International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

EVALUATING DEEP LEARNING MODELS FOR MUSIC EMOTION RECOGNITION

Makarand Velankar, Sneha Thombre, Harshad Wadkar
Department of IT
MKSSS's Cummins College of Engineering for women,
Pune, Maharashtra, India.

Abstract—**Music listening helps people not only for entertainment, but also to reduce emotional stress in their daily lives. People nowadays tend to use online music streaming services such as Spotify, Amazon Music, Google Play Music, etc. rather than storing the songs on their devices. The songs in these streaming services are categorized into different emotional labels such as happy, sad, romantic, devotional, etc. In the music streaming applications, the songs are manually tagged with their emotional categories for music recommendation. Considering the growth of music on different social media platforms and the internet, the need for automatic tagging will increase in coming time. The work presented deals with training the deep learning model for automatic emotional tagging. It covers implementation of two different deep learning architectures for classifying the audio files using the Mel-spectrogram of music audio. The first architecture proposed is Convolutional Recurrent Model (CRNN) and the second architecture is a Parallel Convolutional Recurrent Model (Parallel CNN). Both the architectures exploit the combined features of Convolutional and Recurrent layers. This combination is used to extract features from time and frequency domains. The results with accuracies in the range of 51 to 54 % are promising for both models for a small dataset of 138 songs, considering the large datasets required for training deep learning models.**

Keywords—**Classification, Convolutional Neural Network, Convolutional Recurrent Neural Network, Deep Learning, Mel-spectrogram, Music Emotion Recognition**

I. INTRODUCTION

Music plays a very crucial role in one's life, whether it is gaining pleasure in listening, the emotional response, creating or performing. It is an expression of emotions through musical facets and ornamentation. Music creation is no longer an activity performed only in studios. People create and upload music files through various social media platforms without any tags associated with it. Manual tagging of these music files is tiresome and impossible considering the growth of music files like big data explosion on various platforms.

The songs in the music streaming apps are organized based on different categories such as genre, artist, year, emotion, etc. for a personalized recommendation. Music plays a significant role in altering a person's emotional state of mind. The main disadvantage of these applications is the poor recommendation when it comes to analyzing the music according to the user's preference based on their emotions. Mostly people get frustrated when they don't get the songs which they desire to listen to and hence they end up spoiling their mood and their moment; therefore, it is very crucial that the labeling of music in such music streaming services into different emotional states must be very accurate. The tagging of music into different emotion categories such as romantic, sad, happy, etc. in the online music streaming services is done manually and hence it depends on the perception of the person tagging it. Automating this tagging process will save time and will be more accurate if based on proven machine learning model. It is useful for music recommendation as well.

With these future requirements in mind, this work proposes two different models for the automatic classification of music using a deep learning framework and music audio spectrogram. The two models proposed are namely, Convolutional Recurrent Neural Network (CRNN) and Parallel CNN RNN. In the CRNN model, the convolutional and the recurrent layer extract features one after the other from the spectrogram created from the music audio files to perform classification. In case of the Parallel CNN RNN model, the convolutional and the recurrent layer extract features in parallel from the spectrogram created from the music audio files to perform classification. The two models exploit the architecture that combines Convolutional Neural Network (CNN) as well as a Recurrent Neural Network (RNN) to classify music clips into different emotion categories. The preliminary work is done in three popular emotional categories, namely romantic, sad, and a devotional for the Indian songs. The main contributions of our work are summarized as follows:

1. The dataset is created for Indian (Hindi) songs. Mel spectrogram created for each audio file is divided fairly for training, validation, and testing.
2. Two models Convolutional Recurrent Neural Network (CRNN) and Parallel CNN RNN proposed for music emotion classification. Both models use the

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

Convolutional Neural Network and the Recurrent Neural Network as the basic framework.

3. Evaluation of proposed deep learning models for the music emotion recognition.

II. RELATED WORK

It is quite evident that various kinds of music influence our emotions in various ways. Numerous studies have already been done to investigate the relationship between music and emotions since long time. A comprehensive overview of the emotion recognition task of music was provided by Barthet et al. [1]. Weiczorkowsa et al. [2] were the first ones to propose the emotion recognition task of music as a problem of multi-label classification. Researchers have tried many classifiers for classifying different music into different emotional categories. Some of them include Binary Relevance K Nearest Neighbors, Multi-Label K Nearest Neighbors [3], Back-propagation for Multi-Label Learning, Random k label sets [4], Calibrated label ranking classifier using a support vector machine [5], etc. Out of these classifiers, the Calibrated label ranking classifier using a support vector machine outperforms the rest of the above-mentioned classifiers. Research on the psychological response to the context has shown that there is an effective response to music depending on the context and the environment of listening [6]. It is evident from earlier research that there is a clear difference between the perceptions of emotion induced by the songs and emotion expressed in the songs [7]. Various features have been taken into consideration for the classification of music into different emotion categories. These features include acoustic features [9], rhythmic features [9], timber features [10, 11], spectral features, lyrics [8], etc. The emotion recognition of music is majorly done by extracting these features [14], where acoustic features are used mainly [12]. Most of the speech emotion recognition used Mel Frequency Cepstral Coefficients (MFCC) [13] which also proved to be efficient for classification of audio.

Although there have been a lot of improvements made in the classification of music based on emotions and moods, but the state-of-the-art recognition of emotion in music is still a challenging research problem. The feature extraction using machine learning techniques is bit complex and requires a lot of efforts related to tasks such as feature selection and reduction. In order to overcome these issues, Convolution Neural Network (CNN) architecture was proposed [14]. CNN has remarkable properties to extract and represent high-level music features [15]. It was observed that the CNN used for the conventional image classification could be used the audio features [16]. Though the CNN architecture helped us to provide an easy way of feature extraction the experimental outcome showed that these models were not robust enough. In order to improve the accuracy and provide more statistical information deep residual learning for image classification was introduced [17]. Some techniques also used ANN for genre recognition of music [18]. Since music is a complex blending of frequencies over time, a combination of CNN along with a Recurrent Neural Network was introduced for music recognition [19, 20].

After the remarkable success of deep learning models for speech recognition, recently, emotion recognition in speech using deep learning was explored by researchers [21,22, 23]. Recently similar research was attempted to use deep learning for the music emotion recognition [24, 25, 26]. As per our findings, the use of deep learning models for the music emotion recognition is still in the infancy stage and it will mature over a period of time. Successful use of deep learning for text classification [27] or in recent pandemic situations [28] makes it a promising proven approach. This led us to explore possible deep learning models for music emotion classification.

Music emotions or sentiment analysis has been a topic of research and different approaches have been proposed such as use of lyrics [29], multimodal approach [30] and use of machine learning [31]. Use of soft computing approach with neural network [32] is further extended with deep neural networks. The deep learning approach is promising approach and the work presented here attempts to explore it for the music emotion recognition using 2 models named as 1. CRNN model and 2. Parallel CNN and RNN model.

III. METHODOLOGY

Two strategies for classifying music are proposed into different emotional categories. Figure 1 illustrates the framework of our proposed method. The framework is divided into three main parts: Dataset creation, Data preprocessing, and Classification method. The dataset is created from the existing tags and using a collection of Indian (Hindi) songs. It has 3 emotion categories and has a total of 138 song samples of 30 seconds each. The 46 songs for each category are randomly selected from the websites and popular streaming websites.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

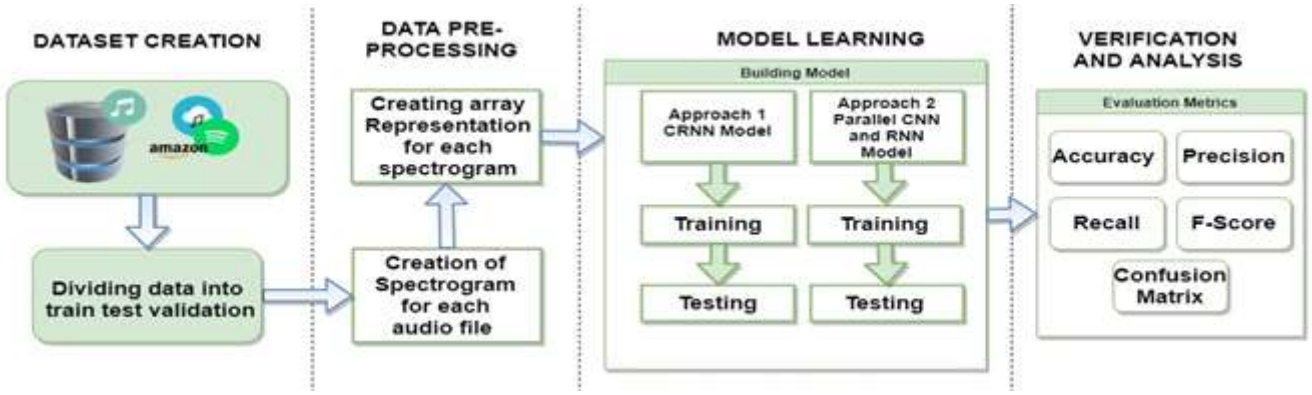

Fig.1. Proposed Framework

A. Data Preprocessing

Each song in the dataset is transformed into a Mel spectrogram using the algorithm and implemented in python. A spectrogram is a visual depiction of the spectrum of frequencies of the audio signals varying with time. It is the squared magnitude of the short-term Fourier transforms (STFT) of the audio signals. The spectrogram is flattened or compressed to convert the audible frequencies into a human-understandable form. This squashing of the spectrogram is done by using the Mel scale. Window length and hop length are the most important parameters used in the transformation. The window of time to execute Fourier Transform on is called the Window length. The number of observations between frames in succession is called the hop length. The shortest reasonable duration for a human ear to recognize an audio signal is 10ms. Hence, for the conversion of Mel spectrogram for each song a window length of 2048 and a hop length of 512 was set.

The python library used produces a Mel spectrogram which is scaled using a log function. The audio data are mapped to the standard logarithmic scale used for assessing the loudness in decibels (dB) as it corresponds to the pitch perceivable by humans. This transformation results in the creation of a Mel spectrogram having shape: 640, 128. In order to accelerate the training process, the dataset is divided into train, validation, and test, each audio file of the dataset is converted into their respective Mel spectrograms and the results are picked.

B. Combination of CNN and RNN

One question which emerges is why one should use the combination CNN and RNN? A spectrogram is a visual depiction of the spectrum of frequencies of the audio signals varying with time. It is like an image having distinct patterns for each song, hence it makes sense to use CNN. RNN succeeds in the interpretation of sequential data by making the hidden state at a time t based on the hidden state at time t-1. The spectrograms have a time dimension. So RNN is likely to do a much better job of determining the audio's temporal long term and short-term characteristics. Therefore, the blend of RNN and CNN allows analyzing the audios in detail, and thus the combination likely to improve the classification accuracy. For music data which is temporal, RNN is more suitable. Use of CNN further likely to improve the automatic feature extraction in the deep learning model used.

C. Convolutional Recurrent Neural Network approach

CNN is mostly exploited for tasks related to image recognition. Instead of matrix multiplication, it performs a convolution operation and for discerning the two-dimensional (2D) layout of the data convolution is usually employed in the initial layers. On the other hand, RNN excels in the interpretation of time series data using the hidden states at different time instances. The neural network built using one dimensional (1D) convolution layers. The convolution operations are performed over the time dimension. As shown in Figure 2 each 1D convolution layer derives characteristics from a small portion of the Mel spectrogram. After convolution, RELU activation is put in use. After applying the activation function, batch normalization [33] is performed and eventually, 1D max-pooling is applied. This reduces the image's spatial dimension and avoids overfitting. The chain of operation, 1D convolution, RELU, batch normalization, and 1D max pooling, is executed thrice. The key configurations exploited here are, 56 filters are used per layer and a kernel size is set to 5.

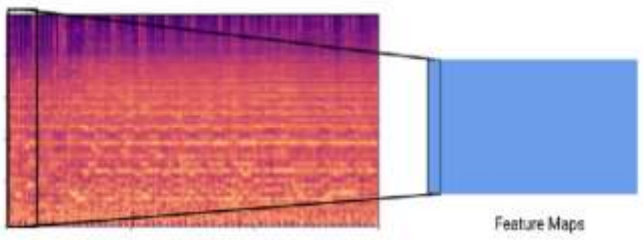

FIG 2. 1D CONVOLUTION BY CRNN MODEL

In order to find the long term and the short-term structure of the audio, the outcome of the 1D convolution layer is augmented into an LSTM [34, 35] having 96 hidden units. The LSTM output is fed into a 64-unit dense layer. The model's

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

final output layer is a dense layer with SoftMax activation. It has 8 hidden units in order to determine the probabilities of the three classes. To prevent overfitting of the model both L2 and dropout [28] regularization was used between all the layers. Figure 3 shows the overall architecture of the model.

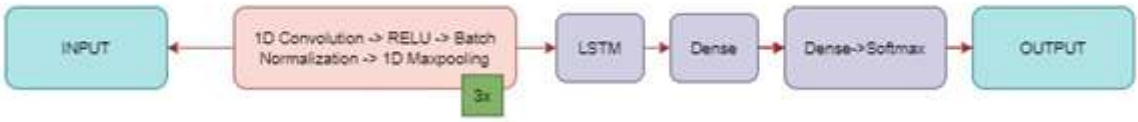

Fig.3. Architecture of CRNN

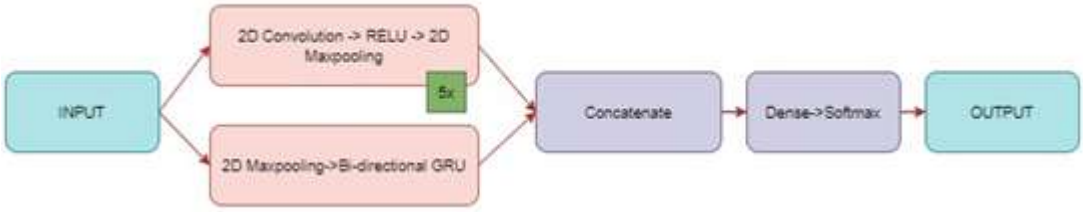

Fig 4: Architecture of Parallel CNN RNN

D. Parallel CNN RNN approach

The key idea behind this approach is that although CRNN has RNN to summarize the temporal feature, it summarizes the temporal features after the CNN output. During the operations with CNN, the temporal relationships of the original audio signals cannot be preserved. In Parallel CNN RNN approach the spectrogram inputs are passed to both, CNN as well as RNN, in parallel. The outputs from both the frameworks are then concatenated and sent through a dense layer having the SoftMax activation function. While the CRNN model uses a 1D convolution and max-pooling layer, the Parallel CNN RNN model uses a 2-dimensional convolution layer followed by a 2-dimensional max-pooling layer. This model has 5 blocks of convolution max-pooling layer. The size of the kernel is 3,1 for all the blocks. For the first block, the size of the filter is 16, for the second block it is 32, and for the remaining blocks, it is 64. After each convolution, RELU activation is applied. The ultimate output is flattened. The final output is a tensor having shape; 256.

With the 2D max-pooling layer of the pool having size (4, 2) the recurrent block starts to reduce the spectrogram size before LSTM operation. This reduction in the feature was mainly performed to accelerate the processing. The diminished image is forwarded to a 64-unit bidirectional GRU [36]. This layer's output is type tensor, 128.

The outputs from both recurrent and convolution frameworks are then concatenated resulting in a tensor having a shape of 384. Eventually, it has a dense layer with SoftMax activation. Figure 4 depicts the Parallel CNN RNN architecture.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The outputs from both recurrent and convolution frameworks are then concatenated resulting in a tensor having a shape of 384. Eventually, it has a dense layer with SoftMax activation. Figure 4 depicts the Parallel CNN RNN architecture.

A. Evaluation metrics

In order to evaluate the performance of the proposed models, parameters such as accuracy, precision, recall, f-score, and confusion matrix are used. Accuracy is the ratio between the correctly predicted outcome and total observation. Precision is the ratio between correctly predicted positive outcomes and the total number of predicted positive outcomes. The recall is the ratio between the correctly predicted positive outcome and the total number of observations in the actual class. The weighted average value of precision and recall is called f1-sore. Both the models were trained for 25 epochs.

B. Experimental results

This section covers the evaluation of the performance of both the proposed models and analyze their results. In the case of the CRNN model, the best-trained model had an overall accuracy of 54% on the test set after hyperparameter tuning. Figure 5.1 depicts the accuracy curve and Figure 5.2 depicts the loss curve for training and validation. Figure 6 depicts the precision, recall and F1 score of the CRNN model on the test set. Figure 7 illustrates the confusion matrix of the CRNN model for the test data. The results show better accuracy in the sad class compared to other emotional classes.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

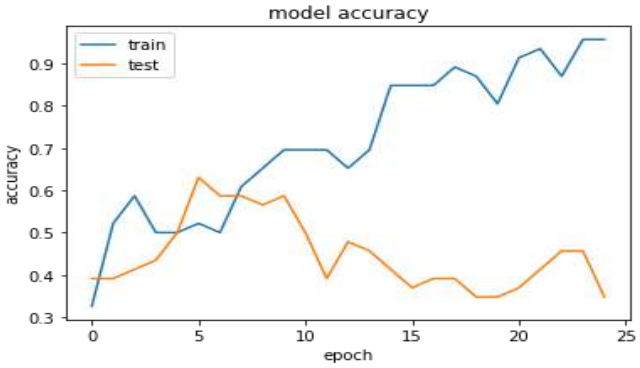

Fig 5.1. Accuracy Curve for CRNN approach

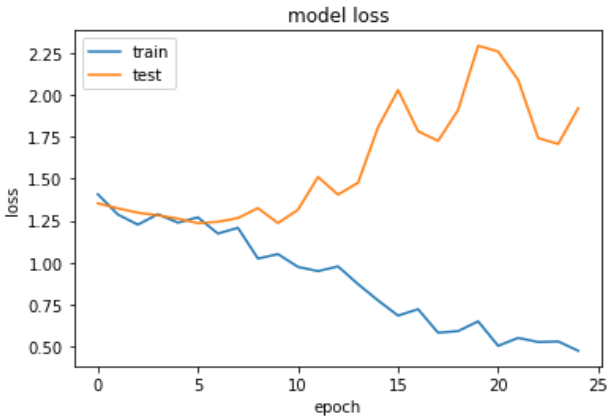

Fig 5.2. Loss Curve for CRNN approach

	precision	recall	f1-score	support
Devotional	0.47	0.47	0.47	15
Sad	0.65	0.69	0.67	16
Romantic	0.50	0.47	0.48	15
accuracy			0.54	46
macro avg	0.54	0.54	0.54	46
weighted avg	0.54	0.54	0.54	46

Fig 6. Precision, Recall and F1 score of CRNN model

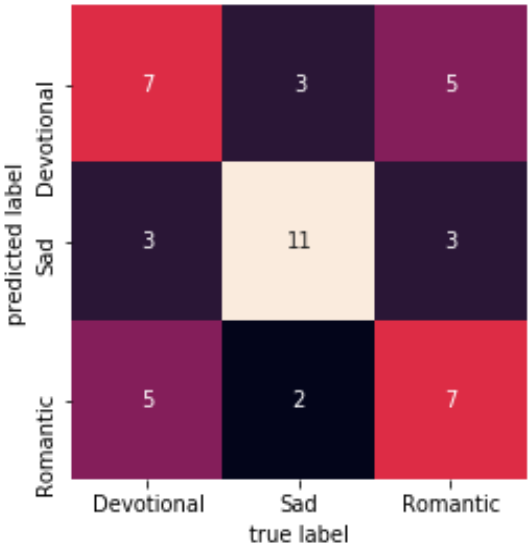

Fig 7. Confusion Matrix for CRNN model

In the case of the Parallel CNN RNN model, the best-trained model had an overall accuracy of 50% on the test set after hyperparameter tuning. Figure 8.1 and Figure 8.2 illustrates the accuracy and loss curves respectively, for training and validation samples. Figure 9 depicts the precision, recall and F1 score of the Parallel CNN RNN model on the test set and Figure 10 illustrates the confusion matrix of the Parallel CNN RNN model. The results show better accuracy in devotional class compared to other emotional classes.

The overall accuracy is better for CRNN model than the Parallel CNN model. The results are promising considering small dataset used for training the model. This classification approach may also give poor results when there is no mere difference in the spectrograms belonging to different categories. For instance, consider a sad song with high beats which displays the grief of the human and a happy song with high beats showing that the human is super excited. Both spectrograms will have similar features that may confuse the neural network during the classification. In such a case, additional features for audio or lyrics may be useful to improve the automatic classification.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 6, ISSN No. 2455-2143, Pages 252-259
Published Online October 2022 in IJEAST (http://www.ijeast.com)

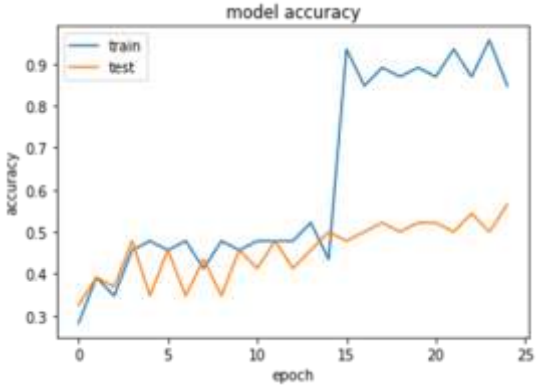

Fig 8.1. Accuracy Curve for Parallel CNN RNN approach

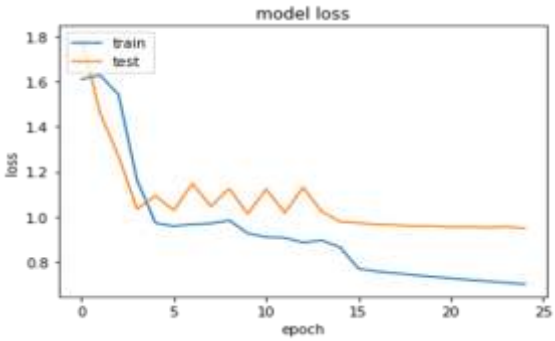

Fig 8.2. Loss Curve for Parallel CNN RNN approach

	precision	recall	f1-score	support
Devotional	0.45	0.60	0.51	15
Sad	0.50	0.44	0.47	16
Romantic	0.58	0.47	0.52	15
accuracy			0.50	46
macro avg	0.51	0.50	0.50	46
weighted avg	0.51	0.50	0.50	46

Fig 9. Precision, Recall and F1 score of Parallel CNN RNN model

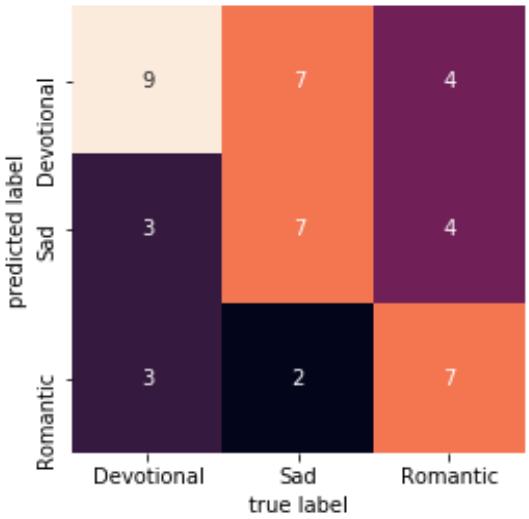

Fig 10. Confusion matrix for Parallel CNN RNN model

V. CONCLUSION AND FUTURE SCOPE

The proposed deep learning models derive essential features from Mel spectrograms which are useful in the classification of music into different emotional classes. This plays a crucial role in the recommendation of music in online music streaming services. Two models proposed based on CNN and RNN frameworks, which are CRNN approach and the parallel CNN RNN approach provides promising initial results for further experimentation. The experimental analyses for the evaluation of these models were done on the small dataset created. The dataset of Indian songs with validated emotional labels is useful for researchers working in the domain of music emotion recognition. The classification accuracy will be further improved with increased number of samples used per category. Typically, Deep learning model uses millions of samples to train the model. The current experiments prove the methodology and the utility of models for the music emotion recognition, with accuracy in the range of 51 to 55% for 3 classes. The trained model with a larger data set and having acceptable accuracies will be used for automatic tagging of large music data generated in future.

The future work will focus on dataset to increase the size of the dataset and which will improve the accuracy of the classification models. In the proposed work instead of only the spectrogram features, more features such as MFCC, lyrics, metadata, etc. can be added for better classification. It definitely does not take longer than 5-10 seconds for humans to decide the emotion of the songs. So, present samples duration of 30 seconds can be reduced further to check the accuracy. The system can be developed based on the proposed model for music recommendation based on emotion.

VI. REFERENCE

[1] Barthet, M., Fazekas, G., & Sandler, M. (2012, June). Music emotion recognition: From content-to context-

based models. In International symposium on computer music modeling and retrieval (pp. 228-252). Springer, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-642-41248-6_13

[2] Wieczorkowska, A., Synak, P., & Raś, Z. W. (2006). Multi-label classification of emotions in music. In Intelligent Information Processing and Web Mining, 307-315. Springer, Berlin, Heidelberg. http://dx.doi.org/10.1007/3-540-33521-8_30

[3] Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 40(7), 2038-2048. http://dx.doi.org/10.1016/j.patcog.2006.12.019

[4] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Random k-labelsets for multilabel classification. IEEE Transactions on Knowledge and Data Engineering, 23(7), 1079-1089. http://dx.doi.org/10.1109/TKDE.2010.164

[5] Fürnkranz, J., Hüllermeier, E., Mencía, E. L., & Brinker, K. (2008). Multilabel classification via calibrated label ranking. Machine learning, 73(2), 133-153. http://dx.doi.org/10.1007/s10994-008-5064-8

[6] Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. the MIT Press.

[7] Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. Journal of new music research, 33(3), 217-238. http://dx.doi.org/10.1080/0929821042000317813

[8] An, Y., Sun, S., & Wang, S. (2017, May). Naive Bayes classifiers for music emotion classification based on lyrics. In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 635-638. IEEE. http://dx.doi.org/10.1109/ICIS.2017.7960070

[9] Misron, M. M., Rosli, N., Manaf, N. A., & Halim, H. A. (2014). Music emotion classification (mec): exploiting vocal and instrumental sound features. In Recent Advances on Soft Computing and Data Mining, 539-549. Springer, Cham. http://dx.doi.org/10.1007/978-3-319-07692-8_51

[10] Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In Ismir (Vol. 270) 1-11.

[11] Schmidt, E. M., Turnbull, D., & Kim, Y. E. (2010). Feature selection for content-based, time-varying musical emotion regression. In Proceedings of the international conference on Multimedia information retrieval, 267-274. http://dx.doi.org/10.1145/1743384.1743431

[12] Eyben, F., Salomão, G. L., Sundberg, J., Scherer, K. R., & Schuller, B. W. (2015). Emotion in the singing voice—a deeperlook at acoustic features in the light of automatic classification. EURASIP Journal on Audio, Speech, and Music Processing, 2015(1), 19. http://dx.doi.org/10.1186/s13636-015-0057-6

[13] Tiwari, V. (2010). MFCC and its applications in speaker recognition. International journal on emerging technologies, 1(1), 19-22.

[14] Liu, X., Chen, Q., Wu, X., Liu, Y., & Liu, Y. (2017). CNN based music emotion classification. arXiv preprint arXiv:1704.05665.

[15] Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In Proceedings ISMIR (Vol. 86), 937-952.

[16] Li, T. L., Chan, A. B., & Chun, A. H. (2010). Automatic musical pattern feature extraction using convolutional neural network. Genre, 10, 1x1.

[17] Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011, June). Flexible, high performance convolutional neural networks for image classification. In Twenty-Second International Joint Conference on Artificial Intelligence.

[18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778. http://dx.doi.org/10.1109/CVPR.2016.90

[19] Juthi, J. H., Gomes, A., Bhuiyan, T., & Mahmud, I. (2020). Music Emotion Recognition with the Extraction of Audio Features Using Machine Learning Approaches. In Proceedings of ICETIT 2019, 318-329. Springer, Cham. http://dx.doi.org/10.1007/978-3-030-30577-2_27

[20] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutional recurrent neural networks for music classification. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2392-2396. IEEE. http://dx.doi.org/10.1109/ICASSP.2017.7952585

[21] Feng, L., Liu, S., & Yao, J. (2017). Music genre classification with paralleling recurrent convolutional neural network. arXiv preprint arXiv:1712.08370.

[22] Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical Signal Processing and Control, 47, 312-323. http://dx.doi.org/10.1016/j.bspc.2018.08.035

[23] Lim, W., Jang, D., & Lee, T. (2016). Speech emotion recognition using convolutional and recurrent neural networks. In 2016 Asia-Pacific signal and information processing association annual summit and conference (APSIPA), 1-4. IEEE. http://dx.doi.org/10.1109/APSIPA.2016.7820699

[24] Satt, A., Rozenberg, S., & Hoory, R. (2017). Efficient Emotion Recognition from Speech Using Deep Learning on Spectrograms. In Inter-speech, 1089-1093. http://dx.doi.org/10.21437/Interspeech.2017-200

[25] Fan, J., Thorogood, M., & Pasquier, P. (2018). Soundscape emotion recognition via deep learning. In Proceedings of the Sound and Music Computing.

[26] Zhou, J., Chen, X., & Yang, D. (2019). Multimodel Music Emotion Recognition Using Unsupervised Deep Neural Networks. In Proceedings of the 6th Conference on Sound and Music Technology (CSMT). 27-39. Springer, Singapore. http://dx.doi.org/10.1007/978-981-13-8707-4_3

[27] Sharma, S. K. and Sharma, N. K. (2019). Text Classification using LSTM based Deep Neural Network Architecture. International Journal on Emerging Technologies, 10(4): 38–42.

[28] Sharma, K. and Bhatia, M. (2020). Deep Learning in Pandemic States: Portrayal. International Journal on Emerging Technologies,11(3):462–467.

[29] Velankar, M., Kotian, R., & Kulkarni, P. (2021). Contextual Mood Analysis with Knowledge Graph Representation for Hindi Song Lyrics in Devanagari Script. arXiv preprint arXiv:2108.06947.

[30] Velankar, M., Khatavkar, V., & Kulkarni, P. (2020). Multimodal Sentiment Analysis of Nursery Rhymes for Behavior Improvement of Children. JUSST, 22(12).

[31] Velankar, M., Deshpande, A. & Kulkarni, P. (2020). 3 Application of Machine Learning in Music Analytics. In R. Das, S. Bhattacharyya & S. Nandy (Ed.), Machine Learning Applications: Emerging Trends (pp. 43-64). De Gruyter. https://doi.org/10.1515/9783110610987-005

[32] Velankar, M., & Kulkarni, P. (2018). Soft computing for music analytics. International Journal of Engineering Applied Sciences and Technology, 3(2).

[33] Liu, H., Fang, Y., & Huang, Q. (2019). Music emotion recognition using a variant of recurrent neural network. In 2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018). Atlantis Press. http://dx.doi.org/10.2991/mmssa-18.2019.4

[34] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[35] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[36] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.12