Classification of Musical Instruments' Sound using kNN and CNN

Srishti Vashishtha
Department of Computer Science and Engineering
GGSIPU
New Delhi, India
srishtidtu@gmail.com

Rachna Narula
Department of Computer Science and Engineering
GGSIPU
New Delhi, India
rachna.narula@bharatividyapeeth.edu

Poonam Chaudhary
Department of Computer Science and Engineering
The NorthCap University
Gurugram, India
poonamchaudhary@ncuindia.edu

***Abstract*— In the field of machine learning, the classification of musical instrument sounds is becoming popular. The aim of this work is to automate the identification and categorization of diverse musical instruments based on their acoustic characteristics. This classification approach entails gathering extensive audio data and employing various machine learning algorithms- kNN (k-Nearest Neighbor) and CNN (Convolutional Neural Network) to extract pertinent features. The proposed models can effectively distinguish between different musical instruments. These classification models find applications in various areas including music information retrieval, recognizing instruments in music production, and analyzing music performance. The kNN model achieves accuracy of 96% and CNN model achieves an accuracy of 88%. The F1- score of kNN and CNN model is 0.96 and 0.93 respectively.**

Keywords- kNN, CNN, Musical Instruments, Machine Learning, Classification

I. INTRODUCTION

Music is a universal language that has been enjoyed by people for centuries. With the advancement of technology, we are now able to not only listen to music but also analyze it with the help of machine learning techniques. One important aspect of music analysis is the classification of musical instrument sounds.

Recently many classification works are popular for textual data- sentiment analysis [21-22] natural language processing [23-24], image data- breast cancer detection [25], medical image segmentation [26], vehicle detection and audio data [27], speech emotion recognition [28], music recommendation [29]. Musical Instrument Classification (MIC) [35] is one of the major important tasks for acquiring the high-level information about the musical signal. The musical notes are divided into monophonic and polyphonic notes. Monophonic note instrument categorization is essentially effective, while polyphonic note annotation of musical signal is more difficult to discern.

For example, in music production, being able to automatically identify the instruments being played in a recording can greatly speed up the process of mixing and mastering a track. Similarly, in audio transcription, automated instrument recognition can be used to transcribe music more accurately and efficiently. In music education, machine learning can help students learn about different musical instruments and their sounds.

However, identifying musical instruments from audio recordings is a complex task that is challenging even for human listeners. This is because the sound produced by each instrument can differ greatly on factors such as the player's technique, the instrument's condition, and the recording environment. Additionally, multiple instruments may be playing at the same time, making it even more difficult to distinguish between them.

To overcome these challenges, researchers have developed machine learning algorithms and deep neural networks that can learn to recognize and classify different musical instrument sounds with a high degree of accuracy. These algorithms can be trained on large datasets of audio recordings of different instruments, allowing them to learn the subtle differences in sound between different instruments. To improve the effectiveness of the model, sound classification systems may be created using the learning capabilities of deep learning architectures. The major contributions of this work are deployment of two machine learning models: KNN and CNN models to correctly classify the sound of musical instruments. Among these two models, KNN model outperforms CNN model with an accuracy of 96%.

In the following section different musical instrument classification works are described. In section III proposed methodology is explained. Results are presented in section IV and last section concludes the paper.

II. RELATED WORK

The present research in computer vision field is on music instrument sound classification of musical instrument. There are several existing music instrument sound classification algorithms and techniques. Musical instrument sound classification [35] has many practical applications in fields such as music production, audio transcription, and music education. Some of the techniques include algorithms like Convolutional Neural Network (CNN), Artificial Neural Network (ANN), Region-Convolutional Neural Network (R-CNN), etc. which can be found in [1]-[11].

Kratimenos et al. [1] proposed a deep Neural Network architecture for music instrumental sound classification as a baseline model which consists of four convolutional layers and two dense layers. Additionally, they suggested using audio data augmentation techniques to address the issue of data scarcity throughout the entire IRMAS dataset. This proposed model greatly improves efficiency by yielding a mean accuracy of 77.9% as a result of the proposed augmentation.

Sawhney et al. [2] proposed a model based on Latent Feature Extraction for Musical Genres from Raw Audio. The amplitude of the raw audio from GTZAN dataset is used to train the model. In this proposed method, the accuracy of the vanilla auto encoder model is 65.3% and the two-layer neural network is 52%. Avramidis et al. [3] proposed an end-to-end approach for Polyphonic Instrument Classification from Monophonic Raw Audio Waveforms on proposed model which is a combination of RCNN and BiGRU model. The residual fully convolutional model's third CNN cell is where the BiGRU module is embedded, and the output of the two models is averaged to create the optimal architecture. On the IRMAS dataset, the proposed combined model came up with the accuracy of 75%.

Zhang et al. [4] proposed a MIDIMC method based on deep learning for the classification of sound. The dataset used in classifying the music instrument sound was MIDI dataset. Bi-GRU and an attention mechanism were used in the design of the MIDI classification network model, and classification is carried out. The classification accuracy of the proposed approach on MIDI dataset is 90.1%.

Hing et al. [5] proposed a CNN model for sound classification on IRMAS dataset. The training accuracy and validation set accuracy came out to be 70.3% and 60.4% respectively. The dataset includes about 1, 000 validation set examples. The multi-class classifier obtains 40% accuracy on train set, a 30% accuracy on validation set, and a 30% accuracy on the IRMAS testing data. IRMAS testing data has five instrument classes.

Blazek et al. [6] suggested a revolutionary method that uses sets of unique convolutional neural networks (CNNs) for each tested instrument to automatically identify all instruments in an audio snippet. The Slakh dataset, which has 2100 audio recordings, aligned MIDI files, distinct instrument stems, and tagging, was employed. Four instruments—bass, drums, guitar, and piano—were chosen for the experiment out of all those that were accessible. Utilizing the Keras framework and useful API, the suggested neural network was put into practice. Using internal Keras algorithms, the model first generates MFCC (mel frequency cepstral coefficients) from the raw audio signal. The results demonstrate that a straightforward convolutional network based on the MFCC can identify the instruments present in a given musical audio snippet having a F1 score of 0.93 and a precision of 0.93. The experiment also demonstrates that the instrument tested affects how well identification works.

Shreevathsa et al. [7] proposed model for single-instrument classification. The accuracy for Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) model was 72.08% and 92.24% respectively on audio samples. Gururani et al. [8] presented a DNN-based Instrument Activity Detection (IAD) system which detects 18 instruments and uses multitrack dataset. Mean AUC for various time resolutions of MLP, CNN and CRNN are 71.1, 80.92 and 80.1 respectively. Electric bass, acoustic guitars and vocals are among the instruments frequently seen in popular music that the CRNN and CNN successfully identify. They surpass MLP architectures in this task. Moreover, it also performs great for under- represented instruments in the dataset such as flute, cello and flute.

Lezhenin et al. [9] proposed a model based on Long Short Term Memory for urban sound classification for its great efficiency in time dependency learning. UrbanSound8k dataset is used to train the model which has the magnitude of the Mel-spectrogram. In this proposed method, the accuracy shown by the CNN model comes out to be 81.67%, and the proposed LSTM model produces the accuracy of 84.25%, although both models produce results that are nearly identical.

Yang et al. [10] proposed the method of combination of SVM (Support Vector Machine) with multihued attention mechanism to replace the recurrent framework which is widely used in sound classification models. Its applicability to various acoustically characterized fields was confirmed using three benchmarked datasets (Urban-Sound8K, IEMOCAP, and GTZAN). The classification accuracy of the proposed approach on UrbanSound8k dataset, GTZAN and IEMOCAP are 94.6%, 88.4% and 62.8% respectively. Comparison of recent works in sound classification is shown in Table I.

TABLE I. COMPARISON OF RECENT WORKS IN SOUND CLASSIFICATION

Reference	Dataset	Modelused	Accuracy
[1]	IRMAS Dataset	Deep Neural Network	77.9%
[2]	GTZAN Dataset	Vanilla Auto Encoder Two layer neural network	65.3% 52%
[3]	IRMAS Dataset	Combination of RCNN and Bi-GRU model	75% 88.4% 62.8%
[4]	MIDI Dataset	MIDIMC method based ondeep learning	90.1%
[5]	IRMAS dataset	CNN	60.4%
[6]	Slakh dataset	CNN	93%
[7]	Multitrack dataset	ANNCNN	72.08% 92.24%
[8]	Urban Sound 8k dataset	MLP CNN CRNN	71.1% 80.92% 80.1%
[9]	PascalVOC dataset	CNN LSTM	81.67% 84.25%
[10]	Urban Sound8k	SupportVector	94.6%

III. PROPOSED METHODOLOGY

In this project, we are watching out for the issue of recognizing the audio produced from differentinstruments and classifying the sound using deep learning. We, at first,

recognize all of the sounds in the audio and subsequently classify them as of which class it belongs to. We proposed to deal with this issue in two stages, which are: pre-processing the audio followed by feature extraction and classification. There are various algorithms in machine learning that can be used for sound classification such as K-Nearest Neighbor (KNN), Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Auto encoders etc. In this research work KNN and CNN models are implemented.

KNN (K-Nearest Neighbors) is a machine learning algorithm used for classification and regression tasks. Being non-parametric, it makes no assumptions on the distribution of the data. Fig.1. represents K- Nearest Neighbor (KNN) classifier. In sound classification, KNN can be used to classify audio samples into different categories based on their acoustic features. The acoustic features can be extracted using techniques such as Mel Frequency Cepstral Coefficients (MFCCs) or Spectrograms. To classify a new audio sample, the KNN algorithm computes the distances between the audio sample and its K nearest neighbor's in the training set. The predicted class for the new sample is then determined to be the most prevalent class among the K nearest neighbor's. CNN is a class of neural networks that specializes in processing data that has a grid-like topology. This model has 3 layers: a convolutional layer, a pooling layer and a fully connected (FC) layer [30]. Fig.2. represents CNN model for audio classification.

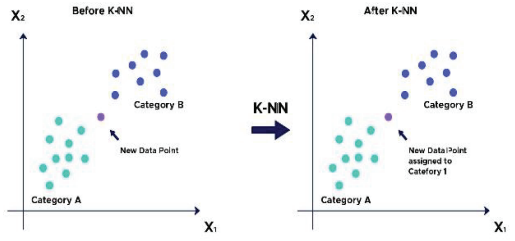

Fig. 1. k- Nearest Neighbor Classifier

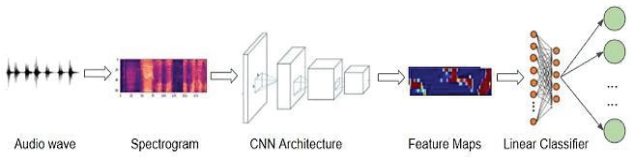

Fig. 2. Convolutional Neural Network

Methodology used for the musical instrument sound classification for KNN and CNN models can be described using the following steps:

A. Choosing the Dataset

For training the model, dataset is required. Subset of audio samples are downloaded from Philharmonic Orchestra Sound Samples which contains 1000s of free sound samples. Dataset consists of 600 audio samples belonging to different classes such as cello, saxophone flute, oboe, viola and trumpet.

In this step, all the audio samples are mono-channelized whether it is stereo or mono audio. The audios are de-sampled from 44100 to 22050 per second and then finally checked if the dataset is uniform or not. Mel- Frequency Cepstral Coefficients (MFCC) is used for feature extraction of the audio samples.

B. Dataset Pre-processing

In this step, all the audio samples are mono-channelized whether it is stereo or mono audio. The audios are de sampled from 44100 to 22050 per second and then finally checked if the dataset is uniform or not. Mel Frequency Cepstral Coefficients (MFCC) is used for feature extraction of the audio samples.

Mel-Frequency Cepstral Coefficients (MFCC)

MFCC feature extraction process includes some important steps such as windowing of the signal, DFT applications, application of inverse DCT and wrapping frequencies on Mel scale. The MFCC summarized the frequency distribution across the window size using Fourier transform, through which we analyzed both the frequency and time characteristics of the sound. These audio representations allowed us to recognize the features from the spectrograms that are obtained for the classification. Feature vectors are calculated after performing MFCC. Some standardizations will be done on feature vector so that the dataset can have zero mean and unit variance.

C. Division of Dataset

The dataset is divided into two parts - 3:1 ratio (75% for training and 25% for testing).

D. Choosing and Training the Model

The model to be used is K-Nearest Neighbor (KNN). KNN classifier is a non-parametric order approach. A new instance is categorized using KNN, a supervised learning approach, based on the nearby training examples that are present in the feature space. It just relies on memory and does not fit using any models. When test information is entered, it is assigned to the class that is often the most fundamental among its k nearest neighbor's. KNN classifier approach requires no parameters. It need not worry about any prior information on the organization of the information while preparing the set. If the new preparing design is incorporated into the current preparing set. Any links can be severed without consideration. The neighborhood classification serves as the prediction value for the new query instance in the KNN algorithm. The local structure of the data has an impact on the KNN algorithm. One of those simple-to-understand algorithms that performs really well in practice is the K- Nearest Neighbor. Activation function such as Softmax, ReLu and SGD optimizer is used. The dataset is trained with 200 epochs.

E. Evaluating the model on the testing dataset

The system will be ready to perform classification tasks, we can pass random audios similar to the trained data to our

system that was not part of the dataset. The accuracy, precision, recall, F1-score will be calculated for evaluation of model and confusion matrix will be used to visualize the model. The overall methodology is depicted in Fig.3.

IV. Results

We have used the K-Nearest Neighbour (KNN) algorithm for Musical instrument sound classification is done using K-Nearest Neighbour (KNN) algorithm and Convolutional Neural Network (CNN) algorithm on around 600 different musical instrument samples including 6 different classes such as cello, saxophone, flute, oboe, viola and trumpet. We have successfully classified different audios produced by different musical instruments according to their classes. Fig. 4, Fig. 5, and Table II represent results of kNN algorithm.

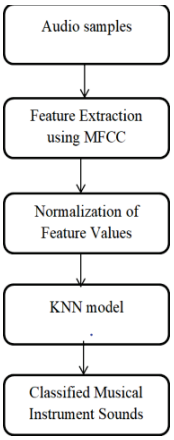

Fig. 3. Flowchart of Musical Instrument SoundClassification

TABLE II. Results of Musical sound classification using KNN and CNN.

Performance Metrics	Results for KNN	Results for CNN
Accuracy	96%	88%
Recall	0.96	0.93
Precision	0.96	0.93
F1-score	0.96	0.93

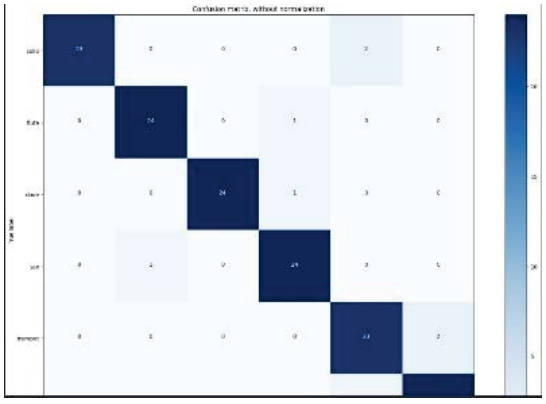

Fig. 4. Confusion matrix of testing dataset

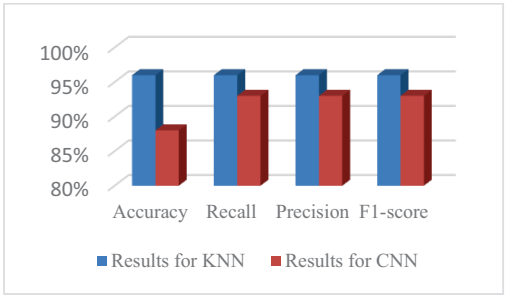

Fig. 5. Comparative Analysis of KNN and CNN models for musical sound classification

V. Conclusion And Future Scope

In this paper, musical instrument sound classification is performed on a philharmonic dataset that includes 600 audio samples by training the model for 6 different classes, i.e., cello, oboe, flute, saxophone, etc. The K- Nearest Neighbor (KNN) algorithm is used as a base model for musical instrument sound classification. The obtained results show that the algorithm K-Nearest Neighbor (KNN) with an average accuracy score of 96%. By training the system for a greater number of epochs and using high-quality sound samples for the training dataset, accuracy, as well as precision, can be enhanced.

Future work will mostly focus on increasing the interpretability of our latent representations. In order to experiment with merging MFCCs and other types offeature extraction, we thus want to increase the number of samples and classes in the dataset as well as thescope of our task brief.

References

[1] A. Kratimenos, K. Avramidis, C. Garoufis, A. Zlatintsi, P. Maragos, Augmentation methods on monophonic audio for instrument classification in polyphonic music. In 2020 28th European Signal Processing Conference (EUSIPCO) (pp. 156-160). IEEE, 2021.

[2] A. Sawhney, V. Vasavada, & W. Wang, Latent feature extraction for musical genres from raw audio. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, QC, Canada (pp. 2-8), 2021.

[3] K. Avramidis, A. Kratimenos, C. Garoufis, A. Zlatintsi, & P. Maragos, Deep convolutional and recurrent networks for polyphonic instrument classification from monophonic raw audio waveforms. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3010-3014). IEEE, 2021.

[4] F. Zhang, Research on music classification technology based on deep learning. Security and Communication Networks, 1-8, 2021.

[5] D.S. Hing & C. J. Settle, Detecting and Classifying Musical Instruments with Convolutional Neural Networks, 2020.

[6] M. Blaszke & B. Kostek, Musical instrument identification using deep learning approach. Sensors, 22(8), 3033, 2022.

[7] P.K. Shreevathsa, M. Harshith, & A. Rao, Music instrument recognition using machine learning algorithms. In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), pp. 161-166, IEEE, 2020.

[8] S. Gururani, C. Summers, & A. Lerch, Instrument Activity Detection in Polyphonic Music using Deep Neural Networks. In ISMIR, pp. 569-576, 2018.

[9] I. Lezhenin, N. Bogach, & E. Pyshkin, Urban sound classification using long short-term memory neural network. In 2019 federated conference on computer science and information systems (FedCSIS), pp. 57-60, IEEE, 2019.
[10] L. Yang and H. Zhao, "Sound classification based on multihead attention and support vector machine," Mathematical Problems in Engineering, pp. 1-11, 2021.
[11] K.J. Piczak, Environmental sound classification with convolutional neural networks. In 2015 IEEE 25th international workshop on machine learning for signal processing (MLSP), pp. 1-6, IEEE, 2015.
[12] J. Sharma, O.C. Granmo, & M. Goodwin, Environment sound classification using multiple feature channels and attention based deep convolutional neural network, 2020.
[13] T. Qiao, S. Zhang, Z. Zhang, S. Cao, & S. Xu, Sub-spectrogram segmentation for environmental sound classification via convolutional recurrent neural network and score level fusion. In 2019 IEEE International Workshop on Signal Processing Systems (SiPS) (pp. 318-323). IEEE, 2019.
[14] Y. Su, Instrument Classification Using Different Machine Learning and Deep Learning Methods. Highlights in Science, Engineering and Technology, 34, 136-142.
[15] S. Prabavathy, V. Rathikarani, & P. Dhanalakshmi, Musical Instrument Sound Classification Using GoogleNet with SVM and kNN Model. In Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021 2 (pp. 230-240). Springer International Publishing, 2022
[16] A. Rosner, & B. Kostek, Automatic music genre classification based on musical instrument track separation. Journal of Intelligent Information Systems, 50, 363-384, 2018
[17] M. Blaszke, & B. Kostek, Musical instrument identification using deep learning approach. Sensors, 22(8), 3033, 2022.
[18] K. Choi, G. Fazekas, M. Sandler, & K. Cho, Convolutional recurrent neural networks for music classification. In 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP) (pp. 2392-2396). IEEE. 2017
[19] Y. Han, J. Kim, & K. Lee, Deep convolutional neural networks for predominant instrument recognition in polyphonic music. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(1), 208-221, 2016
[20] S. Vashishtha, V. Gupta, & M. Mittal, Sentiment analysis using fuzzy logic: A comprehensive literature review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(5), e1509, 2023
[21] S. Vashishtha & S. Susan, Neuro-fuzzy network incorporating multiple lexicons for social sentiment analysis. Soft Computing, 26(9), 4487-4507, 2022
[22] S. Vashishtha & S. Susan, Sentiment cognition from words shortlisted by fuzzy entropy. IEEE Transactions on Cognitive and Developmental Systems, 12(3), 541-550, 2019
[23] S. Vashishtha & S. Susan, Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis. Expert Systems with Applications, 169, 114323, 2021
[24] V. Gupta, H. Gaur, S. Vashishtha, U. Das, V. K. Singh, & D. Hemanth, A fuzzy rule-based system with decision tree for breast cancer detection. IET Image Processing, 17(7), 2083-2096, 2023
[25] N. Beniwal & Vashishtha, Comparative Analysis of U-Net-Based Architectures for Medical Image Segmentation. In International Conference on Advances and Applications of Artificial Intelligence and Machine Learning (pp. 715-725). Singapore: Springer Nature Singapore, 2022
[26] S. Vashishtha, S. Kumar, V. Bothra, V. Singhal, & A. Sharma, Vehicle Detection System using YOLOv4. In 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST) (pp. 1-5). IEEE, 2022
[27] S. Vashishtha, and S. Susan, Unsupervised fuzzy inference system for speech emotion recognition using audio and text cues (workshop paper). In 2020 IEEE sixth international conference on multimedia big data (BigMM) (pp. 394-403). IEEE, 2020
[28] E. Sarin, S. Vashishtha, & S. Kaur, SentiSpotMusic: a music recommendation system based on sentiment analysis. In 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST) (pp. 373-378). IEEE, 2022
[29] A. Dua, A. Bhatia, B. Kalra, & S. Vashishtha, A Novel Recurrent and Convolutional Neural Network Technique for Generating Handwriting from Voice. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1439-1444). IEEE, 2021
[30] R. Narula, V. Kumar, S. Vashishtha, & A. Airan, Health Benefit of Yoga and Gym: A Review. Advancement of Computer Technology and its Applications, 7(2), 5-15, 2023
[31] M. Mehndiratta, N. Jain, A. Malhotra, I. Gupta, & R. Narula, Malicious URL: Analysis and Detection using Machine Learning. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1461-1465). IEEE, 2023
[32] A. Nautiyal, D.K. Bhardwaj, R. Narula, & H. Singh, Real Time Emotion Recognition Using Image Classification. In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing (pp. 8-12), 2023
[33] A. Solanki and S. Pandey, Music instrument recognition using deep convolutional neural networks. International Journal of Information Technology, 14(3), 1659-1668, 2022
[34] A. D. Goswami, G.S. Bhavekar, & P.V. Chafle, Electrocardiogram signal classification using VGGNet: a neural network based classification model. International Journal of Information Technology, 15(1), 119-128, 2023
[35] R. Shashidhar, S. Patilkulkarni, & S.B. Puneeth, Combining audio and visual speech recognition using LSTM and deep convolutional neural network. International Journal of Information Technology, 14(7), 3425-3436, 2022.