Analysis of MFCCs and Data Augmentation in Music Classification by Genre on Deep and Traditional Machine Learning Models.

1st Vukosi Moyane
Department of Computer Science and Applied Mathematics
University of Witwatersrand
Johannesburg, South Africa
vukosi90@gmail.com

***Abstract*—Automatic music classification is an area of research that has been growing in recent years. Deep learning approaches have been shown to be an effective method of accomplishing the task of automatic music classification by genre, however, existing methods make use of different data pre-processing and feature extraction methods, yielding varied performances and a lack of a common baseline for true comparison. It should also be noted that a lot of existing methods are not tested for generalisability through the use of different datasets. These factors considered, framed the focus of this research, which was to propose Mel-frequency Cepstral Coefficients (MFCCs) generated from audio signals that have undergone time-stretching and pitch-shifting augmentation, as a well-performing and generalisable processing pipeline to be used in the task of music classification by genre, as well as to prove the proficiency of deep machine learning models over traditional machine learning models, when trained using my proposed pipeline. This processing pipeline was applied to three datasets, which were used to train nine different models, five traditional and four deep, and the weighted average F1-scores were gathered for comparison. It was found that data augmentation actually worsened the performance of the trained models, however, MFCCs were proven to be a well-performing and generalisable feature set. It was also found that the deep models performed better than the traditional models.**

I. Introduction

Music is one of the most consumed and produced forms of art in existence today, it is a universal human enjoyment and dates back to at least 35000 years ago, and the demand for music is only increasing. [Fit06]. The number of digital music distributors and platforms is also increasing, with the number of songs produced per annum. [Wir22]. This being the case, warrants the need for efficient music organisation methods, for better marketing from the side of companies, and for better discoverability and ease of access to music on the part of consumers.

As noted by Chillara et al. [Chi+19], companies such as Spotify, Apple Music, Deezer, and Tidal, make use of music classification, to recommend music to, and construct playlists for, their customers, and other companies such as SoundHound and Musixmatch use music classification as is, to provide a service to their users. In order for these companies to perform these tasks, the genre of the music pieces would need to be determined, and can be determined through the use of machine learning models, with an approach that has gained popularity in more recent years being the use of deep learning machine learning models.

There have been many attempts to solve the problem of automatic music classification by genre. Feng [Fen14] tackled the problem using the GTZAN dataset, solely, of which a Mel- Spectrogram was extracted from the 30-second long audio signals contained within the dataset, from which Mel-Frequency Cepstral Coefficents (MFCC) were further extracted to train a Deep Belief Neural Network, which was their model of choice. The highest classification accuracy they recorded was 77.94%. Shi, Li, and Tian [SLT19] tackled the problem using the GTZAN dataset too, however, making use of Chroma features, extracted from the 30-second long audio signals contained within the dataset. The Chroma features were then used to train a VGG16 Convolutional Neural Network (CNN), which was able to attain a classification accuracy of 92.12%. They noted from their research that, Chroma features performed better than MFCCs, however, not drastically. Schindler, Lidy, and Rauber [SLR16] tackled the problem using four datasets, namely, the GTZAN, ISMIR2004, Latin Music Database and Million Song Dataset datasets. They extracted Mel-Spectrograms from the audio signals of each dataset and further ran them through a CNN, which was their model of choice, and used the activations of that CNN as features to further propagate through the network and train the model. They were able to attain accuracies ranging from 60.60% as the lowest to 83.22% as the highest, between the four datasets.

The method this research has followed follows an approach similar to those mentioned above, in that I made use of use of three different datasets, namely, the GTZAN, ISMIR2004 Genre and Free Music Archive datasets. I performed data augmentation through time-stretching and pitch-shifting of the audio signals contained within the datasets, then attained MFCC's from the audio signals of the respective datasets, after which I trained my models on each dataset and recorded their weighted average F1-scores. The models implemented were the VGG16 CNN, and the two CNNs proposed by Shi, Li, and Tian [SLT19], Gujjula [Guj], Schindler, Lidy, and Rauber [SLR16], respectively, an LSTM, as well as Support Vector Machines, Logistic Regression, Decision Tree Classifiers, K-Nearest Neighbours classifiers and Neural Networks.

This research aims to present a processing pipeline from an audio signal level to a feature level, that is able to allow models trained using this pipeline to perform proficiently in the classification of music by genre and to also generalise well in terms of their performance on many different datasets. This research also aims to compare the performance of deep learning models to traditional models trained on my proposed pipeline, with the expectation that the deep models will outperform their traditional counterparts.

The following sections of this document will address the related research, methodology, results and discussion, and the conclusion.

II. Related Research

A. Introduction

In order to gain an understanding of the task of music classification by genre, it is important to establish the context of the research. What will be discussed in this section are the various datasets and feature sets used by the authors, the classification performance of their various models and the best performing models, feature sets and performance improvement methods brought forth by the various authors.

B. Datasets used for Music Classification by Genre

Most of the authors reviewed used the GTZAN dataset to train and test their models, however, authors such as Vishnupriya and Meenakshi [VM18], used different datasets such as the Million Song Dataset (MSD), with similar deep learning models to the other authors that used the GTZAN dataset, and produced similar results. More so, authors Schindler, Lidy, and Rauber [SLR16], made use of four datasets, namely the GTZAN, ISMIR, LMD and MSD datasets, all to produce a more well rounded and general consensus on the effectiveness of their trained models in the task of music classification by genre.

C. Features used to train Models

A variety of features were used by the various authors to train their models.

Content-based features were the most widely used set of features for training models, and were used by all reviewed authors except for Shi, Li, and Tian [SLT19], Sigtia and Dixon [SD14]. Content-based features are features that are extracted from a raw audio signal, and can be split into two domains, the time domain, which includes features such as the zero crossing rate, tempo and root mean square energy; and the frequency domain, which contains features such as the spectral roll-off, spectral bandwidth, MFCC and Chroma features. Time domain features are extracted from the audio signal directly and frequency domain features are extracted from an audio signal (in the frequency domain) that has been transformed by a fourier transform.

Another notable approach to extracting features from audio signals would be those of authors Sigtia and Dixon [SD14], who made use of a Rectified Linear Unit (ReLU) Neural Network to learn features from the data, as well as Schindler, Lidy, and Rauber [SLR16], who used CNNs to learn features.

D. Models used for Music Classification by Genre

CNNs are one of the most widely used and best performing models in deep learning, and authors Lau and Ajoodha [LA22], Zhang et al. [Zha+16], Ndou, Ajoodha, and Jadhav [NAJ21], Elbir et al. [Elb+18], make use of them in their respective comparative analyses of deep learning against traditional machine learning approaches in tackling the problem of music classification by genre. It was found that, in general, stand-alone CNN's perform either on par with or actually worse than certain traditional machine learning models such as K-Nearest Neighbours (KNN), Logistic Regression and Support Vector Machines, as shown by authors Lau and Ajoodha [LA22], Ndou, Ajoodha, and Jadhav [NAJ21] and Elbir et al. [Elb+18].

However, many authors were able to achieve classification accuracies greater than that of most traditional machine learning models through the use of various techniques that enhanced their deep learning models. Authors Zhang et al. [Zha+16], enhanced the standard CNN model through the combination of max-pooling and average-pooling, allowing the provision of statistical information to deeper layers of the CNN, and through the use of shortcut connections to skip one or more layers in the CNN, allowing the author to attain a classification accuracy of 87.4%, outperforming traditional machine learning models, which range between 67$\pm$1% and 82$\pm$1%, only falling short of the KNN model which had an accuracy of 92$\pm$1% as shown by authors, Lau and Ajoodha [LA22], Ndou, Ajoodha, and Jadhav [NAJ21]. Authors Shi, Li, and Tian [SLT19] made use of the VGG16 CNN, which also makes use of max-pooling and dropout, to further increase its performance and attain a classification accuracy of 92.12%. Dropout is an important method of improvement in the classification accuracy of a deep neural network as noted by Schindler, Lidy, and Rauber [SLR16], who also achieved a high accuracy with their implementation.

E. Highest Performing Models and Feature Sets

The VGG16 CNN used by Shi, Li, and Tian [SLT19] and the CRNN constructed by Panwar et al. [Pan+17] were the best performing models. The best performing set of features used were the MFCCs and Chroma features, with the Chroma features being proven to be better than the MFCC features when being used as stand-alone features by Shi, Li, and Tian [SLT19].

F. Accuracy Improvement Methods

Data augmentation was proven by Schindler, Lidy, and Rauber [SLR16], to increase the classification accuracy of deep learning networks, as deep CNNs perform better when trained with larger datasets. The authors used two data augmentation methods, namely 'Time Stretching' and 'Pitch Shifting', applied directly to the audio signal, before feature extraction. Time stretching entails either speeding up or slowing down an audio signal, whilst retaining the original pitch information and pitch shifting entails either lowering or raising the pitch of an audio signal, whilst retaining its original tempo information.

G. Conclusion

Various works that deal with the task of music classification by genre have been presented and analysed by the topics: Datasets used, Features used to train Models, Models used, Highest performing Models and Feature Sets, and Improvement Methods. All aiding in this research.

III. Research Methodology

This section details the method utilised to tackle the automatic music classification by genre problem. Starting with dataset descriptions of the three datasets used, the MFCC feature extraction process, the different models used and lastly, the evaluation metrics chosen.

A. Data

I have made use of multiple datasets to gauge the generalisability of my proposed processing pipeline and models on the task of automatic music classification by genre.

*1) **GTZAN**:* This dataset was collected by Tzanetakis and Cook and consists of 1000, 16-bit, roughly 30-second long, song excerpts. These excerpts are almost evenly distributed between ten different genres; namely, Classical, Pop, Metal, Hip-hop, Reggae, Blues, Country, Rock, Jazz, and Disco. This dataset was sourced from Kaggle.com. Some of the faults of this dataset are that it contains exact replicas of entries, it suffers from the 'artist effect' (the artist effect is the observation that a music similarity system can perform significantly worse when artists are disjoint in training and test datasets). Due to the fact that all of the results come from models without a filter for artists, it is expected that the results will be more optimistic, although it is uncertain exactly how optimistic this dataset is, and there are also 59 mislabelings present in the dataset. Another concern that can be argued is that the dataset is not large enough to be an accurate representation of the genres one would wish a music genre classification (MGR) system to be able to discriminate between, as it contains only 1000 excerpts.

*2) **ISMIR2004 Genre**:* This dataset consists of 1458 full length audio clips, classified into one of the following 6 genres: world, jazz blues, rock/pop, classical, electronic and metal/punk. The dataset comes pre-split into training and testing datasets and the exact distribution per genre of the training dataset is as follows: 115 electronic music samples, 122 world samples, 26 jazz blues samples, 320 classical music samples, 101 rock/pop samples and 45 metal/punk samples. The testing dataset has the same number of samples as the training dataset and is distributed across the genres in a similar fashion. The dataset was attained from the Pompeu Fabra University website. One issue that can be raised against the dataset is its size, is only 45.8% larger than the GTZAN dataset, and thus relatively small, raising similar generalisability concerns.

*3) **Free Music Archive**:* The Free Music Archive (FMA) comes in four dataset sizes, 'full', 'large', 'medium' and 'small'. The small dataset, which is the one I made use of, is a balanced subset of the large dataset containing 8000, 30-second long clips with 1000 clips for each one of the 8 root genres. This dataset was used instead of its larger variants because it is quite large despite its name and has a GTZAN-like structure. It is pre-split into training, validation and test datasets, using stratified sampling to preserve the percentage of tracks per genre, using an 80/10/10 ratio. Songs from the same artists are part of one set only. This dataset resolves the problem of there being 'not enough samples' to truly gauge the generalisation capability of the models. The dataset was attained through the official Free Music Archive github repository.

B. Features

Research showed MFCCs to be the best performing sets of features, and thus, were used in this research. MFCCs were extracted from the audio signals of each dataset to train the models. This allows for generalisation of the models, as they are able to be trained with any dataset that contains audio signals with corresponding genres.

*1) **Mel-frequency Cepstral Coefficients**:* Mel-Frequency Cepstral Coeeficients (MFCCs) are short-term spectral-based features extracted from raw audio signals. [Log00].

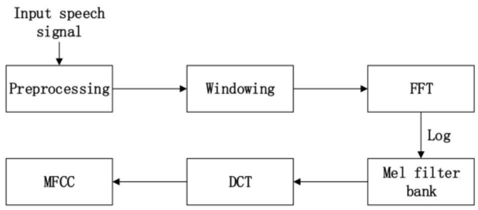


Fig. 1. The MFCC Extraction Process

MFCCs are extracted by firstly dividing the audio signal up into frames through a windowing function, which is usually a Hamming window, at fixed intervals. The windowing function removes edge effects. A cepstral feature vector is then generated for each frame. For each frame, the Discrete Fourier Transform (DFT) is obtained of which the logarithm of the spectrum's amplitude is then taken and preserved whilst all the other information is discarded. This is done because the perceived loudness of the audio signal has been shown to be approximately logarithmic, and the phase data has been shown to be much less important than the amplitude data, and can thus be discarded. Mel scaling and smoothing is then applied to the spectrum through the Mel filter bank, this is done to smooth and emphasize the perceptually meaningful frequencies in the spectrum. The last step that will yield the MFCCs is to apply a transform to the Mel-spectral vectors to decorrelate their components, this transform is known as the Discrete Cosine Transform (DCT). This process will yield 13 or more cepstral-features for each extracted frame, and thus, produce the MFCC feature array. [Log00].

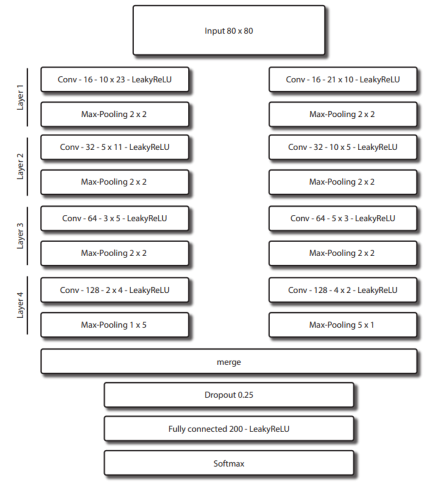
Authors such as Schindler, Lidy, and Rauber [SLR16], Feng [Fen14], Vishnupriya and Meenakshi [VM18], Panwar et al. [Pan+17], have used MFCCs as their features to train their respective models and attained relatively moderate to high accuracies. Schindler, Lidy, and Rauber [SLR16] attained accuracies ranging from 60% to 83.22% on four different datasets. Feng [Fen14] attained accuracies of 61.15% and 77.94% on the GTZAN dataset. Vishnupriya and Meenakshi [VM18] attained an accuracy of 76% on the MSD dataset and Panwar et al. [Pan+17] attained an accuracy of 92.12% on the MagnaTagATune dataset.

Thus, it can be seen that these features do indeed yield good performance overall, but accuracy is not a good performance indicator on datasets with imbalanced classes, and that is something that I have addressed in this research.

C. Models

1) Deep ML Models:

• **Schindler et al. CNN** The model proposed by Schindler, Lidy, and Rauber [SLR16].



- **VGG16 CNN** Used by Shi, Li, and Tian [SLT19].

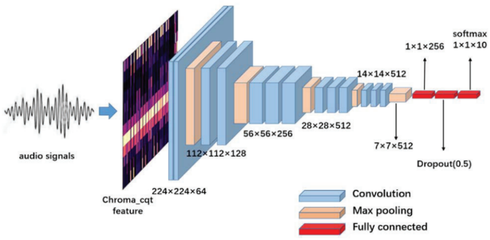
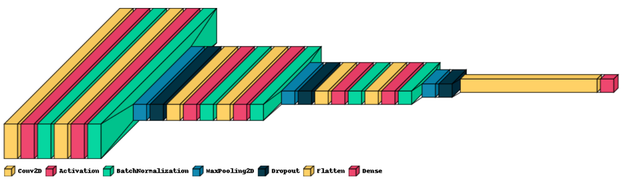


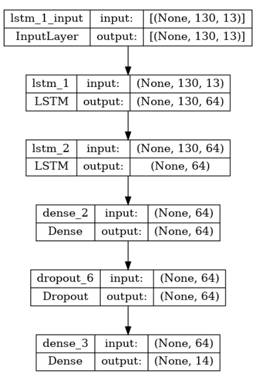
Fig. 2. Neural network structure.

- **Gujjula CNN** Used by Gujjula [Guj].



Each of these models,were implemented as implemented by their respective authors.

- **LSTM**



LSTMs are very proficient with time-series data, and this being the case, advocates for its use with the MFCCs, as they come from audio signals, which are time-series data.

2) *Tradional ML Models:*

- **K-Nearest Neighbours**
- **Decision Tress Classifier**
- **Support Vector Machines**
- **Logistic Regression**
- **Neural Networks**

These models were trained using the gridsearchcv python package, which finds the most optimal hyperparameters for each model.

D. *Evaluation*

F1-Score: The F1-score combines a classifier's precision (which is a measure of a model's efficacy to classify a sample as positive) and recall (which is a measure of a model's efficacy to detect positive samples) into a single metric by taking their harmonic mean. This metric is useful when there are imbalanced classes.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

IV. Results and Discussion

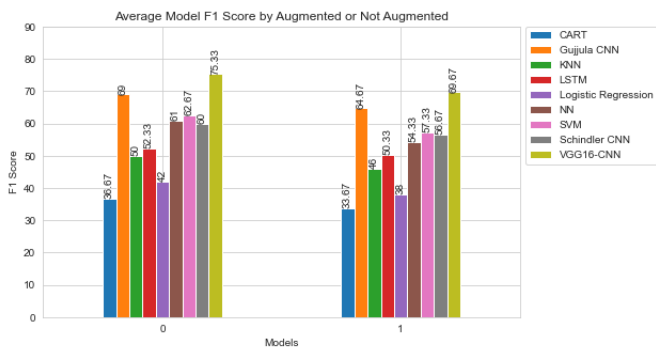
In this section the results from the various models trained on the proposed pipeline and future improvements of the pipeline and research are discussed.

The source code can be found by clicking the following link: CLICK HERE FOR SOURCE CODE

A. *Obtained Results*

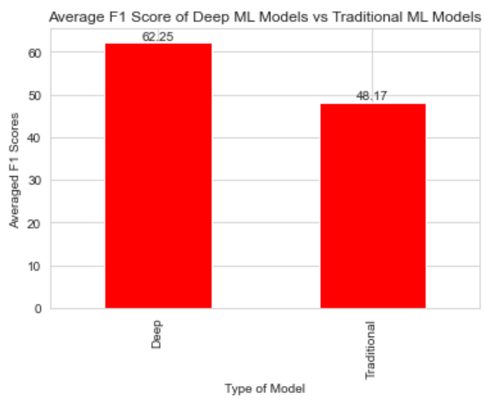
	GTZAN		ISMIR		FMA	
	Not Augmented	Augmented	Not Augmented	Augmented	Not Augmented	Augmented
	Weighted Avg F1-Score	Weighted Avg F1-Score	Weighted Avg F1-Score	Weighted Avg F1-Score	Weighted Avg F1-Score	Weighted Avg F1-Score
Trad ML						
CART	0,27	0,29	0,49	0,42	0,34	0,30
KNN	0,34	0,35	0,58	0,51	0,58	0,52
LogR	0,36	0,35	0,52	0,44	0,38	0,35
SVM	0,51	0,52	0,69	0,60	0,68	0,60
NN	0,52	0,54	0,72	0,60	0,59	0,49
Deep ML						
Gujjula CNN	0,65	0,65	0,76	0,65	0,66	0,64
Schindler et al. C	0,54	0,58	0,68	0,58	0,58	0,54
VGG16-CNN	0,68	0,69	0,80	0,68	0,78	0,72
LSTM	0,42	0,46	0,62	0,54	0,53	0,51

B. *The Effects of Data Augmentation*



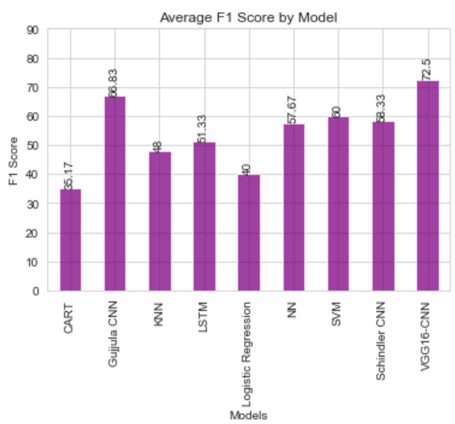
Above is a plot showing the average F1-scores of all the trained models, split by augmented or non-augmented data. It can be seen that data augmentation causes models to perform worse. On the GTZAN dataset the results from its augmented and non-augmented versions remained roughly the same, but for the ISMIR and FMA datasets, the performance drop of the models trained on the augmented datasets compared to that of the non-augmented dataset is very significant with an average F1-score decrease of 14.334% on the ISMIR dataset and 8.789% on the FMA dataset. Now, the audio signals from the GTZAN and FMA datasets came reduced to 30 second long audio segments of the most genre-indicative sections of the original music pieces, whereas the ISMIR dataset came with full length audio samples, which I trimmed to 30 second long segments during feature extraction, but not the most genre-indicative 30 second long segments. This explains why the ISMIR dataset is the worst performer given augmentation, however, does not explain why the FMA dataset performs poorly given augmentation. This indicates an area of improvement in future work.

C. Deep Learning Models vs Traditional Machine Learning Models



The above plot averages the F1-scores of all models, split by deep and traditional. It can be seen that the deep models outperform the traditional models by 14.08%, proving the efficacy of deep learning models over traditional machine learning models to perform well in the task of music classification by genre.

D. Best Performing Models

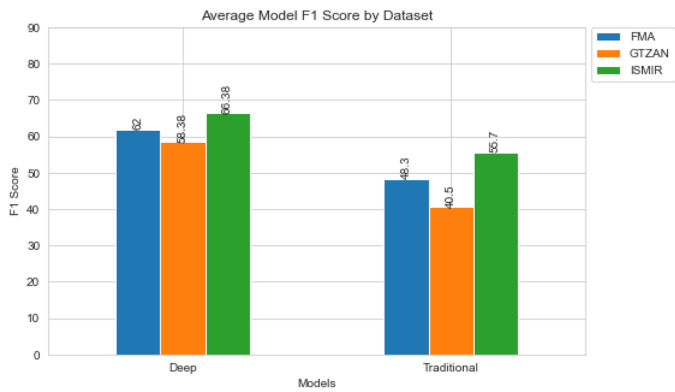


Above is a plot showing the averaged F1-score of all trained models, augmentation agnostic. As shown earlier when comparing the performance of the deep models to the traditional models, the highest performing models are deep models. More specifically, however, the VGG16 CNN was the best performer out all the models, with an average F1-score of 72.5%. It was followed by the Gujjula CNN, with an average F1-score of 66.83%. The third best performing model was the Schindler CNN, which obtained an average F1-score of 58.33%.

A notable mention from the traditional models is the neural network, which obtained an average F1-score of 57.67%, which is arguably negligibly less than the Schindler CNN, as there is only a 0.66% deficit in its average F1-score, compared to that. Now, neural networks because of their structure, namely the presence of hidden layers, are able to learn relevant features from the data, and that is why it performed better than all the other traditional machine learning models, which performed poorly overall, with their average F1-scores ranging between 35.17% in the worst case (CART model) and 57.67%in the best case (Neural Network).

From these results, it can be concluded that the traditional models are not complex enough to learn the relevant features from the data, and as such suffer in performance and are outperformed by the deep models, whose various network propagation and feature learning techniques, such as convolution and max pooling layers in CNNs, allow them to better learn and reduce the MFCCs to their most relevant features and subsequently classify the audio signals they represent better.

E. The generalisability of MFCCs and Data Augmentation



Above is a plot showing the average F1-scores of the models, split by traditional or deep, on the various datasets. It can be seen that the MFCCs were indeed able to serve as a general, easy to compute feature set, extractable from raw audio signals, and so able to generalise to different audio datasets that contain genre labels. The average F1-scores range from 62% to 66.38% on the deep models and from 40.5% to 56.7% on the traditional models. This is impressive as these scores prove the generalisability of MFCCs across datasets and models, even though data augmentation diminished the performance of the models overall.

It can also be seen that the average F1-score of all models on the ISMIR dataset is 61.57%, 55.15%on the FMA dataset and 49.44% on the GTZAN dataset. Given they are different datasets with different dataset sizes, audio signals, genres and number of genres, with the FMA dataset having 14 genres and the others having 10, some variance in the comparative performance was expected, and these results, are quite similar and fairly high with these factors considered, thus further supporting the case of the generalisability of MFCCs.

F. Future Improvements

I would like to have been able to make multiple runs of my models and averaged those runs out to get an even more accurate scoring of the model's performance, as well as to have included the training times of the models as a comparison metric, but running through the training of all models took a bit over a full day, and so having multiple runs to average out was not feasible given the onset of loadshedding. This is something that could be improved in future work, as well as incorporating feature selection on the datasets as part of the preprocessing phase to improve classification performance across all models, and further delving into the intricacies of data augmentation to improve its performance.

V. Conclusion

The first aim of this research was to present a processing pipeline from the audio signal level to the feature level, yielding a set of features that are well-performing and generalisable to different models and datasets, for the task of music classification by genre; as many other works use varied methods feature sets and data preprocessing methods. The secondary aim was to prove the proficiency of deep learning models over traditional machine learning models, when trained with features processed using this pipeline in the aforementioned task. The third aim was to test the generalisability of my proposed processing pipeline on different models and datasets; as many other works use just one dataset to test their models and implementations.

This was accomplished by augmenting the raw audio signals of each of the GTZAN, ISMIR and FMA datasets, through the processes of pitch-shifting and time-stretching, and then obtaining MFCCs from these augmented signals, which were used to train various deep and traditional machine learning models. It was found that indeed the deep learning models performed better than the traditional machine learning models, and that MFCCs do serve as a well-performing and generalisable set of features, from the similar performance of each model on the different datasets. However, it was also found that data augmentation worsened performance overall, and so should be excluded from the proposed processing pipeline. All in all, even with the worsened performance of the models brought forth by the data augmentation, the feature set attained from the proposed processing pipeline coupled with deep learning models, proved to be well-performing and generalisable to different datasets and models, given the consistent and similar high performance of all the deep models on all of the datasets. Future research could benefit from averaging multiple runs of each model, using feature selection practices on datasets and exploring and optimising data augmentation further.

References

[Log00] Beth Logan. "Mel Frequency Cepstral Coefficients for Music Modeling". In: *In International Symposium on Music Information Retrieval*. 2000.

[Fit06] W Tecumseh Fitch. "The biology and evolution of music: A comparative perspective". In: *Cognition* 100.1 (2006), pp. 173–215.

[Fen14] Tao Feng. "Deep learning for music genre classification". In: *private document* (2014).

[SD14] Siddharth Sigtia and Simon Dixon. "Improved music feature learning with deep neural networks". In: *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2014, pp. 6959–6963.

[SLR16] Alexander Schindler, Thomas Lidy, and Andreas Rauber. "Comparing Shallow versus Deep Neural Network Architectures for Automatic Music Genre Classification." In: *FMT*. 2016, pp. 17–21.

[Zha+16] Weibin Zhang et al. "Improved Music Genre Classification with Convolutional Neural Networks." In: *Interspeech*. 2016, pp. 3304–3308.

[Pan+17] Sharaj Panwar et al. "A deep learning approach for mapping music genres". In: *2017 12th System of Systems Engineering Conference (SoSE)*. IEEE. 2017, pp. 1–5.

[Elb+18] Ahmet Elbir et al. "Music genre classification and recommendation by using machine learning techniques". In: *2018 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE. 2018, pp. 1–5.

[VM18] S Vishnupriya and K Meenakshi. "Automatic music genre classification using convolution neural network". In: *2018 international conference on computer communication and informatics (ICCCI)*. IEEE. 2018, pp. 1–4.

[Chi+19] Snigdha Chillara et al. "Music genre classification using machine learning algorithms: a comparison". In: *Int Res J Eng Technol* 6.5 (2019), pp. 851–858.

[SLT19] Leisi Shi, Chen Li, and Lihua Tian. "Music genre classification based on chroma features and deep learning". In: *2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP)*. IEEE. 2019, pp. 81–86.

[NAJ21] Ndiatenda Ndou, Ritesh Ajoodha, and Ashwini Jadhav. "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches". In: *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE. 2021, pp. 1–6.

[LA22] Dhevan S Lau and Ritesh Ajoodha. "Music Genre Classification: A Comparative Study Between Deep Learning and Traditional Machine Learning Approaches". In: *Proceedings of Sixth International Congress on Information and Communication Technology*. Springer. 2022, pp. 239–247.

[Wir22] Business Wire. *Global Recorded Music Market Report 2022-2026 - increasing preference for on-demand music services &; growing demand for music-related leisure activities - researchandmarkets.com*. June 2022. URL: https://www.businesswire.com/news/home/20220606005634/en/Global-Recorded-Music-Market-Report-2022-2026---Increasing-Preference-for-On-demand-Music-Services-Growing-Demand-for-Music-Related-leisure-Activities---ResearchAndMarkets.com.

[Guj] Jahnavi Gujjula. *CNN_keras_cifar10*. Available at https://github.com/JahnaviGujjula/CNN_keras_cifar10/blob/master/cnn_keras.ipynb (2019/06/19).