Music Genre classification Using LSTM and CNN

Abstract

Recommendation systems have become commonplace in recent years, and recommendation components have long been popular, leading to a reliance on the web to find a variety of options. Music companies are attracting as many customers as possible in the competition in the music market by provide individual music to their customers. Music suggestions considering the characteristics of music can increase the user's level of integrity in meeting and proceeding with customers. In this discussion, a music suggestion framework was developed by classifying melodies according to classes. Convolutional neural networks (CNN) and long-short-term memory convolutional neural networks (CNN-LSTM) are used as classification methods. Here CNN is a basic demo and the CNN-LSTM method exention of it. In this study, GTZAN dataset are used. A demonstration of the content-based (CB) proposal is used with cosine similarity and highlight extraction from recordings using Mel spectrograms. The CNN-LSTM method was best suited for CB recommendations for evaluation. In the future, classification can be performed using a hybrid CBCF system with a demonstration of combining the three classification models.

1. Introduction

Digital streaming platforms have captured the considerations of numerous individuals through computerized development and income sharing within the global music industry. In this recruitment advertisement, which is really difficult to apply to a first-class company, the owners of music benefits realized from the general premise that they needed to offer their customers more personalized and modern computerized items and a variety of products. Manufacturing learning, data processing and consulting are also very important in the music industry. Music management settings can be an important issue in this range. Accessibility to the programmed creation of personalized music playlists and how to exchange them between playlists based on specific characteristics. And Mesure will put a lot of effort into personalizing and researching music management.

As the proposed strategy became more common, cold start emergencies became a problem, and in the past it was a common problem where less than a certain number of clients were not found like audits or clicks. Cold start problem could ensure that unused passages are executed in a timely manner due to the immutability of proposals. In this review, we have sorted songs using neural tissue models such as CNNs and CNN-LSTMs. The CNN demonstration (Chiliguano, P. and Fazekas, G. (2016)) has been altered and altered to perform ideally on purpose questions and is considered a primary show. The CNN-LSTM show is based on this standard show with hyperparameter tuning and compares the results to the patterns shown by the CNN.

This includes the GTZAN data set, which also incorporates sound and metadata. The sound record is a wav array, the characteristics of which are extracted as Mel spectrograms. Leverage the CB proposal framework along with similarity metrics to provide suggestions. The reasons for choosing the overdose procedure are given in area 2. Classification precision, log unhappiness, is used to compare classification models. The proposed framework is judged by class correctness. We examine the framework as a whole by measuring how well it fits into the same review of proposals 5 and 10.

Extracts the highlights of a soundtrack within a GTZAN data set within a wave format. Neural network classification models such as CNN-LSTM are used for music classification classification and compared to CNN demos.

Section 2 shows all the operations performed on the recommended music system. Chapter 3 describes the methods used in this study. Section 4 shows the implementation of the model. Section 5 discusses model scoring and makes comparisons using graphs and metrics. Chapter 6 covers the conclusion and future work.

2. Related work

The music business these days is centered around music suggestion frameworks, and the main reason for their existence is to computerize the creation of custom playlists. What we've done is the evolution of the framework that allows us to change the playlist to suit the highlights of the final song, the user's current disposition, region and time. That way you can create original and professional sound applications that will benefit the music industry. This may be a study conducted by several analysts on the music proposal framework, considering employment as the most important part. Highlight extraction and classification models.

2.1. Datasets

Presenting modern models and machine learning computations to solve complex real-world problems is essential in any field of investigation, but comparing and evaluating them with the present state of artisan mentality is essential to the widespread application of innovations. To this end, it is important that the deliberations incorporate reasonably up-to-date information. The FMA data set incorporates both sound and metadata. It offers acceptable quality, sound for its size, unsurprisingly, actionable rich metadata, and easy-to-use, high-quality sound quality (De Gerard et al., 2016). It is also reasonably small and requires metadata. AcousticBrainz, AudioSet, and the Million Melody Dataset (MSD) are all considered contenders for larger datasets. The GTZAN data set contains 1,000 audio clips in 10 styles.

2.2. Feature Extraction

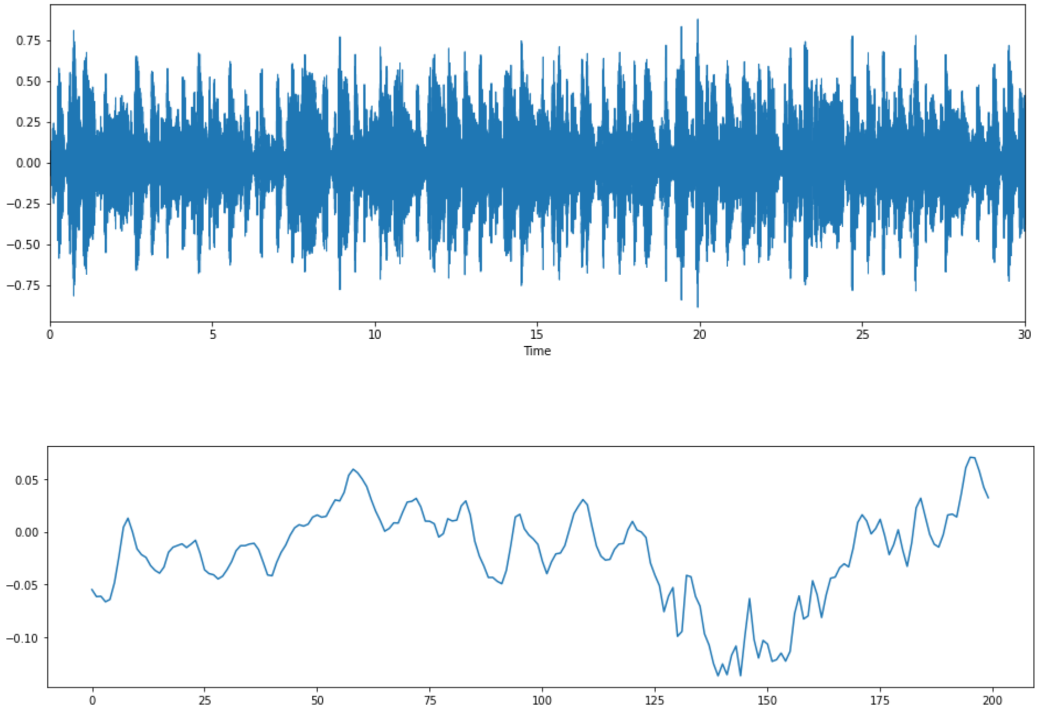

Figure 2. Zoomed Wave Signal

In this strategy, the crude sound flag is split into N windows and then executed M times. Tzanetakis, G. et al. (2002) used centroid, zero-crossing ratio, roll-off, and transmission capacity for characterization. The zero-crossing rate can define the number of times a flag changes sign in given period. Centroid represents the center of gravity of the frequency displayed in the Recursion Canister and is connected to the Recursion Space. Contrast within the weighted normal richness between repeat size and brightness is expressed as bandwidth.

Roll-off is described as standard repetition at the point in which the overall control value of a sound in a silent repetition reaches a certain rate that exceeds the normal range of control. Tzanetakis, G. et al. (2002) also utilizes differentiation, roll-off, and MFCC. The main difference is that the decibel difference at the top and bottom focuses on the signal range.

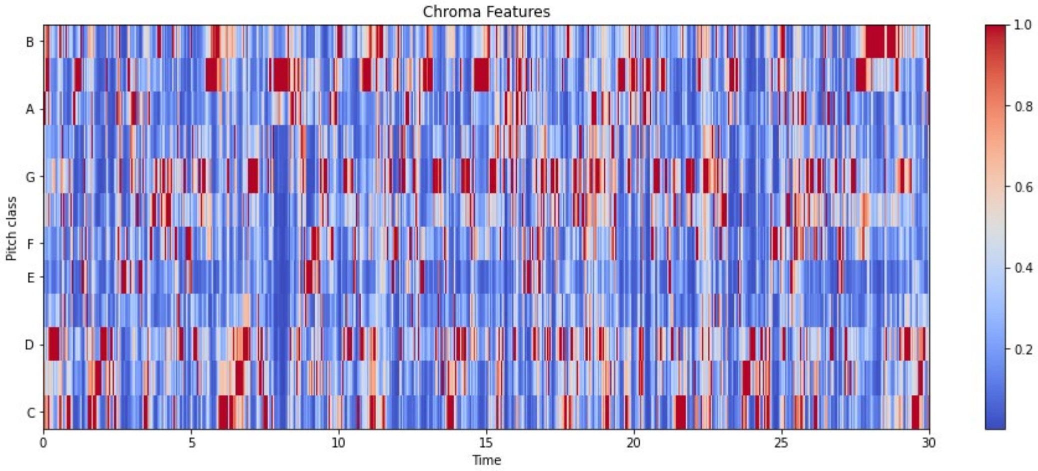

Figure 3. Chroma Features

Provides a point of interest for power changes within audio and audio preparation. MFCC is the ability to clarify the overall appearance of creepy patterns. Build a proposal framework based on MFCC and include similarities (Han et al. (2018)). MFCC quantizes the material and highlights the values. And through this, the characteristics of discourse can be extracted. This is the captured portion, adjusted for 30 seconds, starting with the steps in MFCC. Below that is a line opened using a differential pre-emphasis channel where high repetition sentences are highlighted to compensate for high repetition portions to remove negative flags, and high repetition portions are negative flags. After that, the perimeter and window are ready.

Finally, the discrete cosine change and Mel filter bank steps are performed. Nasrullah (2019) replaced the previously used MFCC to extract repeated material

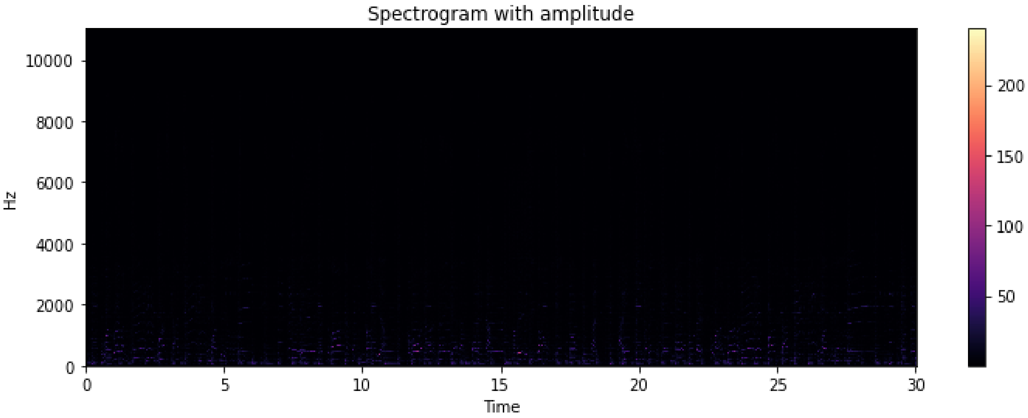

Figure 4. Spectrogram with Amplitude

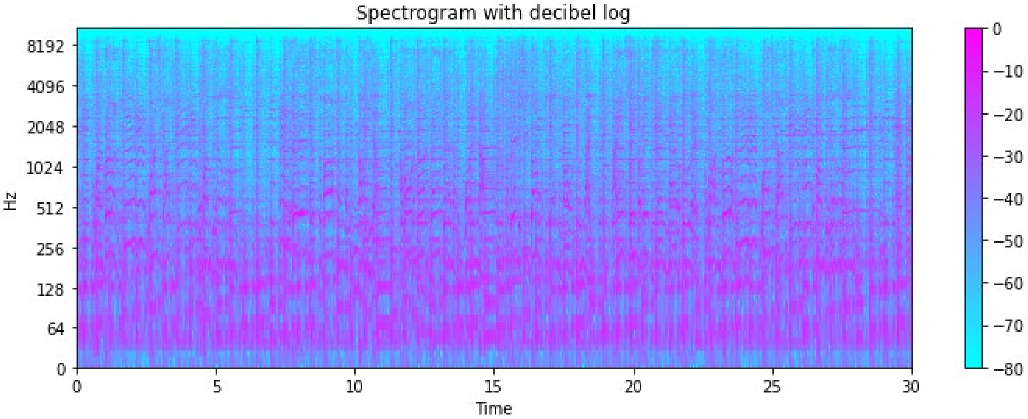

Figure 5. Spectrogram with dB Log

The spectrogram captures the frequency content and temporal changes of a 3-second audio sample. The 30 second soundtrack is divided into 6 parts and tested by Mel-spectrogram. Mel-spectrogram runs a whimsical test involving the extraction strategy and changes to mono channel by stacking each clip rising piece for up to 3 seconds at the test rate (22,050 Hz). Mel-scale driven spectrograms of 128 groups are calculated over 1,024 test windows using 512 test jump estimates for each part. This happens on 130 contour spectrograms with 128 components.

Finally, I changed the spectrogram to logarithmic in decibels. Simpson, A. J et al. (2015) evaluated Mel-spectrogram, MFCC and STFT representations and recently passed the demonstration. Better performance when using Mel-spectrogram than using STFT and MFCC. In conclusion, Mel-spectrogram is utilized within the model to include execution-based extraction and capture more characteristics of the sound.

2.3. Classification Models

Accurate music class classification is paramount in music suggestion and recovery. G. Tzanetakis et al. (2002) proposed a machine learning (ML) method and a profound learning show to classify classes. The suggested neural network provides an accuracy of 74%. This accuracy is higer than the used machine learning models such as KNN and SVM. It shows that neural systems outperform conventional machine learning methods. Dai, J. and Liang, S (2016) proposed to solve the ambiguous classification problem found in classifiers such as Bayes classifier for class classification of music, KNN, SVM, floppy classifier, neural system and quadratic discriminant test. Step 2 Mark the crosshairs. - Breed classifier. When realizing the proposed demonstration, it provides an accuracy of 90% which is inherently higher than other models used.

Simpson, A. J et al. (2015) conveniently proposed a convolutional neural system (FCN) for CB to compute computer music labels. In addition to subsampling layers, various models were evaluated for comparison with 2D convolutional layers. We propose a 4-layer FCN design and compare with four convolutional layers has two max pooling layers. Four-tier engineering is overwhelmed by deep models with extended hierarchies and deep organization that leverage vast amounts of preparatory information.

Asim, M. and Siddiqui, Z.A. (2017) joined deep convolutional neural networks (DCNNs) to extract musical information. DCNN used the soundtrack's metadata and Mel-spectrogram for classification. For faster merging, modified linear units (ReLU) are used for the sigmoid function. Chiliguano, P. and Fazekas, G. (2016) use a CNN approach to classify soundtracks into different classes based on the bits in the sound flags. Using ReLU with MaxPool, we show that an audio signal is transformed into a Mel-spectrogram of 599 contours with 128 repeat containers.

Elbir, A. et al (2020) proposed an method for classification based on the acoustic properties of soundtracks. Extract expressive highlights using new deep neural tissue. Fake dropout highlighting and unused layers included to reduce approval mistakes. Each layer is a compromise between a 2D convolutional layer, a ReLU enactment operation, a dropout layer and a 2D most extreme pooling layer.

Tao et al. (2019) proposes an LSTM show tuned to think about the portability of music and clients based on secular context and ongoing information. The learning rate of the program was scrutinized by the Adam optimizer to extend its accuracy. Sharma, S, 2018 show a hybrid method to classify using SVM and LSTM with extended precision. I got an accuracy of 89%, which is higher than the accuracy. It prepares the models individually and then combines them to display the final predictions. Arbitrary network shapes are used to tune hyperparameters to optimize values.

CLDNN combines CNN, LSTM, and DNN into one combined engineering. CLDNN has been shown to provide a 4% to 6% increase within WER compared to the most valid LSTM of the three models (utilizing Emam, A. et al. (2020)). It is simpler to utilize DNN layers to represent changes in LSTMs in a more isolated space and effectively predict yield targets.

Chang, S. et al., 2018 adopted a Convolutional Repetitive Neural Organizer (CRNN) for highlight extract features. Comparing the designs of CNNs and CRNNs, CRNNs are known to perform much better given their accuracy, reviews, and f1 scores. The CRNN demo includes two RNN layers with gate repeat units (GRUs) to summarize a two-dimensional design with four CNN layer yields. Tzanetakis et al. (2002) receive RNNs for group demonstration and CNNs for sound descriptor training. Layer 2 RNNs are based on LSTMs and incorporate various gates that provide support, demonstrating an understanding of what to ignore and what to keep in mind. Roughly, past testing of the internal state is realized. The wonder of the explosive gradients observed in RNNs and the problem of gradient decay actually tend to be caused by LSTMs.

For classification, Karunakaran et al. (2018) uses a profound residual bidirectional gated iterative neural network based on information augmentation to account for long-term conditions and ensure the legitimacy of data transmission through the remaining associative and bidirectional cells. With this in mind, CNN and CNN-LSTM computations are utilized within the demo for order classification.

2.4. Recommendation Systems

The proposed framework can be divided into two parts: CB channel and auxiliary gadget (CF). The CF grade framework is not based on fabric data. The cold start problem cannot be solved by leveraging CF. The CB rating framework focuses on actionable investigations to approve fabrics. These calculations can help reduce the effects of cold outbreaks by supporting untreated patients. Some frameworks use both cross CF and CB to perform negotiation. This allows the CB character to consider the quality of the fabric and the CF to consider what data the individual will collect. These hybrids can be designed with factors, sizes, specific combinations, combinations, updates, cascading, and meta-level frameworks through strategic combinations (Chiliguano and Fezcas, 2016).

Asim, M. and Siddiqui, ZA (2017) found that far is better for using the calibration relationship coefficients, Pearson relationship coefficients, and k-means. much better; higher; stronger; Found an improvement. This is a clustering calculation. Precision Precision is rated by Root Cruel Square Mistake (RMSE) and Cuminate Medium Blender (MAE). The same grid is used for classification and is used to sandwich some or nearly two components. It can be very important.

Used to determine what is needed (Asim, M. and Siddiqui, Z.A: 2017) mentioned four sizes: the Euclidean eye, the Manhattan eye, the Pearson eye, and the cosine eye. Of these, the cosine measure is the best known and best.

There are many papers using cosine similarity for performance evaluation. Split advanced handles into recommendations based CB, CF and metadata. The content of the metadata manual is conflicting and basic. CF needs a lot of information to figure out network issues and cold start issues. CB Consultation is a sound foundation and requires a small amount of customer data. A visual calculation of the soil mover serves as a visual metric. Finally, with this thought in mind, we took advantage of the CB recommendation by leveraging the well-known cosine image higher than the others due to the needs of the client data.

3. Methodology

3.1. Data Cleaning, Selection and Audio Pre-processing

An understanding of the purpose of a question is gained in the early stages of research, when the accumulated information is transformed into a machine learning problem. This institution reviews the gist of meditation and its overall arrangement. The gist of these considerations is to construct a music suggestion framework using profound learning techniques. The mechanized handle of song suggestions works with the master to utilize manual strategies to solve real-world problems. Melodies, also known as stay tracks, must be entered by the customer, and tracks with similar highlights must be entered and suggested by the performer. Musics companies will increase customer satisfaction and grow their industry by this help.

The GTZAN data set is utilized to be open and easy to open as specified in segment 2.1. Combined with high-quality pre-computed full-length sounds, highlights, user-level and track metadata, and more. A GTZAN small dataset consisting of 500 sound records in wav format. The samples has 30 seconds long and there are 10 balanced genres, and the total size of the data set is 760 MB. The dataset is available at https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification.

A small data set of 500 tracks with 30 tracks each is utilized within a wav format with 10 coordinated alignments. Soundtracks that are corrupted or less than 28 seconds long will be deleted. 6 audio tracks are deleted within the staging. I chose a 28 second cutoff to stretch every track into 10 parts. The soundtrack information of the mp3 composition is changed to a Mel Spectrogram.

Mel-spectrogram is the most powerful compared to other representations as described in section 2.2 (Simpson, A. J et al., 2015).

First, we process the sound flags using a specific inspection rate of 22,050. A window measurement of 2,048 (n fft) tests the input. Each time a deviation of the expected 512 (bounce length) occurs, the future period is tested. It also computes the Quick Fourier Change (FFT) of each window to change the time space into a repeating space.

Second, the Mel scale is produced when the entire repetition range is evenly distributed over the 128 (n Mel) frequencies. In each window decompose the signal magnitude into components in order to generate spectrogram and compare with frequency1 on the Mel scale. Figure 6 shows the test-sorted Mel-spectrogram.

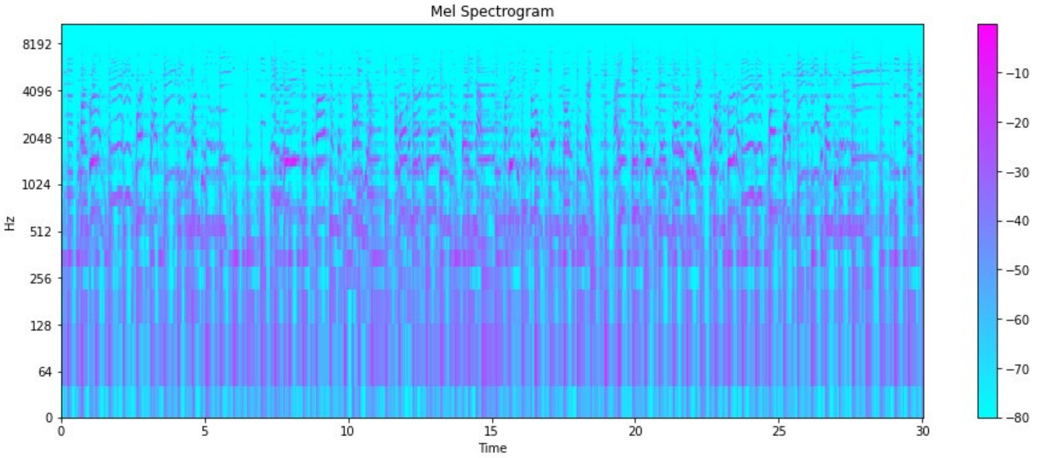

Figure 6. Mel Spectrogram

3.2. Neural Network Classification Models

CNN and CNN-LSTM models are used to highlight builds and classifications in this query. CNN is the primary program utilized for various inquiries, as specified in Area 2.3, and has been tuned for optimal operation. This model is used to extract important highlights within a proposal and classify genres. As seen in Sectin 2.2, they consider the iterative highlights of the spectrogram with time groups. These two models are reviewed below. Classify objects are used in CNN (Chang et al.; 2018; Simpson, A. J et al., 2015) and it contain spatial neighbors. Basically, CNNs consist of three types of iterative layers. That is, convolutional layers, drop layers and pooling layers. Within the convolution layer, numerous channels are utilized as containment locators and convolved after moving over input flags.

A configurable gait level oversees the individual movement of the filter relative to the input flag. Subsequent are downsampled to reduce computational control and overfitting within the pooling layer in measurement. The last layer of CNNs is fully connected(Swapnil et al., 2020). A CNN show is implemented like segment 2 and serves as a good classifier for several considerations. CCN has 24 layers consisting of 2x2 bits, group normalization layer, normal pooling 2D layer, soft layer, convolution layer with thick layer with ReLU and Softmax activations. LSTM.

LSTMs can solve long-term dependency problems, include input connections, and handle entire arrays of information2. The CNN-LSTM demo is run in region 2 where highlights are extracted into the CNN layer and grouping prediction is performed on the LSTM layer. CNN-LSTM consists of CNN and LSTM part. CNN has 26 layers with 23 layers and LSTM has 2 layers with 128 neurons each.

Recently, reconstruction and permutation layers have been included between the CNN and LSTM layers to include measurement reduction, so that the yield of the CNN is passed to the LSTM layer. The yield of the LSTM layer is given as a thicker layer that can produce a better inclusive representational array that can be effectively isolated into different classes as needed (Emam, A. et al., 2020).

3.3. Content-based recommendation model

The content-based suggestion framework suggests objects that are indistinguishable from objects that customers have recently used or rated for a while. This entity infers that the offer is already based on the preferred entity.

The CB proposal framework includes various similarity measures used to compute individual or similarity between two highlight vectors. In segment 2.4, the most common similarity metric, cosine similarity, is used (Asim, M. and Siddiqui, Z.A., 2017). The data set requires client information and the CB proposal framework is used to solve the cold start problem specified in segments 2.1 and 2.4.

Cosine similarity is used to computes the similiarity between two feature vectors. It gives the value as angle betwen two vectors and is calculated as shown in the equation below.

$$\cos(x, y) = \frac{xy}{\|x\|\|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

The function $c(x, y)$ has two values between -1 and 1.

4. Implementation

In this section, overall implementation are describes to develope music recommendation system. To implement the methodology mentioned in section 3, we perform the following procedure.

4.1. Data prepareration

A small data set from GTZAN is utilized and consists of 500 tracks of 30 tracks each within a wav format with 10 tuned classes. The cleaning handle is specified in area 3.1. Using Librosa library, we convert each track to a Melspectrogram. Mel-spectrograms are kept in buffer and cut into 10 parts for faster preparation. The cropped figure is changed at that point to the NumPy array used within the neural network classification model. For ideal utilization of RAM, the buffer is cleared as each track is ready.

The data is divided into three levels. The first thing to do is to deploy it to a NumPy cluster to recently prepare and test it. 500 melodies of all kinds are extracted and placed into NumPy clusters within the test segment. This test data is used to evaluate the precision of the proposed system based on its class. The NumPy cluster's readiness information is split and rearranged into pre-staging and testing within a 9:1 ratio utilizing Sklearn's demo decisions. The test information is used to evaluate the show, and the preparation information proceeds preparation and approval within the 9:1 ratio used in neural network classification models.

4.2. Neural Network Classification Models

The basic show CNN structure consists of 4 layers of 2D convolutions with channels (32-64-128-128). The input shape for the CNN demo is (128, 33, 1), where 33 is the number of highlights extracted. Each convolutional layer has two measures that are used to extract specific highlights from the input. There is also a swarm regularization layer with each convolution to extend and normalize the operation or input. Straight work is used to speed up meetings and reduce fading tilt issues.

The MaxPooling2D layer runs at the same speed while keeping less important data with a specific pool estimate for reduction. The 4 thick layers with neurons (32, 64, 128, 128) contain the linear working capacity for the base 3 layers, and since the classification is multi-class, the ultimate thick layer consists of softmax jobs.

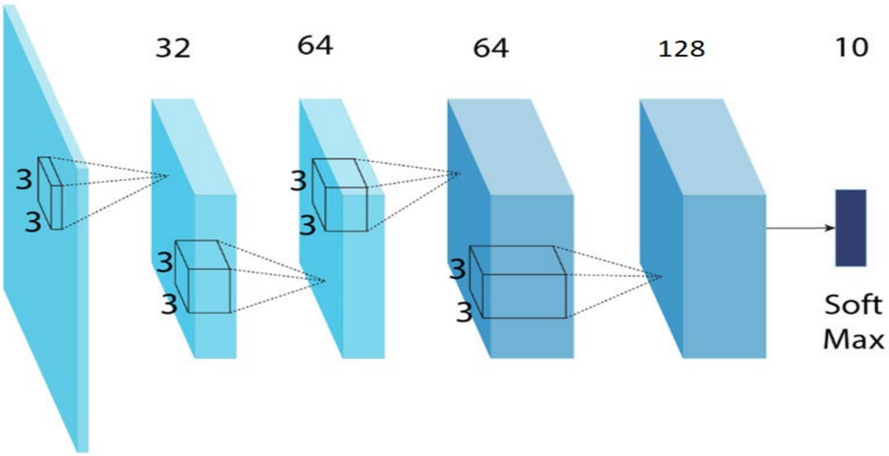

Figure 7. CNN Flowchart

TensorFlow and Keras are used to build CNN models. The streams and layers of a CNN show are shown in Figure 7. The CNN-LSTM demo consists of a CNN structure similar to the one specified above. Two LSTM layers are inserted after the convolutional fragment. To reduce suppression measures, permutation and reconstruction layers are included between the CNN and LSTM layers to ensure that the yield of the CNN layer is synchronized with the input of the LSTM layer.

The input shape of the layer is (128, 128, 1) and there are 128 neurons contained in each LSTM layer. After the LSTM layer are 5 thick layers with channels (1024-256-64-32-8). The CNN-LSTM show was built with Keras and Tensorflow. A CNN-LSTM flowchart show is shown in Figure 8.

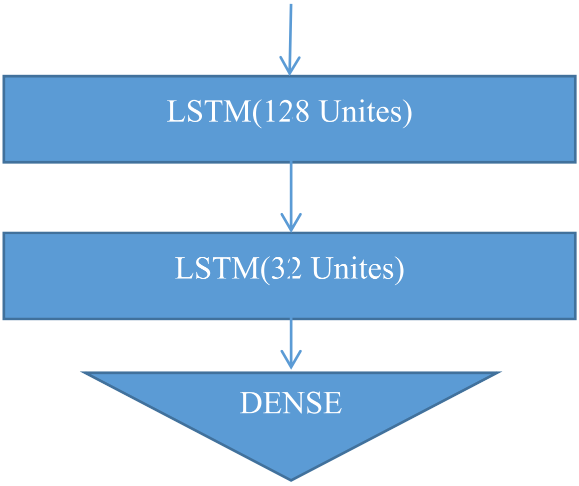

Figure 8: CNN-LSTM Flowchart

When generating the model design, several parameters were set correctly to provide good performance for the demonstration and ideally met the goals of this study. Dropout parameters are set to anticipate demonstrations of overfitting to staging information. Adjusting this hyperparameter prevents overfitting. Multiple dropouts of 0.5, 0.5 are set for CNN-LSTM. This can be a multi-class reflection, so categorical cross entropy is utilized. Adam optimizer takes advantage of this idea, and Modelcheckpoint, which is part of Keras' callback operation, puts the demo into a specific focus while preparing the weights. Shows with high accuracy and ideal unhappiness points are considered. These two models operated in the 50's and 150's.

CB proposal show that how to use an optimal neural network classification model as HDF5 format. Test information is passed to the demo. Affinity scores are calculated for all songs associated with the Grapple soundtrack. The cosine similarity metric suggests a melody similar to taking a soundtrack from information entered into a program. This part is implemented in Google Colab and you can design any number of proposals based on the prerequisites.

5. Evaluation, Results & Discussions

This section is divided into two stages. First, the classification model of neural arrays is evaluated based on the precision of classification, logarithmic misery. Also, the proposal show is evaluated based on the precision of the class with 5 or 10 proposals. Post evaluation is the premise under which the demonstration is evaluated.

5.1. Classification Model Evaluation

Classification exactness and logarithmic misfortune are the basis of assessment of these two neural organize models of classification. Table 1 contains the assessment measurements for the two models which consider the test information. As you can see, the CNN-LSTM show exhibit its excellent exactness between the two with great exactness, exactness.

Table1. Classification model evaluation with test data

Classification Models	Accuracy	Loss
CNN	0.62	2.413
CNN-LSTM	0.68	1.77

With limited information and short arrays of 3 seconds each passed to the LSTM layer, the precision may not increase.

After that, the CNN-LSTM show can be seen as the leading show in class classification. Figure 9 and Figure 10 are algebraic graphs considering the preparation approval information and show the classification accuracy results of CNN and CNN-LSTM models, respectively.

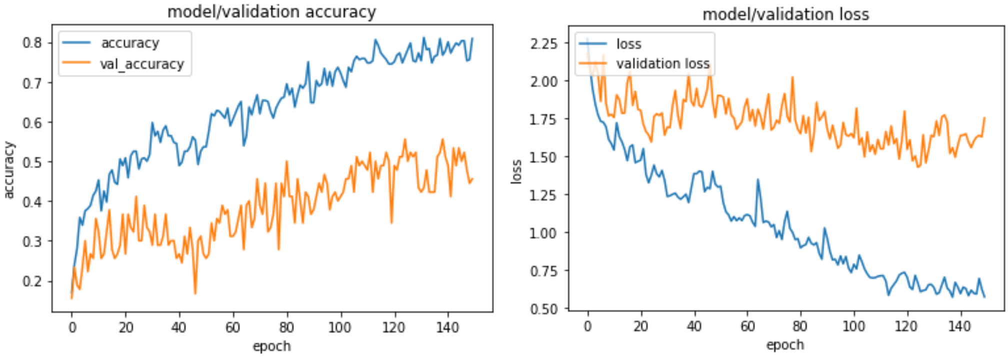

Figure 9. Train-Validation Accuracy and Loss of CNN model

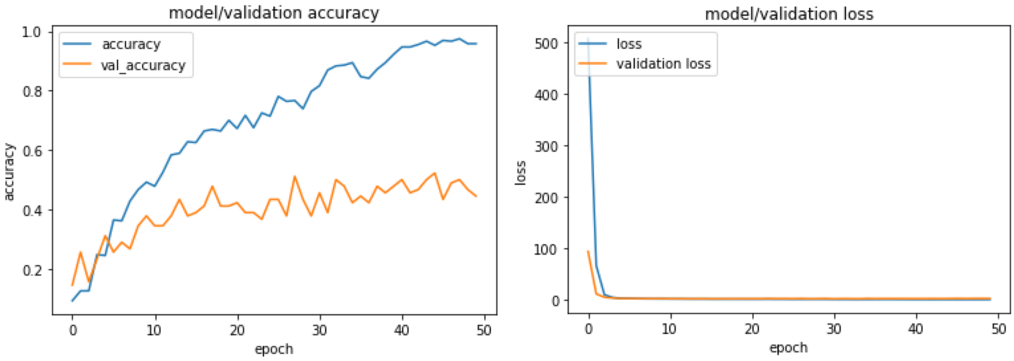

Figure 10. Train-Validation Accuracy and Loss of LSTM model

The blue line indicates train and yellow indicates validation. We observed that the CNN-LSTM show had the highest receptive precision with progressively extended smooth bending.

5.2. Results & Discussions

It showed the best performance compared to CNN-LSTM. The combination of LSTM and CNN to handle grouping provides advanced classification, as shown in the results in section 5.1. Despite the fact that the precision can be extended by using longer arrays for input into the LSTM layer. We leveraged CB's proposal framework due to datasets missing client data.

All models with a recommendation system shared a high level of precision. In any case, instrument grading was less accurate. Because the data used to estimate the proposal is randomly chosen, it is usually incomplete and consequently subject to information asymmetry. Instrumental music classes are also exceptional, with highlights similar to some classes, such as shakes. Therefore, the instrumental sort track can suggest a shake genre track.

6. Conclusion and Future Work

Recommendation systems are becoming increasingly important and widely implemented. This framework can provide a prescribed soundtrack to clients based on the proximity of sound highlights. In this idea, we utilize a dataset from FMA with sound information and extract the highlights using Mel-spectrogram.

Utilize CNN and CNN-LSTM models to classify music that agrees to a class. The CNN show was adapted and utilized as the main show. The CNN-LSTM demo is based on a basic show and results are evaluated. The CNN-LSTM demonstration has been shown to outperform the rest when considering classification accuracy and misfortune. Since there is no client information in the data set, we solve the cold start problem using the CB proposal framework, which has cosine similarity to the proposal.

Music recommendation considering the class accuracy of the two models showed the best results by showing cosine proximity in the CNN-LSTM demonstration. In the future, we propose to utilize the client information as sound highlights and build a crossover proposal framework combining CF and CB. A larger array is entered into the LSTM layer as input to upgrade classification and good accuracy. It can realize a costume show where all the models it provides can be composed of a combination of CNN and CNN-LSTM models.

References

- Asim, M., Siddiqui, Z. A. (2017). Automatic Music Genres Classification using Machine Learning, International Journal of Advanced Computer Science and Applications (IJACSA), vol. 8, no. 8, pp. 337-344.
- Bakhshizadeh, M., Moeini, A., Latifi, M. and Mahmoudi, M. T. (2019). Automated mood based music playlist generation by clustering the audio

features, 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 231~237.

- Chang, S. and Abdul, A. et al. (2018). A personalized music recommendation system using convolutional neural networks approach, 2018 IEEE International Conference on Applied System Invention (ICASI): 13~17.
- Chiliguano, P. and Fazekas, G. (2016). Hybrid music recommender using content-based and social information, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2618~2622.
- Costaa, Y. M. G., Oliveira, L. S., Silla, C. N. (2016). "An evaluation of Convolutional Neural Networks for music classification using spectrograms", Elsevier B.V. 2016 Applied Soft Computing 52, vol. 52, pp. 28-38.
- Dai, J., Liang, S., Xue, W., et al. (2016). Long Short-term Memory Recurrent Neural Network based Segment Features for Music Genre Classification, in 10th International Symposium on Chinese Spoken Language Processing (ISCSLP).
- Elbir, A. and Aydin, N. (2020). Music genre classification and music recommendation by using deep learning, Electronics Letters 56(12): 627~629.
- Elbir, A., Bilal C. and, H., Emre I. M., Ozt¨urk, B. and Aydin, N. (2018). Music genre classification and recommendation by using machine learning techniques, 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1~5.
- Emam, A., Shalaby, M., Aboelazm, M. A., et al. (2020). A Comparative Study between CNN, LSTM, and CLDNN Models in The Context of Radio Modulation Classification, 2020 12th International Conference on Electrical Engineering (ICEENG) pp. 1~8.

- Han, H., Luo, X., Yang, T. and Shi, Y. (2018). Music recommendation based on feature similarity, 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), pp. 650~654.
- Swapnil, S. S., Mani, B. S. (2020). https://www.researchgate.net/publication/341055525_Deep_Convolutional_Bidirectional_LSTM_for_Complex_Activity_Recognition_with_Missing_Data.
- Karunakaran, N., Arya, A. (2018). A scalable hybrid classifier for music genre classification using machine learning concepts and spark, 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), pp. 128~135.
- Sharma, S., Fulzele, P. and Sreedevi, I. (2018). Novel hybrid model for music genre classification based on support vector machine, 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp
- Simpson, A., J, Roma, G. and Plumbley, M. D. (2015). Deep karaoke: Extracting vocals from musical mixtures using a convolutional deep neural network. arXiv preprint arXiv:1504.04658.
- Tao, Y., et al. (2019). Attentive context-aware music recommendation, 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), pp. 54~61.
- Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing, pp. 293~302.
- Tzanetakis, G., Essl, G. and Cook, P. (2002). "Automatic musical genre classification Of audio signals", Speech and Audio Processing IEEE, pp. 293-302.