1. Problem Statement

“If music be the food of love, play on” is a quote from William Shakespeare's play “Twelfth Night”. This phrase has since become a popular saying, often used to express the idea that music has the power to stir emotions and bring people together. Music Information Retrieval (MIR) is a field that focuses on extracting and analysing information from music. Music Genre Recognition (MGR) is an ongoing research area under MIR, which involves the categorization of music into society-defined genres based on the features extracted from the audio signals. Although the number of publications in MGR has been rising from 1995, [1], most of these papers are based on Western music datasets, such as GTZAN, Free Music Archive (FMA) and Million Song Dataset as mentioned in [1], [2]. To put it bluntly, South African music genres are not considered when MGR models are built. South African music genres include Kwaito, Kwela, Mbaqanga, South African house, Isicathamiya, Javia, South African Jazz, Marabi, Hip hop, Rock music, Amapiano and Gqom - 10 Exciting South African Music Genres You May Not know About. The GTZAN dataset has blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock. There only 2 common genres, and we must keep in mind of cultural perspectives of how a particular genre is defined [1]. The features extracted from the common genres between the Western and South African may be different. This gap was noticed [2] and leads us toward developing a music genre classifier suited to South African music genres.

2. Motivation

The most used dataset is GTZAN, which has been used to train models in over 354 of 980 of MIR publications [1]. This paper also mentions the faults that have been discovered in the GTZAN dataset in 2012; however, this dataset is still considered a benchmark dataset. GTZEN was used to train the models in [3], [4], [5]. [1] also gives the two other most used datasets, the Million Song Dataset and the Free Music Archive (FMA). This dataset, due to its huge data, has four various sizes. Namely, FMA Small, FMA Medium, FMA Large, and FMA Full. The FMA Large has 106574 30s audio files that are unequally distributed into 161 genres, while the Full has the same number of tracks but of full length. [3], [6] made use of FMA to model their machine learning. The Million Song Dataset is a collection of audio features and metadata for a million contemporary popular tracks, and the number of genres available is not given from their website, Welcome! | Million Song Dataset. This dataset was only used in one paper [3].

As one of the low-resource nations, it should come as no surprise that finding a proper and ready dataset that contains explicitly South African songs is difficult. Data has been an obstacle in the progress of AI development in Africa [7]. [2] had

to create their own dataset, with their primary source being YouTube videos for training their multimodal Sotho-Tswana music genre classifier.

There are two types of methods that are used in building models that classify a given music into their respective genres. There are traditional methods and Machine Learning methods. The traditional methods include Support Vector Machine (SVC), Logistic Regression, and k-Nearest Neighbour (kNN), while ML tools include Neural Networks (NN) such as Convolutional Neural Network (CNN), Artificial NN (ANN), and Recurrent NN (RNN). [3], [4] shows the incredible results achieved when the CNN and RNN are combined, known as CRNN.

2.1 Aim

This research aims to develop a model that will efficiently classify given South African songs in using machine learning tools and techniques.

2.2 Objectives

The objectives of this research are

i. to find a South African music dataset suitable for training a MGR model,
ii. to preprocess the music dataset,
iii. train the CNN and RNN model by applying Transfer Learning,
iv. to evaluate the performance of the classifying model, and
v. use multi framing strategy for hybrid genres.

2.3 Research Question or Hypothesis

From the aim and objectives, there are questions that arise. These questions include:

i. Does a dataset of South African music exist? And if it does exist, is it accessible? If not, will I have to make one?
ii. Will the use of multi framing, a strategy that allows the extraction of more than one frame from a single song, efficiently handle hybrid genres?

3. Methodology

3.1 Introduction

The purpose of this section is to provide a roadmap of the research, the phases in the building of the classification model for South African music genres using CNN and RNN while taking hybrid genres into consideration.

3.2 Data Collection

A request for the Music4All database [8] was sent to its creators, and an access to this dataset was provided.

[9] A sample rate, which is the number of samples taken from an audio's data per second when converting from analogue audio signal into a digital one of 22.050 Hz will be used to read the audio. This audio will be separated into 3s audio snippets for a length of 30s. To prevent the computer from assuming the audio clips to being unrelated, segment overlapping technique is used. This technique takes half of the earlier segment and append it to the next segment.

There are three features to be extracted from each segment, and they are:

1. Short-Term Fourier Transform (STFT)
   The signal is divided into small, fixed-duration segments called blocks. It then converts the time-domain representation into frequency-domain representation for each block. It will produce a 2D representation of the signal of Time vs Frequency.
2. Mel Spectrogram
   The STFT is then converted into a Mel Spectrogram, by mapping its frequencies onto the Mel scale, which reflects how humans perceive sound. A similar 2D representation is produced, but the frequency axis is warped according to the Mel scale.
3. Mel-Frequency Cepstral Coefficient (MFCC)
   A logarithm will be applied to the Mel Spectrogram to approximate the human perception of sound. To decorrelate the Mel Spectrogram coefficients, Discrete Cosine Transform (DCT) is applied, resulting in a set of coefficients that capture the spectral characteristics of the audio signal [9].

3.3 Classification Methodology

Model selection

1. CNN – this type of deep learning is one that uses pictures to categorize and recognize the genre of a given audio.
2. Long Short-Term Memory – a type of RNN that are effective at capturing temporal dependencies in sequential data.

Training process

The data will be split into three portions, the training data, validation data, and test data with the ratio 8:1:1 respectively. This data will put into a Hybrid Convolutional Neural Network that reshapes the data. After each epoch, which is a single complete pass through the training data,

the validation data will be used to be measure its performance when given unknown data. Once the completion of all trained networks is reached, the test data is to be used to evaluate the performance of the model.

Evaluation metrics

The evaluation metrics [9] for the model to be built are

1. Accuracy - measures of how well the correctly predicts the genre.
2. Precision - measures of how many positive predictions made that are actually true.
3. Recall - measures how many of the actual positive predictions are correctly identified.
4. F1-score - evaluates the accuracy of the model by combining precision and recall.

3.4 Conclusion

Convolutional Neural Network has shown its superiority in music genre classification as seen in [3], [5] and is mostly used for local feature extraction. Long Short-Term Memory is a variant of RNN for temporal patterns. By combining these two techniques, the accuracy of the model can be promising as seen in [4], [9].

4. Scientific Contribution

4.1 Representation of South African Music
This research could expand the scope of MIR research, making it inclusive of non-Western music.

4.2 Development of new datasets
If a suitable dataset appropriate for building the model for SA music genres cannot be found, the research will push us to curate a dataset. This will address the lack of diversity in existing datasets.

4.3 Advancement in feature extraction
South African music has unique rhythmic, melodic and harmonic features. This could lead to the development of new feature extraction techniques tailored to these features.

4.4 Improved Classification models
Using SA music genre could reveal insights into the strengths and limitations of existing techniques.

4.5 Cultural and educational impact
This project could help preserve and promote SA music by making it more accessible to the global audience. It could also help people learn about the cultural and historical significance of SA music genres.

5. Availability of Resources

For building the music genre classification model, there are four accessible datasets. These include GTZAN, FMA, Million Song Dataset and a recently created dataset called Music4All with 109 268 songs, each with a duration of 30s, and with an unknown number of genres. This dataset is yet to be explored to examine whether it would be appropriate for this research.

6. Ethical Considerations

6.1 Copyright and licensing

It is important to ensure that the audio data used in this research comply to copyright laws, by acquiring the necessary permissions to use the music files.

6.2 Transparency and accountability

The methods and datasets should be transparent, ensuring that the findings are reproducible and ethically sound.

7. References

[1] "A Critical Survey of Research in Music Genre Recognition," in *Proceedings of the 25th International Society for Music Information Retrieval*, 2024. Accessed: Feb. 23, 2025. [Online]. Available: https://ismir2024program.ismir.net/poster_149.html

[2] O. E. Oguike and M. Primus, "Multimodal Music Genre Classification of Sotho-Tswana Musical Videos," *IEEE Access*, vol. 13, pp. 28799–28808, 2025, doi: 10.1109/ACCESS.2025.3536026.

[3] "A Study on Music Genre Classification using Machine Learning," ResearchGate. Accessed: Feb. 24, 2025. [Online]. Available: https://www.researchgate.net/publication/370546962_A_Study_on_Music_Genre_Classification_using_Machine_Learning

[4] L. S, T. S, and B. M, "Music Genre Prediction Using Convolutional Recurrent NeuralNetwork," *Elementary Education Online*, vol. 20, no. 1, Art. no. 1, Mar. 2021.

[5] A. Juhukar, Y. Wagh, K. Nakashe, and S. Bane, "Classifico: Music Genre Prediction System using CNN," *IJFMR - International Journal For Multidisciplinary Research*, vol. 5, no. 3, doi: 10.36948/ijfmr.2023.v05i03.4124.

[6] L. Guo, "Music Genre Classification via Machine Learning".

[7] D. V. Marivate, "WHY AFRICAN NATURAL LANGUAGE PROCESSING NOW? A VIEW FROM SOUTH AFRICA #AFRICANLP".

[8] I. A. Pegoraro Santana *et al.*, "Music4All: A New Music Database and Its Applications," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Jul. 2020, pp. 399–404. doi: 10.1109/IWSSIP48289.2020.9145170.

[9] N. Narkhede, S. Mathur, A. Bhaskar, and M. Kalla, "Music genre classification and recognition using convolutional neural network," *Multimedia Tools and Applications*, vol. 84, pp. 1845–1860, Apr. 2024, doi: 10.1007/s11042-024-19243-3.