Urdu Music Genre Classification CS316 - Introduction to Deep Learning Fall 2021

Shafaq Fatima Mughal
Habib University
Karachi, Pakistan
sm05571@st.habib.edu.pk

Shayan Aamir
Habib University
Karachi, Pakistan
sa05583@st.habib.edu.pk

Samarah Asghar Sahto
Habib University
Karachi, Pakistan
ss05563@st.habib.edu.pk

***Abstract*—Music is a form of art whose medium is organised sound. It has the ability of creating a brand of beauty, making one experience positive and deep emotions. It allows one to create, dream, think and connect with oneself and others. Communities from any part of the whole world can be identified by the type of songs they listen to and compose. The various kinds of music is often classified into different genres. The goal of this paper is to discover a machine learning algorithm that can efficiently classify music into genres. Our research focuses on the genre classification of Urdu music, that is native to Pakistan. For the purpose of performance comparison, different classification models were constructed and trained over our Urdu dataset. All models were trained on Mel-Frequency Cepstral Coefficients images of audio tracks and the performances were compared in terms of their validation accuracies.**

***Index Terms*—classfication, convolution neural networks, mel-frequency cepstral coefficients, urdu music genres**

I. Introduction

Music has always played a very integral part in people's lives and now, it has become more dominant than ever [1]. Music is an ever evolving discipline because new additions are being made to it every single day. However, different individuals prefer different types of music and this is what gives birth to **Music Genres**, that is a classification system which places different musics into neat categories based on some metrics [9].

Before diving in further, the commencement of this paper is best off with the fundamentals of music: sound. Sound is a form of audio signals which have many parameters including frequency, bandwidth, wavelength, decibel etc. A basic sound signal is a function of Amplitude and Time. Different formats of such signals exist such as MP3, Windows Media Audio and Waveform Audio File (WAV) which can be understood and speculated by computers [25].

Music classification is commonly used by audio distribution platforms such as Soundcloud, Saavn, Spotify etc. These companies use this classification as a tool that could determine recommendations for their users, or as a product of music search [25]. In order to perform such tasks, it is vital to know the genres of music. Therefore, the analysis of music occurs where machine learning algorithms come in play.

The division of music into classes may be considered subjective [27], however, there are non-cognitive standards which are the basis of Music's analysis, that help determine genres of music. This analysis is on a song's electronic badge for some features that include tempo, rhythm, energy, acoustics, speed etc. Genres of music are set apart by the common attributes shared by the members of each category. These attributes include rhythm, pitch instrumentation, harmony and texture of music [26]. Music genre classification is typically done manually for digitally accessible music. Therefore, procedures for computerized genre classification will prove to be an exceptional incorporation to the digital entertainment industry and audio information retrieval systems [27].

Furthermore, the relevance of music genre classification is prominent in the solving of certain tasks. For instance, creation of music reference, pursuing related music, determining communities or societies that will be interested in a particular music [28] and it may also prove to be helpful in survey schemes.

The classifier aimed to be built in this project will be the first of its kind to be capable of classifying Pakistani Local music, ergo, the chosen set of genres for this project will not only include globally known ones, but some common to Pakistani locals as well.

In our project, we wanted to make a deep learning model that when given a music file, can classify its genre. We wanted to train this model using some data and then test it as well. It was our aim to do so with Pakistan's local Urdu music. The purpose of this paper is to explore the previous work done in this field and then build up on it, concluding with the best approach for our project.

II. Music Genres

Addition to previously mentioned attributes, music exhibits certain generic forms of chord progression, key, melody, lyric and mood. It is by analyzing the musical and lyrical content and structures that music can be categorized into genres. [11] Genres relevant to our research are mainly: Rock, Hiphop, Ghazal and Qawwali.

Rock heavily relies on rhythm sections which creates a 'bombastic' beat. The instruments most commonly heard in rock rhythm are drums, bass guitar, electronic guitar, and acoustic guitars. The speed is usually fast-paced from the get-go. [14]

Hiphop is as a music genre consisting of stylized rhythmic music with beat mixing/matching, juggling, accompanied with rhythmic and rhyming speech. The typical instruments heard in hiphop music are turntable, drum machine, sampler, synthesizer and human beat-boxing. [18]

Qawwali is an art form with energetic musical performance of Sufi Muslim poetry that aims to lead its listeners to a state of religious ecstasy. "Traditional qawwali instruments are harmonium, sarangi, tabla, and dholak. Clapping, an important percussive component in qawwali music, serves as a rhythmic drone and is sustained throughout a song." [15]

Ghazal is a common and popular form of music in the sub-continent with roots in classical Arab poetry [16]. It primarily consists of harmonium, santur, sarangi, sitar, tabla, and rabab [17].

III. Problem Statement and Objectives

The importance of genres had been established earlier. As mentioned before, genre classification is usually done manually and sometimes it can be difficult to figure out the genre of a song this way, so building a deep learning model that classifies genres can prove to be very useful.
In our project we aimed to address the following objectives:

1. Develop an appropriate dataset for this task.
2. Develop deep learning models for the classification task of music into appropriate genres, optimizing them to reach the highest possible accuracy.
3. Test the models and compare accuracies with pre-existing models.

IV. Literature Review

Work related to Music Genre Classification shows that music information retrieval and classification is still an active research topic. There continues to be development in the field using new ideas yielding different results. The Classification models can be divided into two segments as in Figure 1

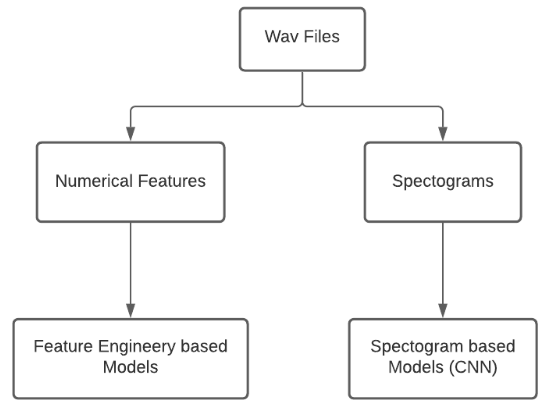

Fig. 1: Classification of Neural Networks

A. Models Based on Numerical Features

The **Detecting Music Genre Using Extreme Gradient Boosting** [3] paper predicts the genre of tracks provided in the FMA Dataset. They divided their approach into two parts, Numerical feature models and Image feature models, to provide a basis for comparison. Their first approach used Numerical features. Three different models were constructed. For the ExtraTrees and XGBoost classifier models they utilized a 5-fold cross validation to tune parameters, with n-estimators. Using Numerical features they also constructed a deep neural network with a random dropout (with p=0.5) at every dense layer. For the second approach a CNN was constructed which was trained on the spectrograms of the tracks. After every pooling layer, a random droupout (with p=0.5) was applied.
The result showed that traditional ensemble approaches outperformed the Neural Networks with XGBoost achieving the lowest lost, and CNN performing the worst with the highest loss, contradicting top tier approaches using neural networks for the same task.
The study yielded a combination of fewer, smaller layers with decreased image size which resulted in the CNN's reduced performance.

In another research paper called **Single-labelled Music Genre Classification Using Content-Based Features** [6], the researchers analyzed features that could be correctly used to classify musical genre. After careful analysis of the features, the researchers eliminated those that did not show a strong correlation with the genre. Using these results they then used the GTZAN data set with 1000 tracks to train six off-the-shelf classifiers with 10-fold cross-validation, including Na¨ıve Bayes, Support Vector Machines, Multi-layer Perceptron, Linear Logistic Regression Models, KNearest Neighbours, and Random Forests.
The results showed that the Linear Logistic Regression

Model provides the best classification score of 81%.

B. Models Based on Spectrograms

In **Music Genre Classification using machine Learning Techniques** [4], the researchers classified their music automatically by providing tags. By using both Neural Networks and Machine Learning algorithms, they were able to compare results and achieve their goal. The first approach uses Convolution Neural Networks, trained on Spectrogram images of audio tracks. The second approach uses various Machine Learning algorithms such as Logistic Regression, Random Forest etc. The researchers manually extracted features like Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, Spectral Centroid etc which were used to train the models.
XGBoost was determined to be the best feature-based classifier, the most important features were also reported. The CNN based deep learning models were shown to outperform the feature-engineered models. We also show that ensembling the CNN and XGBoost model proved to be beneficial.

Alternatively, in **Automatic musical pattern feature extraction using convolution neural network** [5], the researchers made an effort to understand the features that contribute to model for classification. The core purpose of their research was to understand the main features that mainly contributed to build the optimal model for Music Genre Classification. The dataset used for their approach was GTZAN consisting of 10 genres with 1000, 30 second tracks for each genre, sampling at 22050Hz at 16bits. Their results concluded that CNN had a strong capacity to capture informative features. The musical data had similar characteristics to image data, which require very less prior knowledge. As a result, the classifier accuracy was 84%. The resultant accuracy, according to the paper, could be enhanced further by ”increased by parallel computing on different combination of genres.” [8]

Furthermore, the researchers that worked on the research paper called **Improved Music Genre Classification with Convolutional Neural Networks** [7], used Short Time Fourier Transformation (STFT) magnitude spectrum, to visually represent the timbre texture of music. They made use of two different CNN architectures to formulate a comparative analysis between CNN and other forms of Neural Networks. The first CNN they used comprised of 10 layers, with 3 last layers were dense layers. Rectified Linear Units (ReLU) were used in all convolution and dense layers except for the top layer where instead, softmax was applied. The second Neural Network architecture was similar to the first one with the same number of layers and dense layers. The difference therein lay in the use of a residual layer and global max- and average pooling after the residual block, with shortcut connections. The architectures were trained on the GTZAN dataset using Batch Normalization to speed up the training process.
The results showed that both Neural Networks performed exceptionally well with an accuracy of 84.8% and 87.4% respectively, showing that CNN's can be improved by combining max- and average pooling and using shortcut connections, inspired by residual learning.

Model	Accuracy
NNET-2 [7]	87.4%
VGG-16 Fine Tuning [4]	64.0%
XGBoost [3]	78.0%
Linear Logistic Regression [6]	81.0%

TABLE I: Previous Research Models

Table I shows a summary of all the resulting accuracies achieved by previous models.

V. Dataset

As the famous saying goes, garbage in garbage out! So, adequate and reliable data is paramount for accurate results.
The biggest hurdle that arose in the initiation of this project was finding an appropriate dataset for Urdu songs. This was very tough and due to lack of existing datasets, eventually it was decided to make our own dataset from scratch.
The first task was to decide which genres to use, so we went with four most widely known genres of Urdu music: Ghazal, Qawwali, HipHop, Rock.

YouTube was used for downloading songs (songs uploaded by official YouTube channels were downloaded to ensure authenticity and integrity). Unlike the English music datasets that have been used in previous such projects, our Urdu dataset did have songs of variable lengths but uniformity was implemented on this data before processing it further. This will be discussed later in the paper. Table II illustrates details of our dataset, stating genres and the number of songs in each category.

Genre	Count
Ghazal	250
Qawali	250
Hip-Hop	250
Rock	250

TABLE II: Details of Urdu dataset

Manual construction of dataset inevitably led to problems. Firstly, due to time constraints, the number of audio files downloaded were limited. Even if an infinite amount of time was at hand, the number of total songs that have been produced by the Pakistani music industry is deficient and is bounded in comparison to the amount of songs produced by the English music industry that has a global popularity. Another issue faced while building this dataset was that hardly any songs were genre classified. Since there is no unanimous agreement on the taxonomy of genres [2], only those songs were picked that are most widely accepted for each genre.

We embarked on a two-step approach to minimize problems such as biases and unreliability of classification. Firstly, several elements of each genre such as rhythm, progression, and instruments were studied by listening to several songs from the same genre. This gave a basic pattern to identify. All members of our project listened to each song in the dataset and unanimously decided on a genre for an initial rough idea. The authenticity of this manual classification was then confirmed by trusted and legitimate music platforms such as Gaana and Spotify.
Now, for building a deep learning model, problems of having a small dataset still persisted. A major issue of a small dataset is over-fitting. Therefore, to overcome this Data Augmentation was implemented.

Data Augmentation forms new, different examples, using existing data, to train models. This alleviates the need for a larger dataset [13].

Thus, data augmentation was successfully administered in our dataset. This was done by taking several randomly chosen 30 second sections from the entire duration (always varied) of each song opposed to the conventional approach used by previous papers where only one 30 second interval is taken from each song. This increased the information taken from our dataset by many folds and lead to finer results as will be discussed later.

For even better model training, the first 30 seconds of ghazals and qawwalis and first 20 seconds of rock songs were clipped before applying data augmentation. The reason for this was that music from all three of these genres in the beginning was very mellow, almost silent, which would have confused our model, leading to poor results.

VI. Methodology

The first task in the practical implementation was to obtain MFCC images for the songs' .wav files using python's librosa library.

A. *Mel Frequency Cepstral Co-efficient Generation:*

Mel Frequency Cepstral Co-efficients (MFCC) gives us a 2D representation of the audio signals with time on the x-axis and values on the y-axis [4]. The following is a common way to calculate MFCCs:

1) Take (a windowed snippet of) a signal and perform the Fourier transform on it.

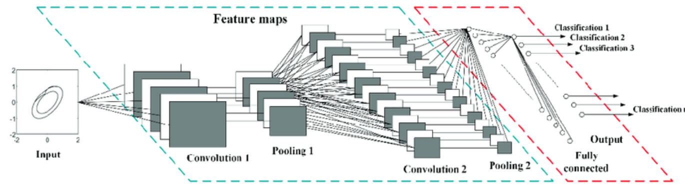

Fig. 2: CNN Framework [29]

2) Using triangular overlapping panes, map the powers of the spectrum acquired above onto the mel scale.
3) Compute the logs of the powers at each mel frequency.
4) As if it were a signal, compute the discrete cosine transform of the list of mel log powers.

The amplitudes of the resulting spectrum are the MFCCs.

For our models, we used the MFCC spectogram with the following parameters:

1) Sampling rate (sr) = 22050 [5]
2) Frequency Scale: MFCC
3) n_mfcc = 32
4) Bins/Segments: 5
5) Sampling time: 30s.
6) Window Size: 1292x13.

Once all MFCCS were generated, the dataset was split randomly into training set (60%), validation set (20%), testing set (20%).

B. *Convolution Neural Networks:*

Convolution Neural Networks (CNN) are special Neural Networks that do in-depth feature engineering using figure 2.

As figure 2 suggests and looking into the previous research, it was concluded that CNN is the best fit for our model, as it mostly gives the highest accuracy.

Pooling: This method is essential as it not only ensures translation in-variance but greatly reduces the dimension of the feature map that is given as output by convolution.

Non-Linear Activation Function: This adds enhanced feature extraction and retrieves more accurate predictions. ReLU function has been exercised in all models as it is quite reliable and avoids gradient vanishing.

Optimizers: They minimize the loss function. For our models we have used Adam optimizer which is a default for most Neural Networks and gives optimum results. [20]

Model Saving: Model Saving allows for the saving of the best possible model. During training, it may be possible to overshoot the lowest possible gradient using a specific learning rate. Model saving allows us to reload models for further evaluation or use. [23]

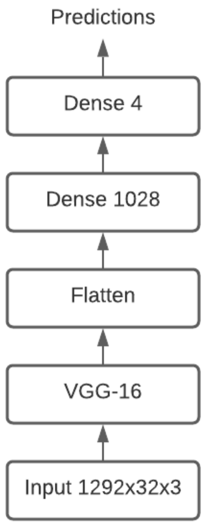

Fig. 3: CNN with Batch Normalization

Batch Normalization: It reduces internal co-variant shift, ensures correct weight initialization, does regularization (avoiding over-fitting), and makes the model less sensitive to hyper-parameter tuning. It also removes the need for dropout. [19]

VII. Evaluation Metrics

The evaluation of the performance of the models was executed using the following two metrics:

1. Accuracy: Literally speaking, this metric evaluates how accurate the model's prediction is in comparison to the true values. This can also be taken as the percentage of test data correctly classified.

2. Loss: It evaluates how well a model performs on the given data. In our models, we have used the Sparse Categorical Cross Entropy loss function. It takes labels as integers, calculates the cross entropy loss between the predictions and actual labels. Thus, one hot encoding is not required. [24]

VIII. Experiments & Results

We built four different models, all based on CNN and then evaluated them using the aforementioned metrics, to arrive at the best model.

A. Models

1) VGG-16 with Transfer Learning: For our experiments, we implemented a vanilla VGG-16 Model using Keras with pre-trained weights on ”Imagenet”. Since VGG-16 was pre-trained the channels of the input layer could not be changed and required an RGB image with a minimum dimension of 32x32x3. To alleviate this problem we replicated out image into 3 dimensions essentially creating a gray-scale image, changing our dimension from 1292x32x1 to 1292x32x3 as seen in Fig 3. Though this implementation consumes a lot of memory, it could not be helped. Furthermore, The output of the VGG-16 was transferred into a flattening layer with RelU activation, and finally into a dense layer with softmax for classification.

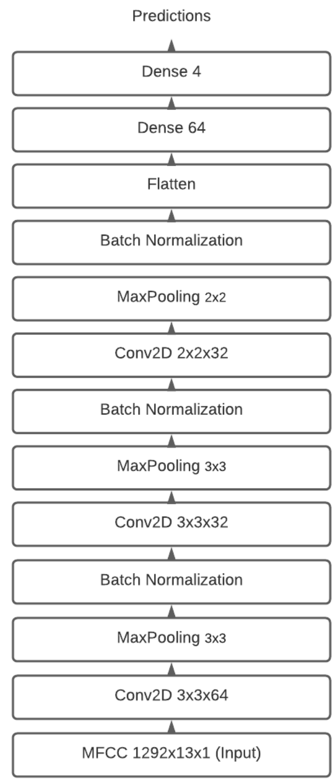

Fig. 4: CNN with Batch Normalization

2) CNN with Batch Normalization: This model follows a block approach. Each block has a convolution layer with ReLU activation followed by max pooling and then Batch Normalization. We have three such blocks in our network and a single dense layer finally applying softmax for classification. Figure 4 shows a visualization of this model's architecture.

This model is based on the model NNET-2 presented in [7]. NNET-2 used a combination of convolution and dense layers, but what makes it stand out is the concatenation of the results of the first convolution layer with the results of the third. The network was inspired by the concept of residual learning proposed by He et al. [30]. The resulting model showed considerable accuracy as shown in Table I.

However, CNN with Batch Normalization took a simplistic approach of deeper and narrower models [31]. Thus, the model consisted of convolutions layers with ReLU activation but also followed by Batch Normalization on every layer.

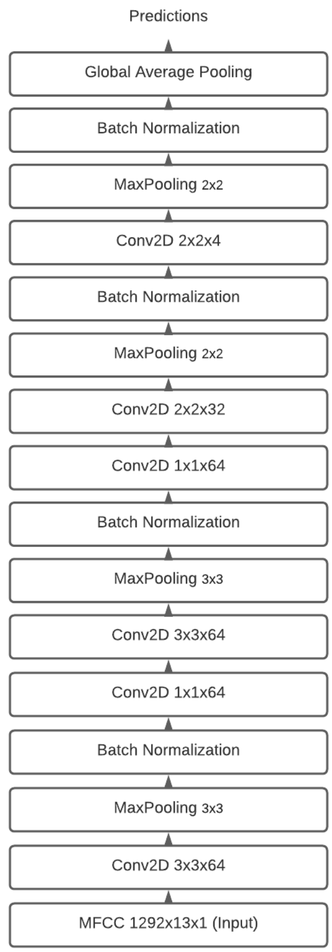

Fig. 5: CNN with Global Average Pooling

3) CNN with Global Average Pooling: For this model, the basic structure of 4 was made use of, but with added 1x1 Convolutions to control spacial depth and decrease the number of weights. These convolutions are followed by either 3x3 or 2x2 convolutions and has Batch Normalization employed. The basic architecture of the model is shown in Figure 5.

To decrease the number of parameters, the dense layers in the final layers are replaced with Global Average Pooling. This helps reduce training time and complexity of our model.

B. Results

All of the models consisted of several thousands of parameters that had to be trained. Each of these models were run for our Urdu dataset for several thousand epochs until the validation accuracies became repetitive and reached a maximum.

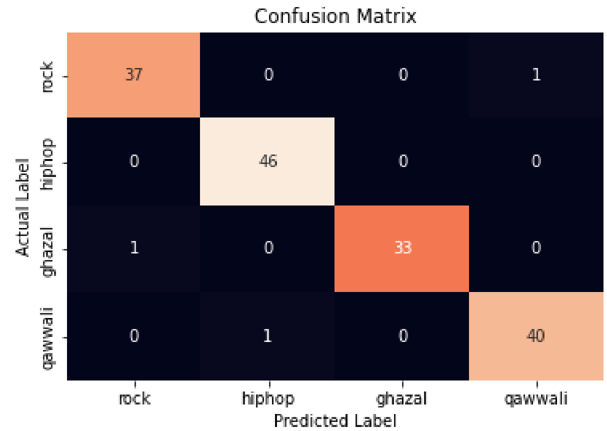

Fig. 6: Confusion matrix of Urdu test data

The results for running the models on our Urdu dataset are illustrated in Table III.

Model	Accuracy	Loss
CNN with Batch Normalization	92.6%	0.0051
CNN with Global Average Pooling	83.1%	0.8079
VGG-16 with Transfer Learning	83.1%	0.0061

TABLE III: Model results and evaluation

The results showed a clear conclusion. Our simplistic model, consisting of Batch Normalization outperformed VGG-16 and other models achieving high validation accuracy and low loss.

Both results conflict from [7] which used residual learning. However, for our Urdu Dataset result, it is important to acknowledge the limitations of our dataset as it is possible to get better and different results with a larger dataset.

To dive further into our best model (CNN with Batch Normalization), it was on testing data that had been separated before the start of our experiments. Figure 6 shows the results in the form of a confusion matrix. The results are quite promising. We see that our model gave a 100% test accuracy on the testing data for hiphop music and only one segment each was incorrectly predicted for the other three genres. This helps us understand the strengths and weaknesses of our model in terms of classification. Altogether, this gives quite a satisfying result.

We tested our model further by making it predict genres on unseen data. Two songs for each genre, not included in the existing Urdu songs dataset, were accessed and acquired correct positive results.

IX. Conclusion & Future improvements

First, we successfully built a dataset of Urdu songs based on 4 genres. Then, MFCCs after some pre-processing of data (including data augmentation) were found. Next, training of four different models were done on this data, keeping in mind the previous work done by people in music genre recognition and what was taught during our Introduction to Deep Learning course. A CNN model with Batch Normalization after each block and dense layers at the end proved to give us the highest validation accuracy.

Further improvements can be made to this classification problem. Firstly, this dataset was not very detailed and large so the creation of a larger dataset with more genres will be helpful and will lead to better accuracy (for better model training) and better testing of this model. Transfer learning can be implemented as it helps reduce learning times and increase accuracy of models as was apparent in [11], [12]. Furthermore, updating our model with Residual Learning, which has been shown to give considerable improvement in classification tasks [30] could also yield better results over a greater number of classes.

X. Acknowledgment

We would like to acknowledge our professor for CS316, Introduction to Deep Learning, Abdul Samad, for guiding and helping us choose our project. He was a constant support and encouraged us throughout the project.

References

[1] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu and N. Komarova, "Musical trends and predictability of success in contemporary songs in and out of the top charts", Royal Society Open Science, vol. 5, no. 5, p. 171274, 2018. Available: 10.1098/rsos.171274.

[2] N. Scaringella, G. Zoia and D. Mlynek, "Automatic genre classification of music content: a survey", IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 133-141, 2006. Available: 10.1109/msp.2006.1598089.

[3] B. Murauer en G. Specht, "Detecting music genre using extreme gradient boosting", in Companion proceedings of the the web conference 2018, 2018, bll 1923–1927.

[4] H. Bahuleyan, "Music genre classification using machine learning techniques", arXiv preprint arXiv:1804. 01149, 2018.

[5] T. L. H. Li, A. B. Chan, en A. H. Chun, "Automatic musical pattern feature extraction using convolutional neural network", Genre, vol 10, no 2010, bl 1x1, 2010.

[6] R. Ajoodha, R. Klein, en B. Rosman, "Single-labelled music genre classification using content-based features", in 2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2015, bll 66–71.

[7] W. Zhang, W. Lei, X. Xu, en X. Xing, "Improved music genre classification with convolutional neural networks", in Interspeech, 2016, bll 3304–3308.

[8] S. Chillara, A. S. Kavitha, S. A. Neginhal, S. Haldia, en K. S. Vidyullatha, "Music genre classification using machine learning algorithms: a comparison", Int Res J Eng Technol, vol 6, no 5, bll 851–858, 2019.

[9] Y. Neuman, L. Perlovsky, Y. Cohen, en D. Livshits, "The personality of music genres", Psychology of Music, vol 44, no 5, bll 1044–1057, 2016.

[10] K. E. Koech, "Cross-entropy loss function," Medium, 25-Feb-2021. [Online]. Available: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e.

[11] D. Beer, "Genre, boundary drawing and the classificatory imagination", Cultural Sociology, vol 7, no 2, bll 145–160, 2013.

[12] K. Choi, G. Fazekas, M. Sandler, en K. Cho, "Transfer learning for music classification and regression tasks", arXiv preprint arXiv:1703. 09179, 2017.

[13] R. L. Aguiar, Y. M. Costa, and C. N. Silla, "Exploring Data Augmentation to Improve Music Genre Classification with ConvNets," 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[14] "Instruments in Rock Music", 2021. [Online]. Available: https://study.com/academy/lesson/instruments-in-rock-music.html.

[15] "Qawwali instruments," Riyaaz Qawwali, 13-Aug-2021. [Online]. Available: https://riyaazqawwali.com/qawwali-instruments/.

[16] "Ghazal Music - Gazal Music India - Ghazals - Indian Ghazals - Ghazal Singers", Culturalindia.net, 2021. [Online]. Available: https://www.culturalindia.net/indian-music/ghazals.html.

[17] "Chandra and David' Gazal (ghazal) Page - Urdu Love Songs", Chandrakantha.com, 2021. [Online]. Available: https://chandrakantha.com/articles/indian_music/gazal.html.

[18] "Hip hop music", Cs.mcgill.ca, 2021. [Online]. Available: https://www.cs.mcgill.ca/ rwest/wikispeedia/wpcd/wp/h/Hip_hop_music.htm.

[19] "Why is Batch Normalization useful in Deep Neural Network?", Medium, 2021. [Online]. Available: https://towardsdatascience.com/batch-normalisation-in-deep-neural-network-ce65dd9e8dbf.

[20] "Various Optimization Algorithms For Training Neural Network", Medium, 2021. [Online]. Available: https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6.

[21] "Demystifying Batch Normalization vs Drop out", Medium, 2021. [Online]. Available: https://medium.com/mlearning-ai/demystifying-batch-normalization-vs-drop-out-1c8310d9b516.

[22] A. El-Sawy, E.-B. Hazem, en M. Loey, "CNN for handwritten arabic digits recognition based on LeNet-5", in International conference on advanced intelligent systems and informatics, 2016, bll 566–575.

[23] "Save and load Keras models — TensorFlow Core", TensorFlow, 2021. [Online]. Available: https://www.tensorflow.org/guide/keras/save_and_serialize.

[24] "tf.keras.losses.SparseCategoricalCrossentropy — TensorFlow Core v2.7.0", TensorFlow, 2021. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/losses

[25] "Music Genre Classification with Python", Medium, 2021. [Online]. Available: https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8.

[26] S. Poria, A. Gelbukh, A. Hussain, S. Bandyopadhyay and N. Howard, "Music Genre Classification: A Semi-supervised Approach", Lecture Notes in Computer Science, pp. 254-263, 2013. Available: 10.1007/978-3-642-38989-4_26

[27] Pretrained Deep Learning Model for Music Genre Classification, 2021. [Online]. Available: http://druckhaus-hofmann.de/gallery/36-wj-june-2020.pdf.

[28] M. Asim and Z. Ahmed, "Automatic Music Genres Classification using Machine Learning", International Journal of Advanced Computer Science and Applications, vol. 8, no. 8, 2017. Available: 10.14569/ijacsa.2017.080844

[29] CNN Framework. 2021 [Online]. Available: https://www.researchgate.net/figure/The-fundamental-framework-of-the-convolutional-neural-network-CNN-model$_fig1_335637427.[Accessed : 19 - Dec - 2021]$

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015.

[31] D. Zhou et al., "Go wide, then narrow: Efficient training of deep thin networks", in International Conference on Machine Learning, 2020, bll 11546–11555.

[32] B. Wu et al., "Shift: A zero flop, zero parameter alternative to spatial convolutions", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, bll 9127–9135.

[33] https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/