Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 6, pp. 1057–1074
DOI : 10.47974/JIOS-1431

A comprehensive study based on MFCC and spectrogram for audio classification

Priyanshu Rawat [@]
Madhvan Bajaj [#]
Satvik Vats *
Vikrant Sharma [‡]
Department of Computer Science and Engineering
Graphic Era Hill University
Dehradun
Uttarakhand
India

Abstract

Music Assortment is a music information retrieval (MIR) function to decide music connotation computationally. In recent years, deep neural networks have been proven to be effective in numerous classification tasks, including music genre categorisation. In this paper, we employ a comparative study between the two different music classification techniques. The first technique uses the audio's spectrogram image and computes the music's genre based on its spectrogram, using the CNN model trained on the spectrograms. The second approach computes the MFCC's (Mel-Frequency Cepstral Coefficients) musical features and utilises them to classify the music using ANN. This paper aims to study the two algorithms closely against different audio signals and check the performance report of the above-mentioned techniques to see which of them is better for music genre classification.

Subject Classification: *68T07.*

Keywords: *Music genre classification, Convolution Neural Network (CNN), Signal processing, Image processing, Spectrogram, Artificial Neural Network (ANN), Audio analysis.*

[@] *E-mail:* priyanshurawat.cse.20011683@gehu.ac.in
[#] *E-mail:* madhvanbajaj.cst.ml.ai.20011249@gehu.ac.in
* *E-mail:* svats@gehu.ac.in (Corresponding Author)
[‡] *E-mail:* vsharma@gehu.ac.in

1. Introduction

In the current era of automation, a huge amount of data is generated by the sensors which are installed in almost every component [1-20]. Audio processing is a complex task in contrast to other classification techniques. Maintaining a database of all the online activity of a user is a challenging task too with managing and simply accessing the songs based on the user's choice. One of the methods for efficient categorisation and organisation of songs is by their genre, which is defined by the musical features. Audio has two types of features: physical features and perceptual features.

Physical features of audio are based on the mathematical calculation of a waveform ie. spectrum, energy function, cepstral coefficient, fundamental frequency etc. Perpetual features are qualitative such as rhythmic structure, harmonic content, loudness, pitch, timbre, etc.

Nowadays, everyone is somehow consuming music in any form whether it be listening to it or producing music, Classifying the music based on its genre with a well-defined label has a great application in music recommender systems (Spotify, iTunes, Soundcloud, Gaana). To implement it we tend to automate the genre classification of audio to minimise human error as machines are robust. Machine Learning & Deep Learning algorithms are used to design the algorithm and produce the desired output [37-38].

In this paper, we propose two DL models to identify and classify the genre of the given audio

The first model uses the Convolution Neural Network (CNN) algorithm trained end-to-end on Mel Spectrograms of the audio files. In the second approach, we generate the Mel Frequency Cepstral Coefficients (MFCC) features from the time and frequency domain of the waveform and train an Artificial Neural Network (ANN) model to detect & classify the genre of the audio file. These models are evaluated against the GTZAN genre collection dataset, a well-known audio collection dataset. It has 1000 audio files belonging to 10 different classes in .wav format [39-41].

The remaining of the article is structured as follows. Literature of the previous research works is reviewed in Section 2. Dataset and methodologies are discussed in Section 3. Section 4 contains the results and analysis followed by the conclusion and future work in Section 5.

2. Literature Survey

Music Genre Classification has been quite popularly used and traversed in the last few years. The spectrogram is a very efficient visual feature to see the temporal change of energy distribution. Lee et al **[21]** trained a Convolutional Deep Belief Network (CDBN) having 2 hidden layers with the goal to learn about the hierarchical features representations on the spectrogram. It motivated researchers and Hamel et al **[22]** used the CDBN with 3 layers onto each other for the spectrogram to feed the extracted features to the SVM. In this paper **[23]** Haggblade explored K-means and KNN. multi-class SVM to classify the classical, metal, jazz, and pop entirely rely on the Mel frequency Cepstral coefficients (MFCCs) to distinguish the data. Costa **[24]** used the texture features that were based on the visual pattern Local Binary Pattern on Latin Music Database (LMD) and ISMIR 2004 dataset to compare the performance of local features with texture features and observed texture features were better than the other. In this paper **[25]** They designed a two-layer network using the manifold learning techniques as non-linear dimensionality reduction methods for music genre classification and differentiate between the output and the MFCCs features. Researchers Proposed a Semi-supervised approach **[26]** where they combined structured and unstructured data in an unbalanced proportion surpassing many ML algorithms and they revealed that long-time features are more noteworthy than the other two as stated by Lee et al **[21]** in his paper. Babu kaji and his fellow researchers [27] used timbral features and rhythmic-content features for the classification and implemented bagging with Extreme Learning Machine Ensemble (ELM) as its foundation on GTZAN and ISMIR2004 datasets. Their Work [28] uses a set of acoustic and visual features fused together into an ensemble to classify the genre of music on these datasets Latin Music Database (LMD), ISMIR2004, and GTZAN dataset. The result outperformed many features with an expedient optimization for parameters. This paper **[29]** implements Recurrent Neural Network (RNN) on a huge dataset of song lyrics due to its hierarchical structure and is used to learn by the hierarchical attention network (HAN). This implementation of HAN surpassed many neural and non-neural network outputs. Feature extraction is a computationally heavy task to execute. Authors in [30] aims the paper to work on symbolic extracted features (How humans perceive music) due to the inherent problems with music. The paper [31] has Mel Frequency Cepstral coefficients (MFCCs) features extracted from each song in the GTZAN dataset, these features act as input

for the Convolutional Neural Network (CNN) which yields a genre for the input music. This Study [32] focuses on building a recommender's system and it is accomplished in 2 steps i.e., Classification and Recommendation. Here CNN is trained on the features and acoustic features are then used to classify the appropriate algorithm to choose for classification to get the best results for the recommendation. The MFCCs-based genre classification is a superior system design, but it has low accuracy. A convolutional Neural Network based on visual features i.e., spectrograms is a better approach. In this paper [33] a new system is proposed that uses Spectrograms as input for a CNN but with a voting scheme. With the new voting scheme, the accuracy has increased by 8.38 %.This paper [34] reviews the research carried out in the field of MGC, the songs are first converted to short segments from a song and formed a labelled spectrogram which acts as input for the CNN having 6 convolutional layers with a final dense layer and activation function as Softmax activation function achieves 85% of accuracy over the GTZAN dataset. This paper [35] estimate and juxtapose the power of musical & non-musical features used in the Deep Learning models for genre classification. In the evaluation, Mel Scaled Features are the most robust features to be found. This Research [36] is a comparative study between the Deep Learning model and the traditional Machine learning model based on their performance on 3-sec audio features and on 30-sec audio features implemented for the respective music.

3. Dataset and Methodologies

3.1 *Dataset*

The dataset used in this paper is one of the most-used public datasets for research in Music Genre Recognition (MGR). This dataset was created by extracting 30 seconds of audio clips from various sources like CDs, radio, and recordings collected between 2000 and 2001. It contains 10 genres with each class containing 100 audio files with spectral representation for each audio file and 2 CSV files containing the features of audio files shown in the following Table 1. This study requires only the audio file data that belong to 10 categories for audio genre classification.

The audio clips were later spat into 5 seconds of audio files to increase the data 6 times for audio classification using deep learning algorithms. The dataset covers all popular genres including classical, country, disco, metal, etc.

Table 1
Number of instances in each genre class

Sno.	Genre	Count
1	Blues	600
2	Classical	600
3	Country	600
4	Disco	600
5	Hip-hop	600
6	Jazz	600
7	Metal	600
8	Pop	600
9	Reggae	600
10	Rock	600
	Total	6000

The data used in the research has been tabulated below. Each audio clip is in the format of a .wav file. The size of each raw data clip is roughly 1.5 MB which has later been spat into chunks of size about 350 KB each. The total data used in this study is approximately 1.4 GB.

3.2 Methodologies

Audio Signal is a function of amplitude and time. The Machine can read the audio in these 3 formats.

a) wav format
b) mp3 format
c) WMA format

3.2.1 Artificial Neural Network

Pre-processing

- We have used several features in the feature extraction phase to gain insight into the audio file, such as MFCCs, Zero-Crossing rate, spectral centroid, Spectral roll-off, Spectral bandwidth, and Chroma Frequencies.
- **Mel-Frequency Cepstral Coefficients** are 20 coefficients that represent the spectral envelope over the spectrum in a succinct

manner. Here the bands are uniformly distributed by a Mel-scale us as shown in eq. (1).

$$Mel\ \ Scale = 2595\ \log_{10}\left(1+\frac{frequency}{700}\right) \tag{1}$$

- **Zero-Crossing rate** is the rate for a signal to move from a positive towards negative crossing zero or vice-versa. It is used to calculate the audio smoothness us as shown in eq. (2).

$$ZCR = \frac{1}{time-1}\sum_{i=1}^{time-1}\{\alpha_{time}\alpha_{time-1} < 0\} \tag{2}$$

- **Spectral Centroid** evaluates amplitude of the centre of Spectrum for a calculated Fourier-Transformation window us as shown in eq. (3).

$$SC = \frac{\sum_x SC(x)F(x)}{\sum_x SC(x)} \tag{3}$$

- **Spectral Roll-Off** is the threshold frequency % where total spectral energy resides.
- **Chroma Frequencies** are 12 distinct semitones of the total spectrum indicating the pitch of each class.
- **Spectral bandwidth** is the width of the light spectrum at 50% of the maximum and shown by the Greek letter (lambda) λsb.

ANN Architecture

ANN is a Deep Learning algorithm used for pattern recognition, classification etc. It takes the features as input and processes them while passing through the hidden layers and assigns weights to the nodes and a bias is associated with each neuron, which is used to calculate the weighted sum (weight + bias) and transferred to the activation function. The activation Function is responsible to produce the output known as forward propagation. The output is then compared with the true output and loss is calculated on the deviation of predicted output to that of true output weights are then updated accordingly propagating backwards to minimise the loss known as backward propagation. Forward propagation initialises the weights for inputs and then based on the loss weights are updated to reduce the error in the output. It is done for a certain number of epochs us as shown in Figure 1 and Figure 3.

- **Input Layer:** It is the initial layer of the Neural Network that inputs the features and transfers them to the hidden layer interconnected to

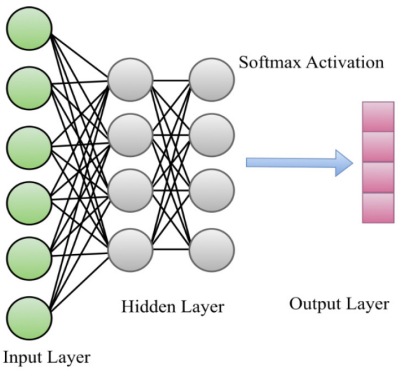

Figure 1

ANN Architecture

it. The number of features decides the number of nodes in the Input Layer. It duplicates the input features to the hidden layer.

- **Hidden Layer:** This Layer pertains to the transformation of the input features using the system of weighted links. It is linked with other hidden layers or the output layer and the summation of the product of the weights of the values is calculated.
- **Output Layer:** It is connected with a hidden layer and returns the predictions based on the processing done in the hidden layers for the input features. The number of nodes in the output layer is determined by the number of output classes.

3.2.2 Convolutional Neural Network

Pre-processing

- **Waveform** is a visual representation of a longitudinal sound wave propagating through any medium, which gives us a visualisation of these fundamental attributes varying over time.

  Wavelength (λ), Frequency (f) or Pitch, Amplitude (A), Velocity ($f * \lambda$)
- **Spectrogram** is a graphical representation of the frequency spectrum(y-axis) with respect to time(x-axis) for a sound or audio clip. It is also known as a spectral heatmap, voicegrams and sonography and is broadly used in signal-processing fields like sonar, linguistics, radar, speech, music, telecommunication, seismology and much more. Spectrograms are basically snapshots of audio waves generally

made by Fast Fourier Transformation of sampled input in the time domain. Each vertical line in the spectrogram is the measurement of magnitude over frequency at the given time and side by side or sometimes slightly overlapped assembling of these spectrums results in a 2D or 3D spectrogram where in a 3D spectrogram the variation of colour depicts the intensity of signal more clearly as 2D spectrogram relies on brightness for the same.

- **Fast Fourier Transform**

  An analogue waveform is to be converted to a discrete signal using Sampler. and is known as Sampling. After the signal is sampled, we apply it to the window. It prevents us from spectral leakage and reduces the computation as the segments do not overlap each other while performing Fourier transformation. If we directly apply Fourier transform to the signal then spectral leakage may be an issue and data may leak resulting in poor prediction. After it we apply FFT to the resultant signal and time domain features are transformed into frequency domain features. Lastly, all the segments are put together to get the full spectrum us as shown in eq. (4) and (5).

  Forward Transform

$$F(\omega) = \int_{T=-\infty}^{\infty} y(T) e^{-k\omega T} dT \qquad y(T) : -\infty < T < \infty \tag{4}$$

  Inverse Transform

$$y(T) = \int_{\omega=-\infty}^{\infty} F(\omega) e^{k\omega T} d\omega \qquad F(\omega) : -\infty < \omega < \infty \tag{5}$$

- **Mel spectrogram** is a type of spectrogram that is more suitable for the audios with low frequency and amplitude generally in which humans hear the sound. Spectrograms are inefficient in accounting for all the information which is required to differentiate sound at a human level that's why Mel spectrograms are used extensively for speech and music recognition and processing. Instead of frequency, the Mel spectrogram uses **Mel Scale**(y-axis) with respect to time(x-axis) and uses **Decibel Scale** to represent colours instead of the intensity or amplitude of the signal.
- **Mel Scale** is a scale constructed perceptually in a way that the sound is represented in pitch, logarithmic way instead of linear manner which is the most accurate representation of human hearing as

change between 100 to 500 Hz is obvious whereas the change in 7500 and 8000 is barely observable to us as shown in eq. (6).

$$Mel\ \ Scale = 2595\ \log_{10}\left(1+\frac{frequency}{700}\right) \tag{6}$$

- **Decibel Scale** is a logarithmic scale for amplitude or intensity of audio instead of linear order. Its non-linear nature of it provides a realistic approach and information from the perspective of human hearing as shown in eq. (7).

$$Decibel\,Scale = 10\log_{10}\left(\frac{I}{I0}\right) \tag{7}$$

CNN Architecture

CNN is a Deep Learning algorithm that in part works with images efficiently. It consists of multiple layers including convolutional layers, pooling layers and fully connected dense layers. CNN is in fact ANN with additional layers to pre-process the image by extracting key features and down sampling the resolution before feeding it to the underlying neural network as shown in Figure 2 and Figure 3.

- **First Convolution Layer:** The input size of the image is 256x256x3 pixel resolution. It is first processed against 16 filter weights in the first convolutional layer of kernel size (3x3) pixels. Convolutional layers(2D) with a 3x3 filter on an RGB image will be convolved to generate a 2D feature map, because of having multiple filters in each layer, the result is a 3D output i.e., a 2D map per filter mapped together. Convolving the image will enhance and highlight the presence of crucial features within the image.
- **First Pooling Layer:** by taking the highest value in the input window of size(2x2) and sliding along the input image which results in down sampling the image dimensions.
- **N^{th} Convolution Layer:** The (n-1)$^{\text{the pooling}}$ layer's output will be used as the input for the (N^{th}) convolution layer. This layer is using 32 filter weights of size (3x3 pixels each).
- **N^{th} Pooling Layer:** The down-sampled image has its resolution downscaled by using a maximum pool size of 2x2.
- **Flatten Layer:** This Layer is utilised to flatten the input while keeping the batch size unaffected. It flattens each batch to a single dimension.

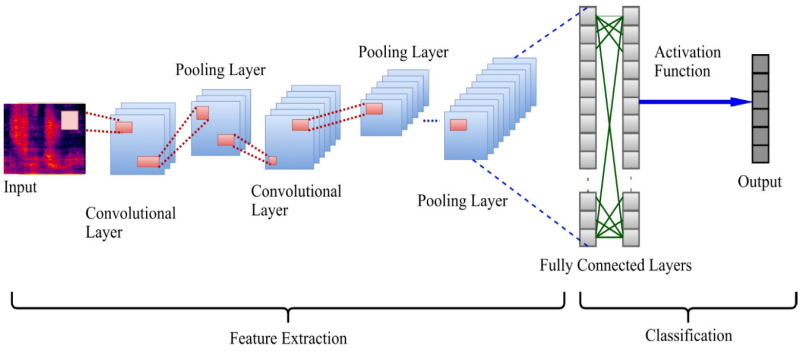

Figure 2

CNN Architecture

If the inputs are not oriented with a feature axis, flattening adds an extra stream dimension and the resulting shape is (batch, 1).

- **First Densely Connected Layer:** The output of the flattened layer serves as this layer's input. The flattened inputs are now sent to the first densely connected layer, which has 1024 fully connected neurons. The output is transmitted to the second densely connected layer. To prevent overfitting, a dropout layer with a value of 0.1 is used.
- **N-1th Densely Connected Layer:** The output from the previous densely connected Layer is now sent into a layer with 64 fully connected neurons.
- **Final layer:** The previous densely connected layer's output feeds into the final layer, which has as many neurons as classes being classified ie.10(music genres).

3.2.3 Activation Function

- **Sigmoid Activation Function**

  It is used in cases where we have to calculate output as the probability (0-1) referring to the following eq. (8).

$$\frac{1}{1+e^{-x}} \tag{8}$$

- **Tanh Activation Function**

  It ranges from [-1,1]. It is similar to the Sigmoid function but it can be used to map the negative values referring to the following eq. (9).

$$\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{9}$$

- **ReLU (Rectified Linear Unit) Activation Function**

  ReLU is the most widely used activation function. It is a partial rectification i.e.; it yields 0 when the value is <=0 referring to the following eq. (10).

$$R(v) = \max(0, v) \tag{10}$$

- **SoftMax Activation Function**

  It is the final layer of the DL model that is used to scale numbers into probabilities referring to the following eq. (11).

$$\frac{e^{y(i)}}{\sum_{j=1}^{n} e^{y(j)}} \tag{11}$$

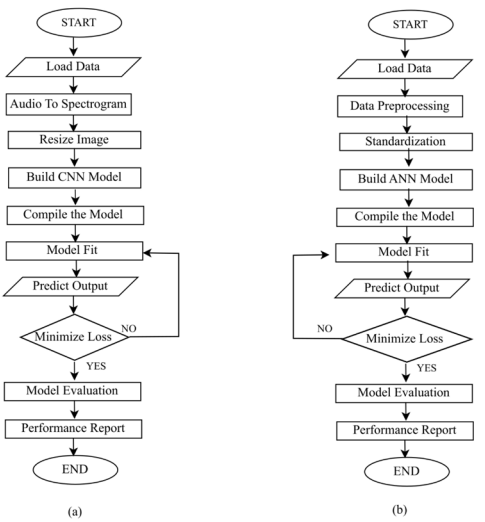

Figure 3

Flow Chart, (a). CNN, (b) ANN

4. Result and analysis

Here we study the performance of Artificial Neural Network (ANN) And Convolutional Neural Network (CNN) against the GTZAN dataset (MNIST of sounds) by Andrada Olteanu, the most commonly used dataset

in the domain of music genre classification. The experiments revealed that the CNN model has better accuracy and predicted the results precisely, over the ANN model. CNN predicted the outcome based on the frequency domain features of a Mel-spectrogram whereas ANN worked on MFCCs and spectrum features.

4.1 *Binary Classification*

Binary music genre classifier is trained for 2 classes. The accuracy for the CNN model is **97.6%** for Mel-spectrograms whereas the accuracy for the ANN model is **96.5%** for the MFCC feature vector and spectral features as shown in Figure 4 and Figure 5.

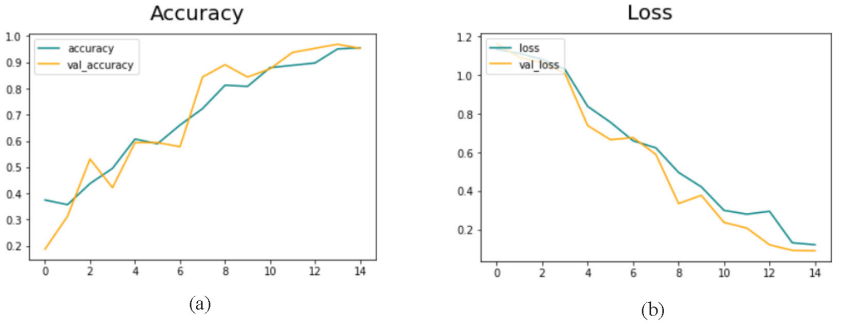

Figure 4

Validation Curves in CNN Model, (a) Accuracy, (b) Loss

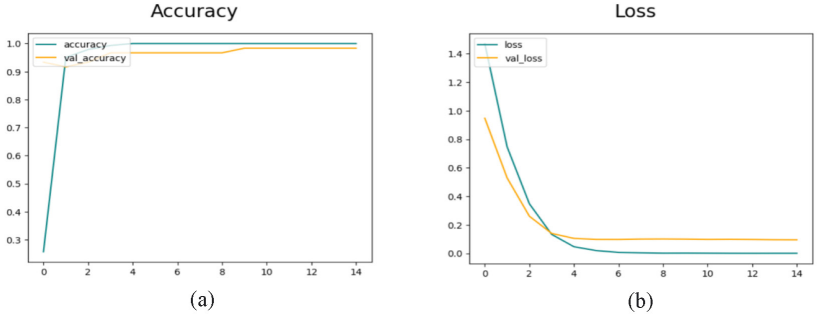

Figure 5

Validation Curves in ANN Model, (a) Accuracy, (b) Loss

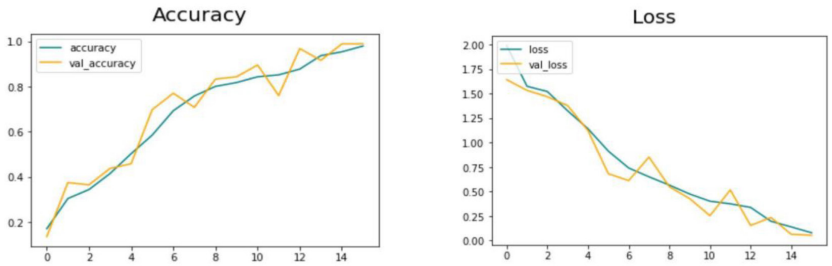

Figure 6

Validation Curves in CNN Model, (a) Accuracy, (b) Loss

4.2 *Multiclass Classification for 5 classes*

Multiclass music genres classifier is trained for 5 classes. The accuracy for the CNN model is **93.6%** for Mel-spectrograms whereas the accuracy for the ANN model is **82.4%** for the MFCC feature vector and spectral features as shown in Figure 6 and Figure 7.

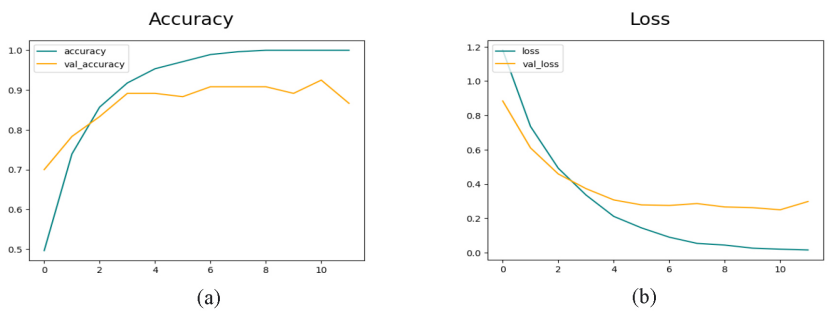

Figure 7

Validation Curves in ANN Model, (a) Accuracy, (b) Loss

4.3 *Multiclass Classification for 10 classes*

Multiclass music genres classifier is trained for 10 classes. The accuracy for the CNN model is **76.2%** for Mel-spectrograms whereas the accuracy for the ANN model is **61.4%** for the MFCC feature vector and spectral features as shown in Figure 8 and Figure 9.

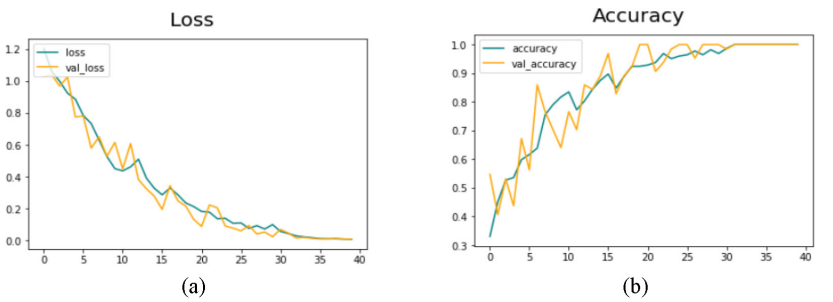

Figure 8

Validation Curves in CNN Model, (a) Accuracy, (b) Loss

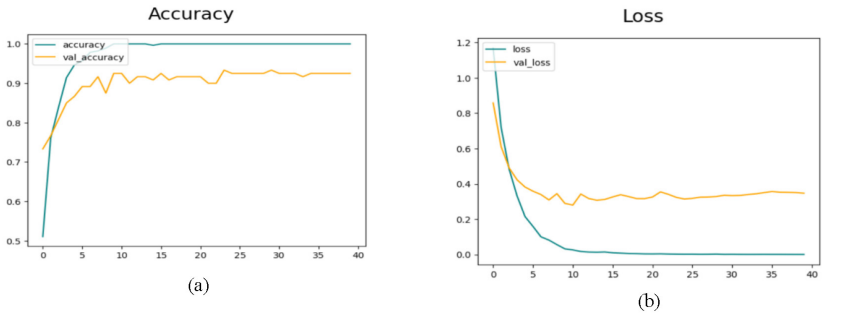

Figure 9

Validation Curves in ANN Model, (a) Accuracy, (b) Loss

5. Conclusion and Future work

In this paper, we used two distinct approaches of feature extraction for the two models used here for music genre classification. Librosa is used for audio analysis as it is a very handy python package for audio data having a rich collection of functions for audio.

According to the above analysis, CNN using Mel-spectrograms for music genre classification performs better in terms of accuracy with large numbers of classes and data as compared to ANN using MFCC (Mel Frequency Cepstral Coefficient) features. In the upcoming work, we will implement a hybrid model with Recurrent Neural Network and ensemble learning having lyrics as one of the features with the existing features to predict the mood and genre of the music. This can be used in the music recommender system as a filter based on lyrics and music style to give more personalised results or recommendations.

References

[1] R. Agarwal, S. Singh, and S. Vats, "Implementation of an improved algorithm for frequent itemset mining using Hadoop," Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2016, pp. 13–18 (Jan. 2017), doi: 10.1109/CCAA.2016.7813719.

[2] S. Vats, S. K. Dubey, and N. K. Pandey, "Criminal Face Identification System Criminal Face Identification."

[3] S. Vats and B. B. Sagar, "Data lake: A plausible big data science for business intelligence," Commun. Comput. Syst. Proc. 2nd Int. Conf. Commun. Comput. Syst. ICCCS 2018, pp. 442–448, Oct. 2019, doi: 10.1201/9780429444272-70/DATA-LAKE-PLAUSIBLE-BIG-DATA-SCIENCE-BUSINESS-INTELLIGENCE-SATVIK-VATS-SAGAR.

[4] M. Bhatia, V. Sharma, P. Singh, and M. Masud, "Multi-Level P2P Traffic Classification Using Heuristic and Statistical-Based Techniques: A Hybrid Approach," Symmetry 2020, Vol. 12, Page 2117, vol. 12, no. 12, p. 2117 (Dec. 2020), doi: 10.3390/SYM12122117.

[5] S. Vikrant, P. R. B, and B. H. S, "Policy for planned placement of sensor nodes in large scale wireless sensor network," KSII Trans. INTERNET Inf. Syst., vol. 10, no. 7 (2016), doi: 10.3837/tiis.2016.07.019.

[6] D. Prasad, A. Kumar, V. Sharma, and D. Prasad, "SDRS: Split Detection and Reconfiguration Scheme for Quick Healing of Wireless Sensor Network," *Int. J. Adv. Comput.*, vol. 46, no. 3, pp. 2051–0845 (Accessed: Feb. 03, 2023). [Online]. Available: https://www.researchgate.net/publication/263661007.

[7] D. Arora, A. Singh, V. Sharma, H. S. Bhaduria, and R. B. Patel, "Hgs-Db: Haplogroups Database to understand migration and molecular risk assessment," *Bioinformation*, vol. 11, no. 6, p. 272 (Jun. 2015), doi: 10.6026/97320630011272.

[8] V. Sharma, R. B. Patel, H. S. Bhaduria, and D. Prasad, "Policy for random aerial deployment in large scale Wireless Sensor Networks," Int. Conf. Comput. Commun. Autom. ICCCA 2015, pp. 367–373 (Jul. 2015), doi: 10.1109/CCAA.2015.7148445.

[9] S. Vats, B. B. Sagar, K. Singh, A. Ahmadian, and B. A. Pansera, "Performance Evaluation of an Independent Time Optimized Infrastructure for Big Data Analytics that Maintains Symmetry," *Symmetry* 2020, Vol. 12, Page 1274, vol. 12, no. 8, p. 1274 (Aug. 2020), doi: 10.3390/SYM12081274.

[10] S. Vats, S. Singh, G. Kala, R. Tarar, and S. D. FRCR, "iDoc-X: An artificial intelligence model for tuberculosis diagnosis and localization,"

https://doi.org/10.1080/09720529.2021.1932910, vol. 24, no. 5, pp. 1257–1272 (Jul. 2021), doi: 10.1080/09720529.2021.1932910.

[11] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "NADS: Neighbor Assisted Deployment Scheme for Optimal Placement of Sensor Nodes to Achieve Blanket Coverage in Wireless Sensor Network," *Wirel. Pers. Commun.*, vol. 90, no. 4, pp. 1903–1933 (Oct. 2016), doi: 10.1007/S11277-016-3430-6/METRICS.

[12] D. Prasad, A. Kumar, V. Sharma, and D. Prasad, "Distributed Deployment Scheme for Homogeneous Distribution of Randomly Deployed Mobile Sensor Nodes in Wireless Sensor Network," IJACSA) *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 4 (2013), Accessed: Feb. 03, 2023. [Online]. Available: www.ijacsa.thesai.org.

[13] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Deployment schemes in wireless sensor network to achieve blanket coverage in large-scale open area: A review," *Egypt. Informatics J.*, vol. 17, no. 1, pp. 45–56 (Mar. 2016), doi: 10.1016/J.EIJ.2015.08.003.

[14] S. Vats and B. B. Sagar, "An independent time optimized hybrid infrastructure for big data analytics," https://doi.org/10.1142/S021798492050311X, vol. 34, no. 28 (Jul. 2020), doi: 10.1142/S021798492050311X.

[15] S. Vats and B. B. Sagar, "Performance evaluation of K-means clustering on Hadoop infrastructure," https://doi.org/10.1080/09720529.2019.1692444, vol. 22, no. 8, pp. 1349–1363 (Nov. 2020), doi: 10.1080/09720529.2019.1692444.

[16] R. Agarwal, S. Singh, and S. Vats, "Review of parallel apriori algorithm on mapreduce framework for performance enhancement," Adv. Intell. Syst. Comput., vol. 654, pp. 403–411 (2018), doi: 10.1007/978-981-10-6620-7_38/COVER.

[17] S. Vikrant, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Glider assisted schemes to deploy sensor nodes in Wireless Sensor Networks," Rob. Auton. Syst., vol. 100, pp. 1–13 (Feb. 2018), doi: 10.1016/J.ROBOT.2017.10.015.

[18] V. Sharma et al., "OGAS: Omni-directional Glider Assisted Scheme for autonomous deployment of sensor nodes in open area wireless sensor network," ISA Trans., (Aug. 2022), doi: 10.1016/J.ISATRA.2022.08.001.

[19] J. P. Bhati, D. Tomar, and S. Vats, "Examining Big Data Management Techniques for Cloud-Based IoT Systems," pp. 164–191 (2018), doi: 10.4018/978-1-5225-3445-7.CH009:

[20] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Pneumatic Launcher Based Precise Placement Model for Large-Scale Deployment in Wireless Sensor Networks," IJACSA) *Int. J. Adv. Comput. Sci*. Appl., vol. 6, no. 12 (2015), Accessed: Feb. 03, 2023. [Online]. Available: www.ijacsa.thesai.org.

[21] Lee, H., Pham, P., Largman, Y., & Ng, A. : Unsupervised feature learning for audio classification using convolutional deep belief networks. *Advances in neural information processing systems*, *22* (2009).

[22] Hamel, P., & Eck, D. : Learning features from music audio with deep belief networks. In *ISMIR*, Vol. 10, pp. 339-344 (2010, August).

[23] Haggblade, M., Hong, Y., & Kao, K. : Music genre classification. *Department of Computer Science, Stanford University* (2011).

[24] Costa, Y. M., Oliveira, L. S., Koerich, A. L., Gouyon, F., & Martins, J. G. : Music genre classification using LBP textural features. *Signal Processing*, 92(11), 2723-2737 (2012).

[25] Rajanna, A. R., Aryafar, K., Shokoufandeh, A., & Ptucha, R. : Deep neural networks: A case study for music genre classification. In 2015 IEEE 14th international conference on machine learning and applications (ICMLA), pp. 655-660 (2015, December). IEEE.

[26] Poria, S., Gelbukh, A., Hussain, A., Bandyopadhyay, S., & Howard, N. : Music genre classification: A semi-supervised approach. In Pattern Recognition: 5th Mexican Conference, MCPR 2013, Querétaro, Mexico, June 26-29, 2013. Proceedings 5, pp. 254-263 (2013). Springer Berlin Heidelberg.

[27] Baniya, B. K., Ghimire, D., & Lee, J. : Automatic music genre classification using timbral texture and rhythmic content features. In 2015 17th International Conference on Advanced Communication Technology (ICACT), pp. 434-443 (2015, July). IEEE.

[28] Nanni, L., Costa, Y. M., Lumini, A., Kim, M. Y., & Baek, S. R. : Combining visual and acoustic features for music genre classification. Expert Systems with Applications, 45, 108-117 (2016).

[29] Kumar, A., Singh, K., & Khan, T. : L-RTAM: Logarithm based reliable trust assessment model for WBSNs. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(6), 1701-1716 (2021).

[30] Khan, T., & Singh, K. : Resource management based secure trust model for WSN. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(8), 1453-1462 (2019).

[31] S. Vishnupriya and K. Meenakshi, "Automatic Music Genre Classification using Convolutional Neural Network," 2018 International

Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1-4 (2018), doi: 10.1109/ICCCI.2018.8441340.

[32] A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydin, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Adana, Turkey, pp. 1-5 (2018), doi: 10.1109/ASYU.2018.8554016.

[33] Sugianto, S., & Suyanto, S. : Voting-based music genre classification using mel spectrogram and convolutional neural network. In 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 330-333 (2019, December). IEEE.

[34] Pelchat, N., & Gelowitz, C. M. : Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, 43(3), 170-173 (2020).

[35] Singh, Y., & Biswas, A. : Robustness of musical features on deep learning models for music genre classification. *Expert Systems with Applications*, 199, 116879 (2022).

[36] N. Ndou, R. Ajoodha and A. Jadhav, "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, pp. 1-6 (2022), doi: 10.1109/IEMTRONICS52119.2021.9422487.

[37] Hore, P., & Sharma, A. : Code-switched end-to-end Marathi speech recognition for especially abled people. *Journal of Discrete Mathematical Sciences and Cryptography*, 25(3), 771-784 (2022).

[38] Umamaheswaran, S., John, R., Deepthi, S. S., & Dharmarajlu, S. M. : Caption positioning structure for hard of hearing people using deep learning method. *Journal of Discrete Mathematical Sciences and Cryptography*, *25*(3), 623-633 (2022).

[39] Gupta, V., Juyal, S., Singh, G. P., Killa, C., & Gupta, N. : Emotion recognition of audio/speech data using deep learning approaches. *Journal of Information and Optimization Sciences*, 41(6), 1309-1317 (2020).

[40] Jaber, H. Q., & Abdulbaqi, H. A. : Real time Arabic speech recognition based on convolution neural network. *Journal of Information and Optimization Sciences*, 42(7), 1657-1663 (2021).

[41] Poulose, A., Reddy, C. S., Dash, S., & Sahu, B. J. R. : Music recommender system via deep learning. *Journal of Information and Optimization Sciences*, 43(5), 1081-1088 (2022).