Drum Instrument Classification Using Machine Learning

Anubhav Chhabra
Guru Tegh Bahadur Institue of Technology
Guru Gobind Singh Indraprastha University,
Dwarka, New Delhi, India
chhabra.anubhav1997@gmail.com

Aryan Veer Singh
Guru Tegh Bahadur Institue of Technology
Guru Gobind Singh Indraprastha University,
Dwarka, New Delhi, India
indiaaryanveer310@gmail.com

Ritesh Srivastava
Computer Science & Engineering Department
Galgotias College of Engineering and Technology (GCET)
Greater Noida, Uttar Pradesh, India
ritesh21july@gmail.com

***Abstract*— Music is a way to express our creativity. As an art form, music can go beyond the limits of human imagination. When one hears a piece of music or sounds, human brain releases chemical dopamine. Hearing sounds again and again repetitively allows us to remember characteristics and nature of sound in a very efficient way. This is known as auditory learning and is believed to occur in our day to day life which helps us in identifying, memorizing and classifying various sounds. It allows, for example, immediate recognition of sounds or voices which become familiar through experience. The exact same principle can be implemented using Machine Learning. Music and Mathematics are strongly correlated with each other, whether it be the waveform or the sequence in which the melody is being played. In this paper, a Drum Instrument Classification Model is implemented using Machine Learning. The data is self prepared by recording samples and by using a Drum Simulator. The initial dataset contains only audio files in .wav format. The pivotal task is to perform Feature Extraction from the audio files and using them to train the Machine Learning model. Finally, a model is created which is capable of classifying various drum instruments when provided with an audio input.**

Keywords— Feature Extraction from Audio, Instrument Classification, Machine Learning, Music Information Retrieval

I. INTRODUCTION

Machine Learning has created a boom in each and every industry we can think of, whether it be healthcare sector, financial sector or Cyber Security just to name a few. It is being used to tackle complex problems, which if manually programmed could take enormous amount of time. The music industry is also being revolutionized using machine learning to handle problems like Audio Classification, Music Transcription, Audio Fingerprinting, Music Retrieval and Music Recommendation.

Sound classification till date remains one of the most challenging problems to the domain of machine learning. Sound Classification using Machine Learning is very much similar to the human auditory learning. We as humans learn to distinguish between various sounds with our experience. The human brain is capable of recognizing sounds by somehow analyzing the various features of the sound. We implicitly, are unaware of what exactly those features are, but somehow our brain does the work for us.

Same process is followed in Machine Learning: a training dataset is provided (audio data here), then features are extracted from that data and model is trained based on that data. This data for training also contains a column that tells which instrument made that sound. Many different Machine Learning algorithms can be used to get a model for a dataset (algorithm gets that data and creates a model as an output). The resultant model is a classifier for new audios it gets. It gets the audio, processes it, and understands features of that audio. Sound Classification can also be done by using the feature vectors and applying Hidden Markov Model [1].

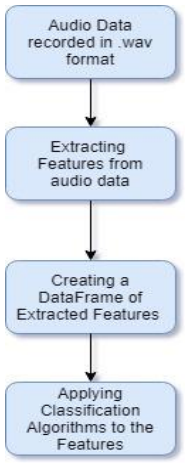

Fig. 1. Flowchart for understanding the workflow step-by-step

The main focus of this paper is on classification of various Drum sub instruments: Bass Drum, Snare Drum, Toms and Cymbal. The initial input is the audio recordings of the various sub parts of the Drum Kit in WAV file format. The audio recordings are self-prepared by recording instruments using proper recording equipment in a sound-proofed room so as to minimize noise interference. Feature Extraction is then performed on these audio recordings and various features are extracted and are recorded in a tabular format, having rows for each entry and columns for various features. One of the columns is the Class of the instrument to which it belongs. It is simply an integer mapped to an instrument (making the use of Label Encoder). The Flowchart below in Fig. 1 shows the workflow and the steps carried out for this paper. After introduction we have the section of Related Work that shows most of the papers and

978-1-7281-0211-5/19/$31.00 2019 ©**IEEE**

researches already done in the domain of music, which is followed by the section of Proposed Work that describes in detail how the data was collected, what all classifiers were used, how Feature Extraction was done and what all features were used to train the classifiers. Finally the Results show a 3-d plot of instruments classified based on the features and we have also calculated the Precision, Recall and f1 score.

II. Related Work

Machine Learning is often used to solve problems based on regression and classification, and there are various algorithms for the same. A very common classification problem which can be thought of is the email classification (Spam vs Not Spam). Various algorithms have been implemented and have been reviewed [11]. A similar classification can be performed on the audio data format as well. Machine Learning promises to enable a natural communication between the musicians and machines [3]. Audio samples can be classified on the basis of various features. This approach of classification was applied by Dannenberg, Thom and Watson [2], where they classified and recognized various music styles. Zhang and Kuo [1] also have already studied sound classification using Hidden Markov Model. They classified the sound effects and retrieved corresponding videos present in the database. Extracting features from music is an important task and Han, Zin and Tun [6] presented a journal for Extraction of audio features from music based on emotions. These days, genre classification is also a popular problem statement of this domain. Both machine learning and deep learning are being used for the same. With increase in the number of musicians and the huge increase in the number of Genres and Sub Genres it is definitely very much difficult to prepare a model which can easily classify a huge number of genres. Feng [4] has presented a genre classification model using Restricted Boltzmann machine algorithm. This work shows positive results, but these positive results are bounded to just a limited number of genres classes. Herrera, Yeterian and Gouyon [9] have studied drum sound classification in context of reducing the most relevant features of the refined preliminary set of descriptors and testing different classification techniques. As a general criticism, they used a set of 20 features to perform the classification, whereas in this paper we performed classification with only 3 features and achieved a high accuracy with the same. A more recent study which is apart from the Supervised, Jondya and Iswanto [8] presented Cluster Analysis using Unsupervised Training in Audio Features in which they clustered Indonesia's Traditional Music. Recent works in the same domain include intelligently generating style-specific music. The increase in the use of Neural Networks has also motivated researchers to use Deep Learning for this purpose as well and Mao, Shin and Cottrel [5] have worked on an end-to-end generative model that is capable of composing music.

III. Proposed Work

A. Data Collection

The initial dataset consists of self-recorded audio files in WAV format. The audio recordings then are parsed and converted into a list and various features are extracted from the same. Individual samples of drum instruments were recorded using a condenser microphone, an audio interface and Logic Pro X (Digital Audio Workstation). The dataset contains a total of 160 Wav format audio files, obtained from both live recordings as well as drum simulator. Each and every instrument was played multiple times and every time it was recorded simultaneously. As of yet we just have the audio files in .wav format. In this project, we are using a Python library 'Librosa' which is used for music and audio analysis [12]. Every audio is converted to its respective samples and using the result we are able to plot that audio file into a 2-dimensional plot which helps us understand sound wave better and easy. All the samples recorded were 16 bit Wav and having a sampling rate of 22.05 KHz. Table I below shows the distribution of the audio data files and their share percentage.

TABLE I. NUMBER OF SAMPLES IN THE DATASET

INSTRUMENT	NO. OF SAMPLES	PERCENT DISTRIBUTION
KICK DRUM	40	25%
SNARE DRUM	40	25%
TOMS	40	25%
CYMBALS (OH)	40	25%
TOTAL	160	100%

B. Feature Extraction

The Wav audio files have a sampling rate of 22.05 KHz, that means in one second of time there are 22050 samples in an audio list created, similarly for a duration of 2 seconds there will be a total of 44100 samples in an audio list parsed. This means that we have 160 various parsed audio lists, each of which has 44100 numbers in them. The list of each audio is then used to extract the various possible features from them [14], which includes: Zero Crossing Rates, Spectral Centroid and RMS Energies [13].

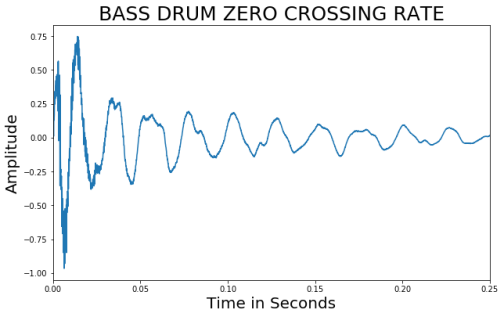

Fig. 2. Amplitude versus Time plot of a Bass Drum. (used for calculating the zero crossing rate here, by observing the number of sign changes in the wave)

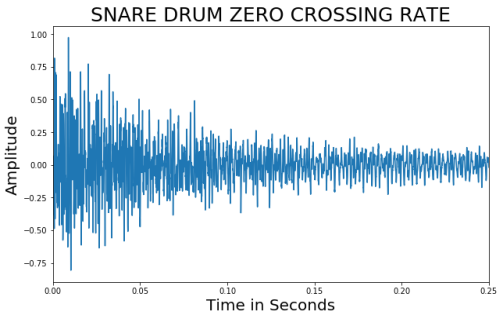

Fig. 3. Amplitude versus Time plot of a Snare Drum. (here the sign changes are occuring more often than the Bass Drum plot)

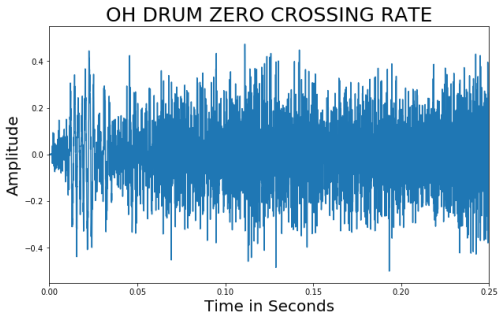

Fig. 4. Amplitude versus Time plot of Cymbals. (highest sign changes occurs in this plot)

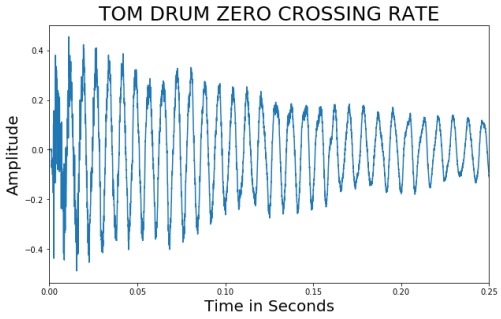

Fig. 5. Amplitude versus Time plot of a Tom Drum. (as the frequency is low, the number of times sound wave crosses the X-axis can also calculated by just looking at it carefully)

1) Zero Crossing Rate

Zero Crossing Rate is the number of times the sound wave cross the X - axis, (that is the Zero Reference Line) in unit time. This simply means that more the frequency of the sound wave more will be the zero crossing rate of the sound wave [10]. This is an important feature whose measure can be used in the classification process. The plots below show the zero crossing rates of various drum kit parts.

In Fig. 2, 3, 4 and 5 it is clearly observable that all the 4 parts have completely different waveforms and how the Zero Crossing Rate can help the model classify the instrument accurately. There is also relationship between Frequency and ZCR, higher the frequency: higher the ZCR.

2) Spectral Centroid

Spectral Centroid is another important feature of a sound wave which highly correlates with the brightness of the sound. Theoretically, it gives us the point where the 'centre of mass' of a spectrum is located. Mathematically, Spectral Centroid is the weighted mean of the frequencies present in a signal [7]. This feature, for all the 4 parts of the drum kit behaved quite differently for all. Thus, this is the most unique of them all and it makes it easier for algorithms to classify the parts.

3) Root Mean Square Energy

The root mean square energy of a wave is the square root of summation of distance of sampling points from the Zero Reference Line [18]. This feature of a sound wave reflects the energy stored in a sound wave. This feature thus has a close relation with the decay factor of a wave. Suppose if we have a sound wave which fades out quickly or decays very fast, (for example Kick Drum decays really very fast), then the total amount of energy would be lesser than a sound wave which decays after a long duration. Kick Drum and Snare Drum decay much faster than Toms and Overheads / Cymbals. Thus this decay can be also used as a classifying tool for the same purpose. The plots below show various samples of all 4 instruments in which the decay of the wave can be clearly seen.

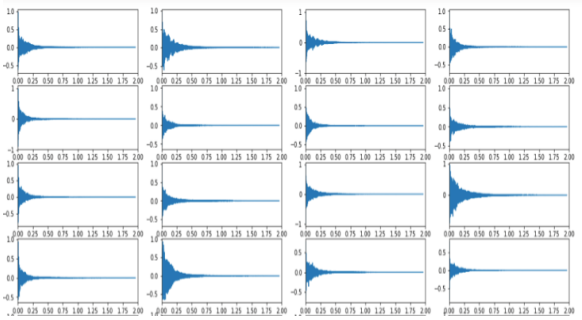

Fig. 6. Amplitude versus Time plots of a Snare Drum. (RMSE is the square root of summation of distance of sampling points from the Zero Reference line, and as we can observe that fequency and amplitude are not high in this plot so this has low RMS Error)

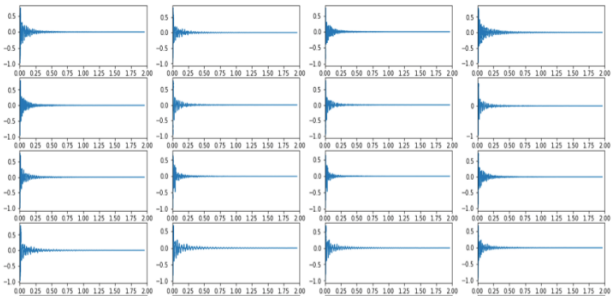

Fig. 7. Amplitude versus Time plots of a Kick Drum. (Kick and Snare Drums almost have the same frequency and amplitude, thus having almost same RMSE)

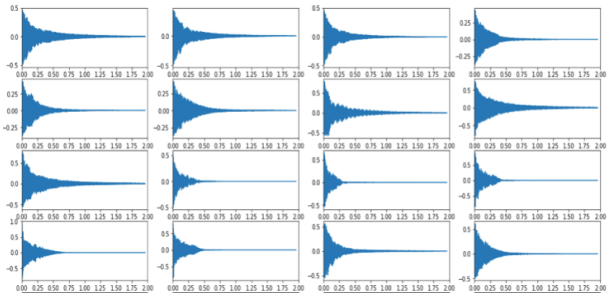

Fig. 8. Amplitude versus Time plots of Tom Drums. (these plots show higher frequency and amplitude, thus having higher RMSE than the plots above)

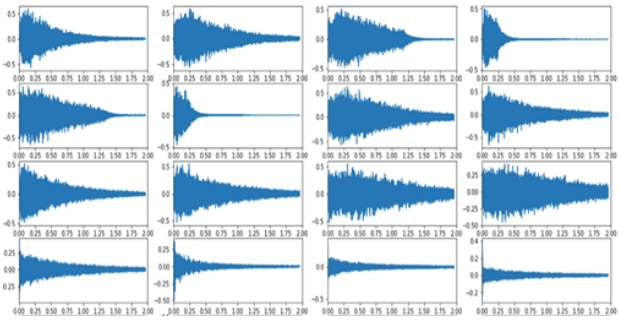

Fig. 9. Amplitude versus Time plots of Cymbals (these show the highest RMSE)

In Fig. 6, 7, 8 and 9 above, 16 samples from each category are taken of 2 seconds each. It is clearly visible that the last section of plots (that is the overhead plot) is the most crowded amongst all. Hence, this feature is also capable of distinguishing the class of instruments.

C. Classification Techniques

We worked using various classifiers including K Nearest Neighbors, Random Forests and Naive Bayes. Further tuning of the hyper parameters was done accordingly so as to increase the accuracy score.

1) K-Nearest Neighbors

The K nearest neighbors algorithm, takes into account the features of the testing data point and computes 'k' nearest neighbors (k here is the no. of neighbors specified) on the basis of the distance between them [15]. After computing the 'k' nearest neighbors which ever class has a higher no. of members present in these 'k' neighbors, is the final outcome class which will be predicted by the model.

2) Random Forest

Random Forest is a type of ensemble learning methods capable of performing classification as well as regression tasks [16]. In case of classification, it does so by constructing multiple decision trees at the training time and then predicting the class which is the mode of classes of individual decision trees. It is prone to over fitting, but in that case pruning can be done so as to avoid over fitting.

3) Naïve Bayes

The Naive Bayes classifier is based on Bayes' Theorem with a Naive assumption that the features are independent of each other. In other words, it assumes that the features are uncorrelated and normally distributed. This classifier is based on a simple mathematical equation (1) of Probability that is the Bayes Theorem which is explained by Zhang in [17] and is given as following:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \qquad (1)$$

Further because of the naive assumption we have:

$$P(c|x) = P(x1|c) * P(x2|c) * P(x3|c) \ldots . \\ \ldots P(xn\,|\,c) * P(c) \qquad (2)$$

This Method is easy to understand, implement and it performs really well on classification tasks.

IV. Results

All the classifiers mentioned above were trained on the training data, with a percentage of 80% as training data and rest 20% as testing data. The classifiers were able to distinguish between various sub instruments sounds easily. The 3-dimensional plot shown in the Fig. 10 shows the representation of various samples of data based on their features (3 features).

All the three classifiers performed really well and all of them gave an accuracy percentage of above 90%. Out of all Naive Bayes performed the best. Table II shows the accuracy of all the classifiers used.

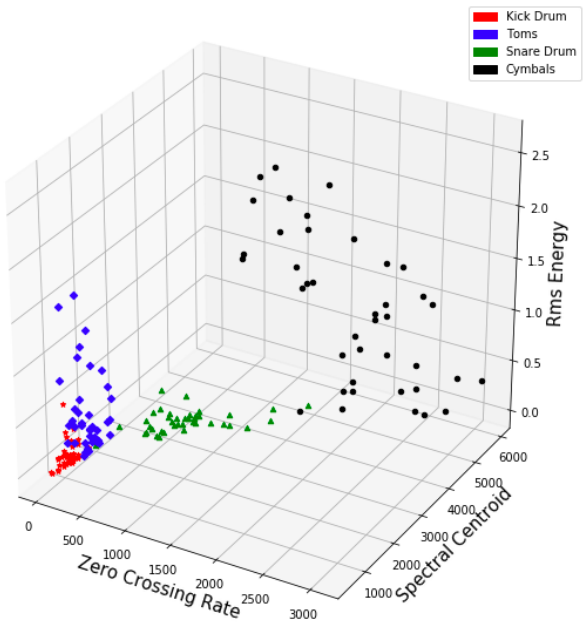

Fig. 10. A 3-dimensional plot of different drums for 3 three different features. (Zero Crossing Rate on the X-axis, RMSE on the Y-axis and Spectral Centroid on the Z-axis)

TABLE II. ACCURACY FOR VARIOUS CLASSIFIERS

Classifier	Accuracy
K Nearest Neighbors Classifier	93.75%
Random Forests Classifier	93.75%
Naive Bayes Classifier	96.875%

TABLE III. PRECISION, RECALL AND F1 SCORE

Class	Precision	Recall	F1 score
0	0.89	1.00	0.94
1	1.00	1.00	1.00
2	1.00	0.88	0.93
3	1.00	1.00	1.00

Apart from calculating the accuracy of the classifiers we also calculated the precision recall and f1 score which according to us is a better way of verifying the prepared classifier. Table III below shows the same. The values of these are calculated on the basis of values predicted by the Naïve Bayes Classifier.

Table III shows the Precision, Recall and f1 score and all three came out to be 0.97 for the self-prepared dataset.

V. Conclusion and Future Scope

Machine Learning models offer a really sophisticated and efficient way to perform various tasks related to Music Information Retrieval. In this work, the classifiers performed really well, because the features of the instruments didn't overlap each other. But so as to deal with such situations as well where the features used overlap or are in the same range, we will have to look for various other features of an audio which typically could distinguish each instrument from the other category. Here, a data of 160 wav files was self-prepared, features were extracted from them and were recorded in tabular form. This data was then fed to machine learning models so as to train them. A total of 3 classifiers were used which included K Nearest Neighbors,

Random Forests and Naive Bayes. Out of the three, Naive Bayes performed the best giving an accuracy of 96.875%. Furthermore, precision, recall and f1 score were calculated and all the three turned out to be 0.97, which depicts that the classifier performed really well. Further study and work is required so as to extract more features from the audio files so that the model can be scaled to a much larger range of instruments. The main focus would be laid upon the classification of various Indian Classical Instruments.

REFERENCES

[1] Zhang, Tong, and C-CJ Kuo. "Classification and retrieval of sound effects in audiovisual data management." In Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on, vol. 1, pp. 730-734. IEEE, 1999.

[2] Dannenberg, Roger B., Belinda Thom, and David Watson. "A Machine Learning Approach to Musical Style Recognition." In ICMC. 1997.

[3] Dannenberg, Roger B. "Artificial intelligence, machine learning, and music understanding." In Proceedings of the 2000 Brazilian Symposium on Computer Music: Arquivos do Simpsio Brasileiro de Computao Musical (SBCM). 2000.

[4] Feng, Tao. "Deep learning for music genre classification." private document (2014).

[5] Mao, Huanru Henry, Taylor Shin, and Garrison Cottrell. "DeepJ: Style-specific music generation." In Semantic Computing (ICSC), 2018 IEEE 12th International Conference on, pp. 377-382. IEEE, 2018.

[6] Han, Kee Moe, Theingi Zin, and Hla Myo Tun. "Extraction Of Audio Features For Emotion Recognition System Based On Music." International Journal of Scientific & Technology Research 4, no. 8 (2015): 53-56.

[7] Kua, Jia Min Karen, Tharmarajah Thiruvaran, Mohaddeseh Nosratighods, Eliathamby Ambikairajah, and Julien Epps. "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition." In Odyssey, p. 7. 2010.

[8] Jondya, Aisha Gemala, and Bambang Heru Iswanto. "Indonesian's Traditional Music Clustering Based on Audio Features." Procedia Computer Science 116 (2017): 174-181

[9] Herrera, Perfecto, Alexandre Yeterian, and Fabien Gouyon. "Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques." In Music and Artificial Intelligence, pp. 69-80. Springer, Berlin, Heidelberg, 2002

[10] Gouyon, Fabien, François Pachet, and Olivier Delerue. "On the use of zero-crossing rate for an application of classification of percussive sounds." In Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy. 2000

[11] Awad, W. A., and S. M. ELseuofi. "Machine Learning methods for E-mail Classification." International Journal of Computer Applications 16, no. 1 (2011).

[12] Librosa.github.io. (2018). LibROSA — librosa 0.6.2 documentation. [online] Available at: https://librosa.github.io/librosa/.

[13] "Index".2018. Musicinformationretrieval.Com. https://musicinformationretrieval.com/.

[14] Jensen, Jesper Højvang. Feature extraction for music information retrieval. Multimedia Information and Signal Processing, Aalborg University, 2010.

[15] Zhang, Min-Ling, and Zhi-Hua Zhou. "A k-nearest neighbor based algorithm for multi-label classification." In Granular Computing, 2005 IEEE International Conference on, vol. 2, pp. 718-721. IEEE, 2005.

[16] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[17] Zhang, Harry. "The optimality of naive Bayes." AA 1, no. 2 (2004): 3.

[18] Chai, Tianfeng, and Roland R. Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature." Geoscientific model development 7, no. 3 (2014): 1247-1250.