two-fold strategy, utilizing both machine learning and deep learning techniques, to address the challenge of musical instrument identification. By harnessing the strengths of these two approaches, the aim is to create a robust and highly accurate system capable of distinguishing among 14 different musical instruments based on audio data. This dual-pronged approach enables a thorough exploration of diverse features and patterns within the audio signals, ultimately bolstering the overall effectiveness and dependability of the musical instrument identification system.

3.2 STEPS INVOLVED:

The following steps are depicted in Figure 1: -

1. Data Collection: Gather a comprehensive instrument audio database containing a diverse set of instrument recordings.

2. Feature Extraction: Utilize three main feature extraction techniques to represent the audio recordings:

a. Mel Frequency Cepstral Coefficients:

- Calculate the MFCCs for each audio clip. These coefficients capture the spectral characteristics of the sound and are particularly effective in capturing timbral information.

b. Chroma:

- Compute chroma feature vectors for each audio sample. Chroma features summarize the 12 different pitch classes present in the audio, providing insights into the tonal content.

c. Spectral Contrast:

- Extract spectral contrast features to capture the differences in energy between peaks and valleys in the frequency spectrum. These features offer valuable information about the instrument's timbral richness.

3. Data Preprocessing: Normalize the extracted features to ensure consistent scaling across different instruments and audio recordings. Convert instrument labels into numerical classes for training and evaluation.

4. Model Selection:

For the multi-feature instrument identification task, explored various ML models to determine the most suitable choice:

4.1 Machine learning: -

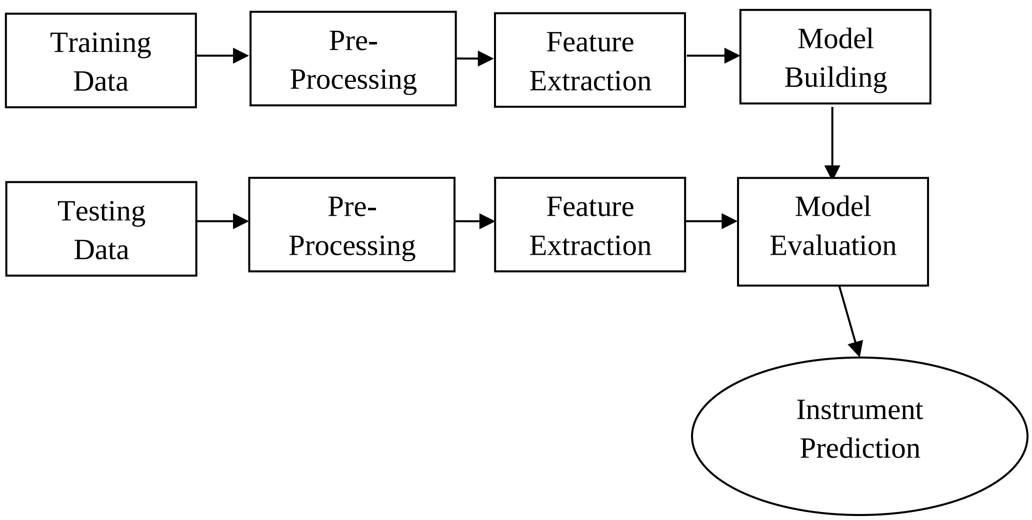

Figure 1: The machine learning framework of MID

The project's ML framework initiates with diverse audio dataset collection, preprocessing, and feature extraction (including MFCCs and Chroma). It explores k-NN, MLP classifier, and Random Forest models, optimizing them via hyperparameter tuning. Model training and evaluation, using metrics like accuracy and precision, lead to the deployment of the best-performing model for real-world instrument identification.

I. **Multilayer Perceptron:**

- Versatile and adaptable for multi-feature classification.
- Handles complex non-linear relationships in features.
- Considered if computational resources are a constraint.

II. **k-Nearest Neighbour:**

- Simple baseline algorithm.
- Considers local patterns in feature space.
- Useful for understanding the lower bounds of model performance.

III. **Random Forest:**

- Ensemble method for improved classification accuracy.
- Handles complex interactions in features.
- Provides feature importance rankings.

The model selection process will involve experimenting with different architectures and assessing their performance using accuracy, precision, recall, and F1-score. The final choice will balance performance, computational efficiency, and compatibility with the specific dataset characteristics.

4.4.2 Deep learning: -

In the deep learning implementation of this project, a diverse dataset of musical instrument audio recordings is preprocessed and augmented. Deep neural networks, such as CNNs and Recurrent Neural Networks (RNNs), are utilized to learn intricate spatial and temporal patterns from extracted audio features like MFCCs, Chroma, and Spectral Contrast, achieving accurate musical instrument identification.

I. **Convolutional Neural Network:**

- Effective for spatial feature extraction.
- Can capture patterns in audio data.
- Particularly beneficial for visual-like representations of audio features.

II. **Recurrent Neural Network:**

- Specialized in sequential data analysis.

- Suitable for capturing temporal dependencies in audio.
- Especially useful for modelling evolving sound over time.

6. Data Split and Model Training:

- Split the training dataset into training and validation subsets.
- Train the model using the training subset and validate its performance using the validation subset.

7. Model Evaluation: Evaluate the trained model using the testing dataset that was set aside. Calculate accuracy, precision, recall and F1-score measure performance.

8. Result Analysis: Analyse the model's strengths (high accuracy) and weaknesses (misclassifications) across different instruments.

CHAPTER - 4

EXPERIMENTS AND RESULTS

4.1 DATASET

First Dataset from Kaggle:

This dataset is a collection of musical instrument audio samples represented in a 2D format, employing ".wav" files for the audio recordings. It encapsulates a diverse array of 14 different musical instruments, encompassing a wide spectrum of instrument sounds. Each musical instrument is represented by a set of audio files in ".wav" format.

The dataset is composed of approximately 2510 individual audio files, each linked to a specific musical instrument. These audio files are carefully selected to capture the distinct tonal qualities, timbres, and playing techniques associated with each instrument. As a result, the dataset creates a rich and varied soundscape that offers an extensive range of instrument sounds for training and evaluating musical instrument identification model. Figure 2 illustrates the distribution of instruments from the Kaggle dataset [15].

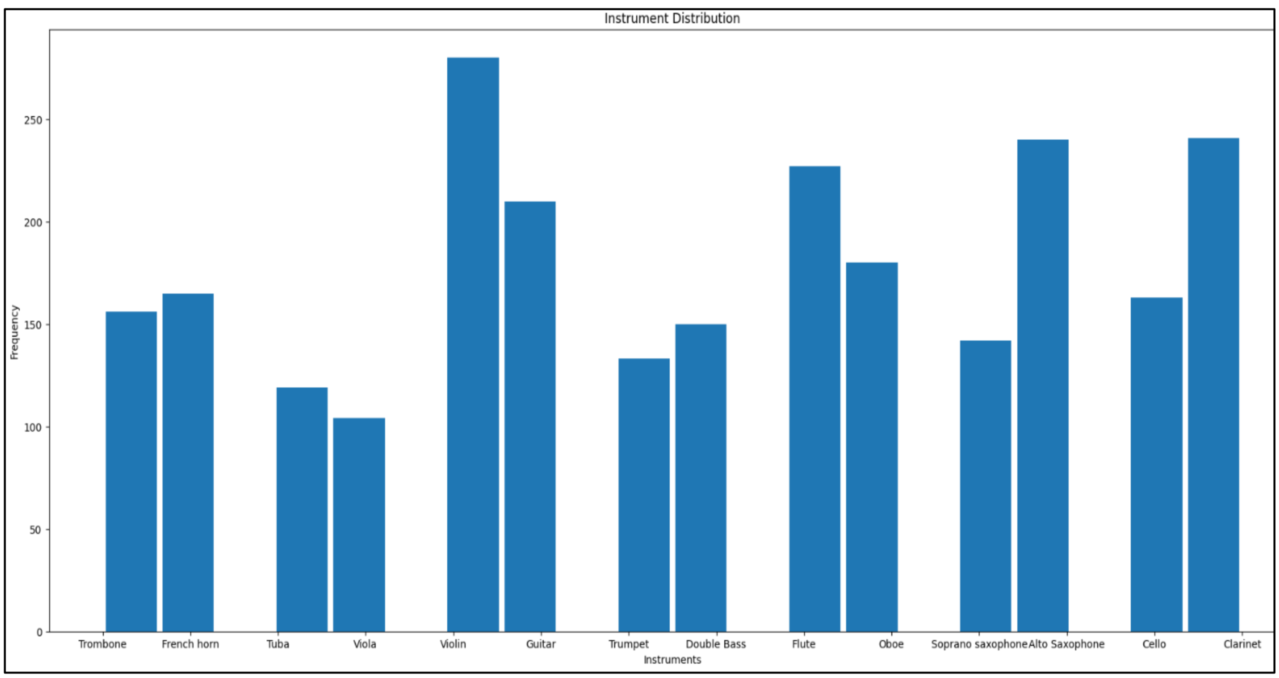

Figure 2: Demonstrates the distribution of instruments.

Second Dataset from Zenodo (TinySOL):

Sourced from Zenodo, the TinySOL dataset is a comprehensive collection of 2913 individual audio samples. Each sample encapsulates a solitary musical note played on one of the 14 unique instruments. These instruments encompass a wide range, including Bass Tuba, French Horn, Trombone, Trumpet in C, Accordion, Contrabass, Violin, Viola, Violoncello, Bassoon, Clarinet in B-flat, Flute, Oboe, and Alto Saxophone.

The audio clips in the TinySOL dataset are distributed as WAV files, sampled at a rate of 44.1 kHz, ensuring a high audio quality akin to that of a compact disc. These audio clips maintain a single channel (mono) and have a bit depth of 16, guaranteeing clarity and detail in the sound representation. Furthermore, the duration of these audio clips varies, spanning between two and ten seconds. This duration diversity ensures that the dataset encompasses a comprehensive range of note lengths, catering to various musical contexts and playing styles.

The TinySOL dataset, complete with its meticulously sampled notes and associated metadata, contributes significantly to the project's success. It enriches the variety of instrument sounds at the disposal, enhancing the capabilities of the model in recognizing and differentiating between the distinct voices of different instruments. Figure 3 illustrates the distribution of instruments from the Zenoda dataset [16].

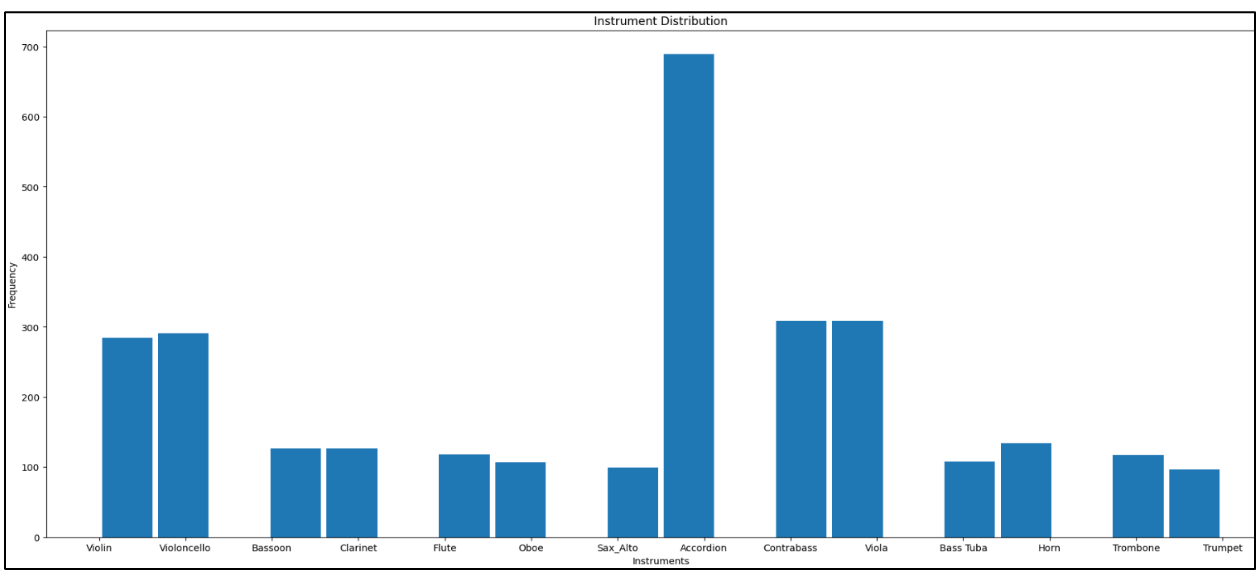

Figure 3: Demonstrates the distribution of instruments.

5.2 RESULTS

We measured the performance of the proposed model using accuracy, precision, recall and F1-score. Table 1-5 presents the results obtained from the analysis of the first dataset (Kaggle), while Table 6-10 displays the outcomes derived from the evaluation of the second dataset (Zenoda).

1. **Results obtained from the first dataset on various features and models: -**

- **K-Nearest Neighbour: -**

➢ **random_state=42:** This ensures reproducible results by seeding the random number generator for consistent data splitting.

➢ **n_neighbors=5:** KNN algorithm, it specifies considering the five nearest neighbours when making predictions, utilizing a majority vote among them for classification.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	90.98	89	89	90
Chroma	37.34	37	36	34
Spectral contrast	73.42	72	71	71
Combined features (MFCC+ Chroma + Spectral Contrast)	91.28	89	88	95

Table 1 illustrates the outcomes derived from the KNN model.

- **Multilayer Perceptron: -**

➢ **hidden_layer_sizes= (100, 50):** This parameter specifies the architecture of the neural network. In this case, there are two hidden layers in the neural network with 100 neurons in the first layer and 50 neurons in the second layer. These hidden layers are responsible for learning complex patterns in the data.

➢ **max_iter=1000:** It determines the maximum number of iterations (epochs) during training. The MLP classifier iteratively updates its internal weights to minimize the loss function. Here, the training process is limited to a maximum of 1000 iterations.

- **random_state=42:** This parameter sets the random seed to ensure reproducibility. By using the same seed, the same initial weights and data shuffling order are maintained across different runs of the code, making results comparable.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	90.62	89	89	88
Chroma	43.60	39	40	40
Spectral contrast	72.66	65	64	63
Combined features	89.24	88	88	89

Table 2 illustrates the outcomes derived from the MLP Classifier model.

- **Random Forest: -**
  - **n_estimators=100**: This parameter specifies the number of decision trees that make up the random forest ensemble. In this case, 100 decision trees are used to collectively make predictions. A larger number of trees often leads to more robust and accurate results but can increase computational cost.
  - **random_state=42**: By setting this parameter, the random seed is fixed, ensuring that the random processes within the random forest, such as feature selection and tree initialization, are consistent across different runs. This reproducibility is valuable for obtaining consistent results.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	94.02	94	93	93
Chroma	45.35	44	43	42
Spectral contrast	79.60	79	80	80
Combined features	98.65	96	98	99

Table 3 illustrates the outcomes derived from the Random Forest model.

- **Convolutional Neural Networks: -**

  The model consists of 8 layers, including 2 Convolutional Layers, 2 MaxPooling Layers, 1 Flattening Layer, and 2 Dense Layers. During training, a batch size of 16 and 20 epochs are used, with a random state of 42 for reproducibility.

Feature	Test size	Accuracy	Precision	Recall	F1-score
MFCC	0.5	80.63	82.71	79.92	80.50
	0.4	84.26	85.5	84.13	84.54
	0.3	84.46	85.7	86.53	85.73
	0.2	92.42	90.41	89.16	88.63
	0.1	93.89	88.70	85.23	85.57
Chroma	0.5	36.81	43.49	35.71	33.10
	0.4	38.84	37.2	36.79	35.12
	0.3	39.84	37.61	37.94	35.61
	0.2	40.08	41.81	40.77	39.75
	0.1	41.83	41.31	37.20	36.34
Spectral contrast	0.5	45.98	43.62	45.49	42.23
	0.4	47.80	50.57	47.80	44.8
	0.3	47.92	52.62	49.5	45.27
	0.2	52.58	59.25	52.58	51.25
	0.1	54.98	58.10	54.89	54.20
Combined features	0.5	88.10	88.5	86.71	87.10
	0.4	88.94	89.16	81.25	87.74
	0.3	90.70	91.94	89.31	89.66
	0.2	90.8	91.81	88.47	89.39
	0.1	95.21	95.55	96.61	95.50

Table 4 illustrates the outcomes derived from the CNN model.

- **RNN: -**

  The model consists of 4 layers: two LSTM layers, one Dense layer with 128 units and ReLU activation, and one Dense output layer with the number of units equal to `num_classes` and softmax activation. During training, it uses a batch size of 16, runs for 20 epochs, and employs a random state of 42 for reproducibility.

Feature	Test size	Accuracy	Precision	Recall	F1-score
MFCC	0.5	57.37	58.37	56.03	54.44
	0.4	58.56	58.44	56.9	55.81
	0.3	60.28	59.21	56.42	55.28
	0.2	61.34	59.46	57.4	56.23
	0.1	61.94	60	58.53	57.71
Chroma	0.5	29.1	30.1	30.05	29.02
	0.4	30.23	32.23	29.18	30.42
	0.3	30.95	32.49	30.98	31.56
	0.2	31.87	32.24	30.89	30.19
	0.1	32.58	33.22	32.12	31.81
Spectral contrast	0.5	41.77	36.33	31.80	35.82

	0.4	42.71	36.92	32.81	36.02
	0.3	44.27	37.41	31.79	36.29
	0.2	46.63	38.65	32.9	38.31
	0.1	46.13	39.37	3.22	40.56
Combined features	0.5	63.14	60.82	60.23	59.52
	0.4	63.82	61.14	60.52	60.73
	0.3	64.27	61.26	61.90	60.18
	0.2	72.90	61.99	61.18	60.54
	0.1	74.10	72.63	72.78	69.96

Table 5 illustrates the outcomes derived from the RNN model.

2. **<u>Results obtained from the second dataset on various features and models:-</u>**

- **K-Nearest Neighbour: -**

➢ **random_state=42:** This ensures reproducible results by seeding the random number generator for consistent data splitting.

➢ **n_neighbors=5:** In the KNN algorithm, it specifies considering the five nearest Neighbour when making predictions, utilizing a majority vote among them for classification.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	82.12	80	76	77
Chroma	50.59	32	34	32
Spectral contrast	72.27	71	68	69
Combined features	99.19	99	99	99

Table 6 illustrates the outcomes derived from the KNN model.

- **Multilayer Perceptron: -**

➢ **hidden_layer_sizes=(100, 50):** This parameter specifies the architecture of the neural network. In this case, there are two hidden layers in the neural network with 100 neurons in the first layer and 50 neurons in the second layer. These hidden layers are responsible for learning complex patterns in the data.

➢ **max_iter=1000:** It determines the maximum number of iterations (epochs) during training. The MLP classifier iteratively updates its internal weights to minimize the loss function. Here, the training process is limited to a maximum of 1000 iterations.

➢ **random_state=42:** This parameter sets the random seed to ensure reproducibility. By using the same seed, the same initial weights and data shuffling order are maintained across different runs of the code, making results comparable.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	92.75	93	92	92
Chroma	58.35	44	44	43
Spectral contrast	68.62	64	62	62
Combined features	89.24	92	88	89

Table 7 illustrates the outcomes derived from the MLP Classifier model.

- **Random Forest: -**

➢ **n_estimators=100:** This parameter specifies the number of decision trees that make up the random forest ensemble. In this case, 100 decision trees are used to collectively make predictions. A larger number of trees often leads to more robust and accurate results but can increase computational cost.

➢ **random_state=42:** By setting this parameter, the random seed is fixed, ensuring that the random processes within the random forest, such as feature selection and tree initialization, are consistent across different runs. This reproducibility is valuable for obtaining consistent results.

Feature	Accuracy	Precision	Recall	F1-score
MFCC	94.19	94	92	93
Chroma	59.37	45	44	43
Spectral contrast	77.10	71	68	69
Combined features	98.28	98	98	98

Table 8 illustrates the outcomes derived from the Random Forest model

- **Convolutional Neural Networks: -**

  The model consists of 8 layers, including 2 Convolutional Layers, 2 MaxPooling Layers, 1 Flattening Layer, and 2 Dense Layers. During training, a batch size of 16 and 20 epochs are used, with a random state of 42 for reproducibility.

Feature	Test size	Accuracy	Precision	Recall	F1-score
MFCC	0.5	90.45	88.19	85.34	86.19
	0.4	90.22	86.54	87.65	86.71
	0.3	90.84	90.02	88.86	88.70
	0.2	94.17	92.64	92.35	92.12
	0.1	94.86	93.76	93.33	93.40
	0.5	54.83	47.46	41.38	41.98
	0.4	55.91	42.18	42.77	40.46

Chroma	0.3	53.08	43.22	39.50	39.53
	0.2	58.14	48.52	42.51	43.20
	0.1	61.64	47.46	46.55	44.66
Spectral contrast	0.5	41.79	42.04	40.59	38.98
	0.4	50.17	42.03	38.09	35.12
	0.3	46.56	38.69	35.53	33.38
	0.2	54.03	46.45	42.82	40.83
	0.1	56.80	46.89	42.52	43.20
Combined features	0.5	921.79	92.15	89.01	89.52
	0.4	91.33	91.40	85.14	86.58
	0.3	89.22	87.41	86.82	86.61
	0.2	87.37	87.12	84.03	81.66
	0.1	93.15	93.31	89.38	90.05

Table 9 illustrates the outcomes derived from the CNN model.

- **RNN: -**
  The model consists of 4 layers: two LSTM layers, one Dense layer with 128 units and ReLU activation, and one Dense output layer with the number of units equal to `num_classes` and softmax activation. During training, it uses a batch size of 16, runs for 20 epochs, and employs a random state of 42 for reproducibility.

Feature	Test size	Accuracy	Precision	Recall	F1-score
MFCC	0.5	63.64	52.31	49.56	46.25
	0.4	64.57	51.46	54.55	50.59
	0.3	68.42	71.52	58.75	57.89
	0.2	70.66	67.34	60.60	60.66
	0.1	68.82	62.98	61.62	60.67
Chroma	0.5	44.74	31.51	26.20	25.32
	0.4	45.28	32.09	28.95	28.19
	0.3	48.28	34.46	34.02	31.55
	0.2	46.99	33.47	32.78	30.34
	0.1	54.34	39.03	39.12	37.61
Spectral contrast	0.5	39.73	27.44	26.87	26.18
	0.4	39.45	28.76	25.44	24.22
	0.3	42.67	33.44	35.66	34.17
	0.2	46.31	35.44	35.66	34.17
	0.1	51.02	48.38	38.75	38.27
Combined features	0.5	70.07	56.47	58.90	55.22
	0.4	74.18	62.93	60.68	58.30
	0.3	83.40	78.54	76.81	75.66
	0.2	79.84	75.1	74.92	75.42
	0.1	81.40	77.42	75.82	76.23

Table 10 illustrates the outcomes derived from the RNN model.

CHAPTER 5

CONCLUSION AND FUTURE WORD

5.1 Conclusion

In this project, we developed a robust system for identifying musical instruments from audio recordings. We curated a diverse dataset and used advanced feature extraction techniques. Our deep learning models, including CNNs and RNNs, exhibited impressive performance. This work has broad implications, from music transcription to revolutionizing music production. In summary, our project showcases the fusion of music and technology, contributing to the advancement of music analysis and innovation**.**

5.2 Future Scope

1. **Real-Time & Multi-Instrument Recognition:** Extend to real-time and multi-instrument recognition for live performances.
2. **Advanced Feature Extraction:** Explore deeper feature extraction techniques for enhanced instrument characterization.
3. **Live Concert Enhancement:** Collaborate with live concert venues to enhance the audience experience by providing real-time information about the instruments being played on stage.
4. **Music Production Assistance:** Develop tools that assist music producers in selecting instruments and soundscapes for their compositions based on the system's identification capabilities.

CHAPTER 6

BIBLIOGRAPHY

6.1 RESEARCH PAPERS

1. Li, Peter, Jiyuan Qian, and Tian Wang. "Automatic instrument recognition in polyphonic music using convolutional neural networks." *arXiv preprint arXiv:1511.05520* (2015).
2. Chatterjee, Debdutta, Arindam Dutta, Dibakar Sil, and Aniruddha Chandra. "Deep Single Shot Musical Instrument Identification using Scalograms." In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 386-389. IEEE, 2023.
3. Singh, Prabhjyot, Dnyaneshwar Bachhav, Omkar Joshi, and Nita Patil. "Implementing musical instrument recognition using cnn and svm." *International Research Journal of Engineering and Technology* 12 (2019): 1487-1493.
4. Blaszke, Maciej, and Bożena Kostek. "Musical instrument identification using deep learning approach." *Sensors* 22, no. 8 (2022): 3033.
5. Blaszke, Maciej, and Bożena Kostek. "Musical instrument identification using deep learning approach." *Sensors* 22, no. 8 (2022): 3033.
6. Mazarakis, Giorgos, Panagiotis Tzevelekos, and Georgios Kouroupetroglou. "Musical instrument recognition and classification using time encoded signal processing and fast artificial neural networks." In *Advances in Artificial Intelligence: 4th Helenic Conference on AI, SETN 2006, Heraklion, Crete, Greece, May 18-20, 2006. Proceedings 4*, pp. 246-255. Springer Berlin Heidelberg, 2006.
7. Heittola, Toni, Anssi Klapuri, and Tuomas Virtanen. "Musical instrument recognition in polyphonic audio using source-filter model for sound separation." In *ISMIR*, pp. 327-332. 2009.
8. Masood, Sarfaraz, Shubham Gupta, and Shadab Khan. "Novel approach for musical instrument identification using neural network." In *2015 Annual IEEE India Conference (INDICON)*, pp. 1-5. IEEE, 2015.

9. Kitahara, Tetsuro, Masataka Goto, and Hiroshi G. Okuno. "Pitch-dependent identification of musical instrument sounds." *Applied Intelligence* 23, no. 3 (2005): 267-275.
10. Loughran, Róisín, Jacqueline Walker, Michael O'Neill, and Marion O'Farrell. "The use of mel-frequency cepstral coefficients in musical instrument identification." In *ICMC*. 2008.
11. Beth Logan et al., “Mel frequency cepstral coefficients for music modeling.,” in ISMIR T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. Okuno. Musical instrument recognizer "instrogram" and its application to music retrieval based on instrumentation similarity. International Symposium on Multimedia, pages 265–274, 2006., 2000.
12. Banchhor, Sumit Kumar, and Arif Khan. "Musical instrument recognition using zero crossing rate and short-time energy." *Musical Instrument* 1, no. 3 (2012): 1-4.
13. Das, Orchisama. "Musical Instrument Identification with Supervised Learning." *Comput. Sci* (2019): 1-4.
14. Blaszke, Maciej, and Bożena Kostek. "Musical instrument identification using deep learning approach." *Sensors* 22, no. 8 (2022): 3033.
15. Dataset one: https://www.kaggle.com/datasets/dibakarsil/music-instruments-and-2d-figures
16. Dataset two: https://zenodo.org/record/3685367