Web-Based Music Genre Classification for Timeline Song Visualization and Analysis

K.Vijay Kumar[1], [2], M Ramyakrishna[2], P Hari Krishna[2], S Sai Chandra[2], J Rajini[2]

[1,2]Department of Computer Science and Engineering

[1,2] Sree Dattha Institute of Engineering and Science, Sheriguda, Telangana

ABSTRACT:

This paper presents a web application that retrieves songs from YouTube and classies them into music genres. The tool explained in this study is based on models trained using the musical collection data from Audioset. For this purpose, we have used classiers from distinct Machine Learning paradigms: Probabilistic Graphical Models (Naive Bayes), Feed-forward and Recurrent Neural Networks and Support Vector Machines (SVMs). All these models were trained in a multi-label classication scenario. Because genres may vary along a song's timeline, we perform classication in chunks of ten seconds. This capability is enabled by Audioset, which offers 10-second samples. The visualization output presents this temporal information in real time, synced with the music video being played, presenting classication results in stacked area charts, where scores for the top-10 labels obtained per chunk are shown. We briey explain the theoretical and scientic basis of the problem and the proposed classiers. Subsequently, we show how the application works in practice, using three distinct songs as cases of study, which are then analyzed and compared with online categorizations to discuss models performance and music genre classication challenges.

Index-Terms: Music Genre Classification, Machine Learning, Naive Bayes, Neural Networks, Support Vector Machines, Multi-label Classification, Audioset, Temporal Classification, Visualization.

1.INTRODUCTION

Research in Music Information Retrieval (MIR) [1] comprises a broad range of topics including genre classication, recommendation, discovery and visualization. In short, this research line refers to knowledge discovery from music and involves its processing, study and analysis. When combined with Machine Learning techniques, we typically try to learn models able to emulate human abilities or tasks, which, if automated, can be helpful for the nal user. Computational algorithms and models have even been applied for music generation and composition.

Music genre classication (MGC) is a discipline of the music annotation domain that has recently received attention from the MIR research community, especially since the seminal study of Tzanetakis and Cook [5]. The main objective in MGC is to classify a musical piece into one or more musical genres. As simple as it sounds, the eld still presents challenges related to the lack of standardization and vague genre denitions. Public databases and ontologies do not usually agree on how each genre is dened. Moreover, human music perception, subject to opinions and personal experiences, makes this agreement even more difcult. For example, when a song includes swing rhythms, piano, trumpets and improvisation, we would probably dene it as jazz music. However, if we introduce synthesizers in the same song, should the song be classied as electronic music as well? If we only consider acoustic characteristics, the answer is probably yes. But different listeners can perceive the piece from their own perspective. Whereas some might categorize the song as jazz, others might consider it electronic music or even a combination of both.

In an effort to provide a tool that gives more insights about how each genre is perceived, we have trained several classi- cation models [6] and embedded them in a web application that allows the user to visualize how each model ``senses" music in terms of music genre, at particular moments of a song. Note that experimentation details for each model are beyond the scope of this article and can be found in [6]. These models have been built using common machine learning techniques, namely, Support Vector Machines (SVM), Naive Bayes classiers, Feed forward deep neural networks and Recurrent neural networks. Whereas Bayesian and SVM methods have historically delivered good results as generalpurpose machine learning models, the results achieved with deep learning techniques in articial perception (articial vision, speech recognition, natural language processing, among others) have delivered remarkable results, approaching human-like accuracy [7]. By comparing deep learning with more traditional machine learning techniques, we also aim to compare its performance for music genre classication.

MACHINE LEARNING FRAMEWORK

Machine Learning (ML) is an area of Computer Science that involves the application of Articial Intelligence techniques to learn from data. In our case, we perform the task of supervised classication. Taking a set of songs as input, labeled by genre, we have learned different models. The songs are characterized by specic features and the labels will guide the learning process. In this case, one song can be labeled with multiple genres, and they are classied in excerpts, as we will explain later. So, the problem that we approach in this work is the annotation of music genres present in a music clip, with the purpose of comparing the performance of different machine learning models when applied to this specic problem. To this end, we use the Audioset repository and its music genre samples to train the following set of models.

2.LITERATURE SURVEY

2.1 Music Information Retrieval

Authors: Roberto Raieli, in MultimediaInformation Retrieval, 2013

The status of AR systems is covered in the Survey of Music Information Retrieval systems, presented at the Sixth International Conference on Music Information Retrieval in 2005.27 In illustrating a summary of 'Music Information Retrieval (MIR)', a distinction is made between the content-based search systems of general 'audio data' and search systems for 'music based on the notes'. Alongside these are the 'hybrid' systems, which in the early treatment of any type of audio data were converted into a symbolic version of the notes.

With reference to music databases, content- based search has different perspectives. Search-by-humming allows users to search for pieces by humming, or strumming from memory. The traditional search-by-example, according to the type of similarity required, is useful for musicologists searching for pieces inspired by a melody. Lastly come searches orientated towards comparing whole soundtracks or their parts, proving useful in 'investigations' for copyright purposes into cases of plagiarism or quotation. AR techniques have numerous practical applications: identifying songs transmitted by broadcasters, also via a 'common receiver' connected to a treatment system; search for 'suspicious' sounds recorded by surveillance systems; and sound analysis of video and any type of application in television, radio or other media industry archive. Despite the novelty of its application, AR is making tasks faster and more efficient, and its applications are now present in a lot of commercial equipment.

The survey moves on to describe the two techniques, AR or MIR, relative to 'musical data' structured on notes and 'audio data' in general. For musical data it is still necessary to distinguish between 'monophonic and polyphonic melodies'. The most important issues in both cases are measuring differences between the compared data of the notes, which the system must be able to carry out automatically, and the construction of the data index, automatically or semi- automatically. 'Distance measure' and 'indexing' are processes closely linked to the degree of matching, set each time for the document's retrieval, and the more broad and generic it is, the more the system can easily estimate the similarity between the parameters of the notes being compared, or between a parameter and indexing terms used.

For audio data not based on systems of notes, other features need to be singled out, even by 'segmenting' sound tracks into parts representative of their structure. These automatically detectable features are those typical to each sound object, namely tempo, frequency, amplitude, timbre, tone etc. The problem is in finding a scheme capable of composing the results of a track's analysis in order to obtain a satisfactory and reliable enough model of its audio features. This is feasible, for example, by composing vectors such as audio-fingerprint, or as it is known, a 'Self-organizing Map (SOM)'. This panorama continues with quick descriptions and comparisons of the 17 most advanced AR systems, and the differing needs and characteristics of users. The authors take into account three classes of user, namely 'industrial, professional, and general consumers'. These classes, to varying degrees of research, need single sound outputs, full tracks, information about composers, musical genres and classes of sounds. Objectives can be varied: copyright protection; search for music based on tastes and styles; search for the works of a given artist; and identification of tracks, etc.

2.2 The bach doodle: Approachable music composition with machine learning at scale

Authors: Cheng-Zhi Anna Huang, Curtis Hawthorne, Adam Roberts, Monica Dinculescu, James Wexler, Leon Hong, Jacob Howcroft.

To make music composition more approachable, we designed the first AI- powered Google Doodle, the Bach Doodle, where users can create their own melody and have it harmonized by a machine learning model Coconet (Huang et al., 2017) in the style of Bach. For users to input melodies, we designed a simplified sheet-music based interface. To support an interactive experience at scale, we re-implemented Coconet in TensorFlow.js (Smilkov et al., 2019) to run in the browser and reduced its runtime from 40s to 2s by adopting dilated depth-wise separable convolutions and fusing operations. We also reduced the model download size to approximately 400KB through post-training weight quantization. We calibrated a speed test based on partial model evaluation time to determine if the harmonization request should be performed locally or sent to remote TPU servers. In three days, people spent 350 years worth of time playing with the Bach Doodle, and Coconet received more than 55 million queries. Users could choose to rate their compositions and contribute them to a public dataset, which we are releasing with this paper. We hope that the community finds this dataset useful for applications ranging from ethnomusicological studies, to music education, to improving machine learning models.

2.3 Deep learning techniques for music generation A survey

Authors: Jean-Pierre Briot, Gaëtan Hadjeres, François-David Pachet

This paper is a survey and an analysis of different ways of using deep learning (deep artificial neural networks) to generate musical content. We propose a methodology based on five dimensions for our analysis: Objective - What musical content is to be generated? Examples are: melody, polyphony, accompaniment or counterpoint.

- For what destination and for what use? To be performed by a human(s) (in the case of a musical score), or by a machine (in the case of an audio file). Representation - What are the concepts to be manipulated? Examples are: waveform, spectrogram, note, chord, meter and beat. - What format is to be used?

Examples are: MIDI, piano roll or text. - How will the representation be encoded? Examples are: scalar, one-hot or many-hot. Architecture - What type(s) of deep neural network is (are) to be used? Examples are: feedforward network, recurrent network, autoencoder or generative adversarial networks. Challenge - What are the limitations and open challenges? Examples are: variability, interactivity and creativity. Strategy - How do we model and control the process of generation? Examples are: single- step feedforward, iterative feedforward, sampling or input manipulation. For each dimension, we conduct a comparative analysis of various models and techniques and we propose some tentative multidimensional typology. This typology is bottom-up, based on the analysis of many existing deep-learning based systems for music generation selected from the relevant literature. These systems are described and are used to exemplify the various choices of objective, representation, architecture, challenge and strategy. The last section includes some discussion and some prospects.

2.4 Piano automatic computer composition by deep learning and blockchain technology

Authors: Huizi Li

To explore the automatic computer composition, investigate the copyright protection and management of digital music, and expand the application of deep learning and blockchain technologies in the generation of digital music works, piano composition was taken as a sample. First, through the elaboration of the neural network methods based on deep learning, the Recurrent Neural Network (RNN), Long- Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) networks were introduced, and the deep learning-based GRU-RNN automatic composition model was constructed. Second, the blockchain technology was analyzed and expressed, and the problems in the traditional copyright protection and management of digital music were analyzed. The three aspects, i.e., ownership, right of use, and right protection, were fully considered, and the blockchain technology was integrated into the copyright protection and management of digital music. Finally, the manual analysis evaluation and pause analysis were selected as the indicators to analyze and characterize the music composition quality of the GRU-RNN model, as well as analyzing the development of the digital music market integrated with blockchain technology. The results show that the GRU-RNN model shows satisfactory effects in manual analysis evaluation or in the pause analysis of the passage. The deep learning method has great potential for application in automatic computer composition of digital music; the integration of blockchain technology has played a promotive role in the expansion and popularization of the digital music market. However, in the meantime, it still faces some technical and policy challenges. The results have a positive effect on promoting the development and application of deep learning methods and blockchain technology in digital music.

2.5 Musical genre classification of audio signals

Authors: G. Tzanetakis and P. Cook

Musical genres are categorical labels created by humans to characterize pieces of music. A musical genre is characterized by the common characteristics shared by its members. These characteristics typically are related to the instrumentation, rhythmic structure, and harmonic content of the music. Genre hierarchies are commonly used to structure the large collections of music available on the Web. Currently musical genre annotation is performed manually. Automatic musical genre classification can assist or replace the human user in this process and would be a valuable addition to music information retrieval systems. In addition, automatic musical genre classification provides a framework for developing and evaluating features for any type of content-based analysis of musical signals. In this paper, the automatic classification of audio signals into an hierarchy of musical genres is explored. More specifically, three feature sets for representing timbral texture, rhythmic content and pitch content are proposed. The performance and relative importance of the proposed features is investigated by training statistical pattern recognition classifiers using real-world audio collections. Both whole file and real-time frame-based classification schemes are described. Using the proposed feature sets, classification of 61% for ten musical genres is achieved. This result is comparable to results reported for human musical genre classification.

3.EXISTING SYSTEM

we have used classiers from distinct Machine Learning paradigms: Probabilistic Graphical Models (Naive Bayes), Feed-forward and Recurrent Neural Networks and Support Vector Machines (SVMs). All these models were trained in a multi-label classication scenario. Because genres may vary along a song's timeline, we perform classication in chunks of ten seconds. This capability is enabled by Audioset, which offers 10-second samples. The visualization output presents this temporal information in real time, synced with the music video being played, presenting classication results in stacked area charts, where scores for the top-10 labels obtained per chunk are shown. We briey explain the theoretical and scientic basis of the problem and the proposed classiers. Subsequently, we show how the application works in practice, using three distinct songs as cases of study, which are then analyzed and compared with online categorizations to discuss models performance and music genre classication challenges.

4.PROPOSED SYSTEM

In this paper author is using various machine learning algorithms such as Linear SVM and Ensemble Decision Tree and have also experiment with deep learning algorithms such as Feed Forward Neural Networks and LSTM (long short term memory) to classify music genre (type of music like HIP HOP, JAZZ, Disco or etc. In all algorithms LSTM is giving better accuracy. To implement this project author has used YouTube dataset called AUDIODATASET and we are also using same dataset to implement this project.

5. IMPLEMENTATION:

MODULES:

1) User Login:

Using this module user can login to application and after login can train with SVM, LSTM and then classify music genre

2) New User Signup Here:

Using this module user can signup with the application and then can login

3) Train SVM:

Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with SVM and then will calculate accuracy, average precision, AUC and recall with confusion matrix graph. Here extracted features dataset will be split into train and test where 80% data used for training and 20% for testing

4) Train Decision Tree:

Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with Decision Tree and then will calculate accuracy, average precision, AUC and recall with confusion matrix graph. Here extracted features dataset will be split into train and test where 80% data used for training and 20% for testing.

5) Train LSTM:

Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with LSTM and then will calculate accuracy, average precision, AUC and recall with confusion matrix graph. Here extracted features dataset will be split into train and test where 80% data used for training and 20% for testing.

6)Train Feed Forward Network:

Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with Feed Forward Neural Network and then will calculate accuracy, average precision, AUC and recall with confusion matrix graph. Here extracted features dataset will be split into train and test where 80% data used for training and 20% for testing.

7) Music Genre Classification:

Using this module user can upload test audio files from 'testMusicFiles' folder and then LSTM will predict/classify type of that uploaded music Genre

6.SYSTEM ARCHITECTURE DIAGRAM

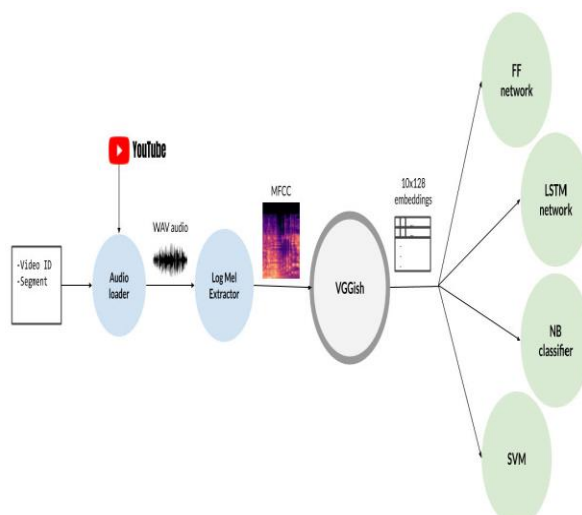

Figure 1.System Architecture Diagram

7.SCREENSHOTS

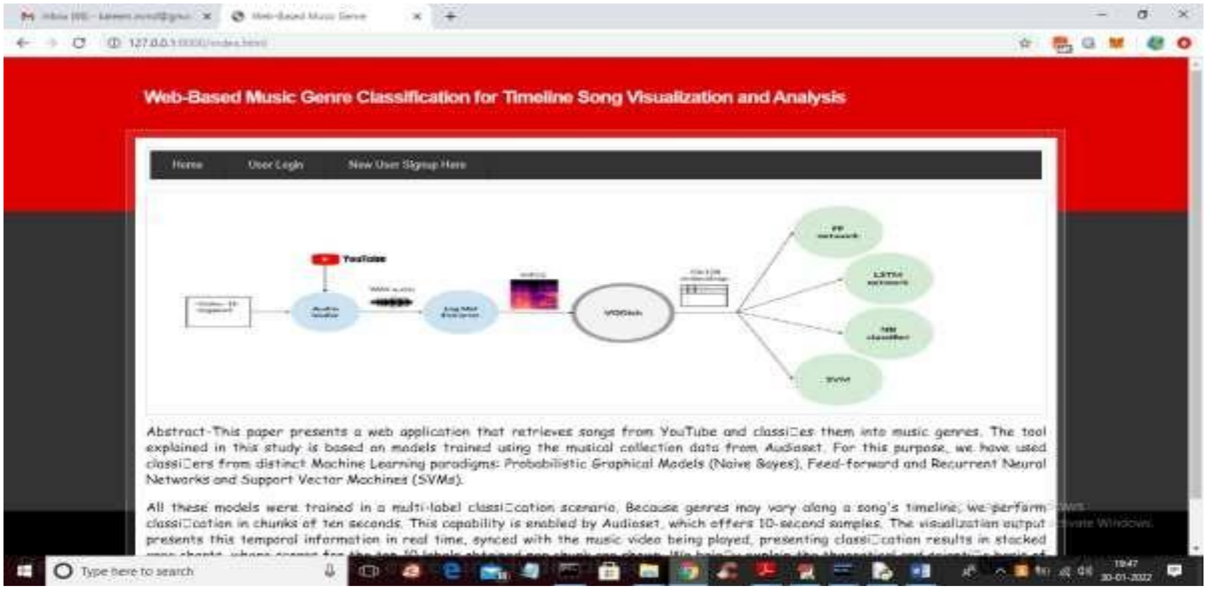

Figure.2 Home Screen

In above screen click on 'New User Signup Here' link to get below screen

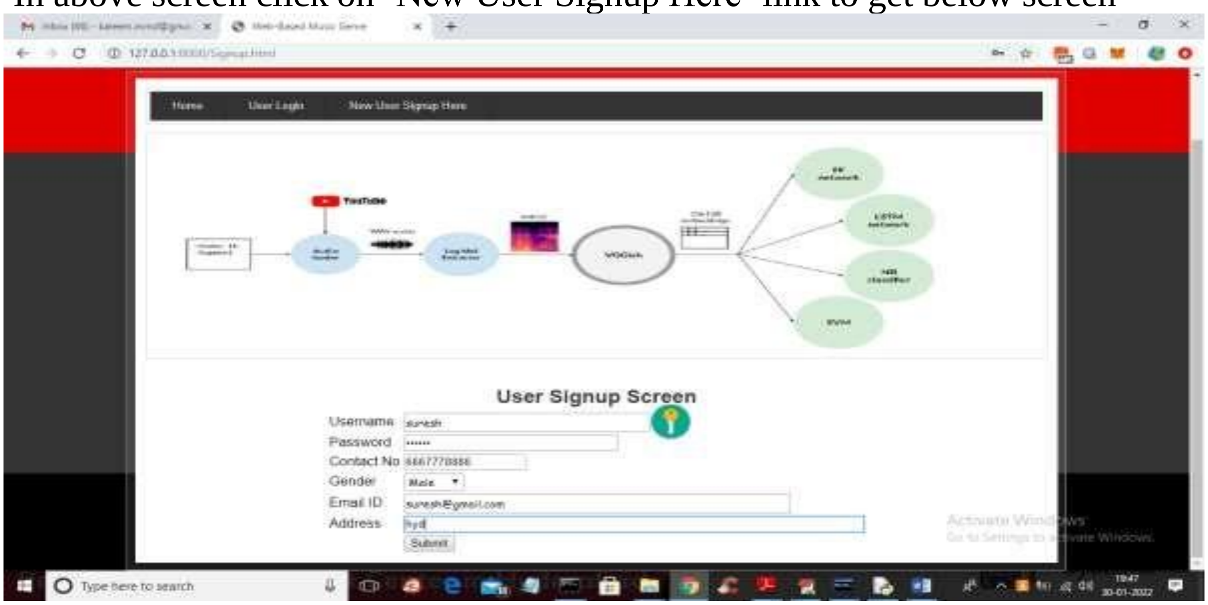

Figure.3 User Signup Screen

In above screen user is entering signup details and then click on ‘Submit’ button to get below screen

Figure.4 User Signup Screen

In above screen signup task completed and now click on ‘User Login’ link to get below login screen

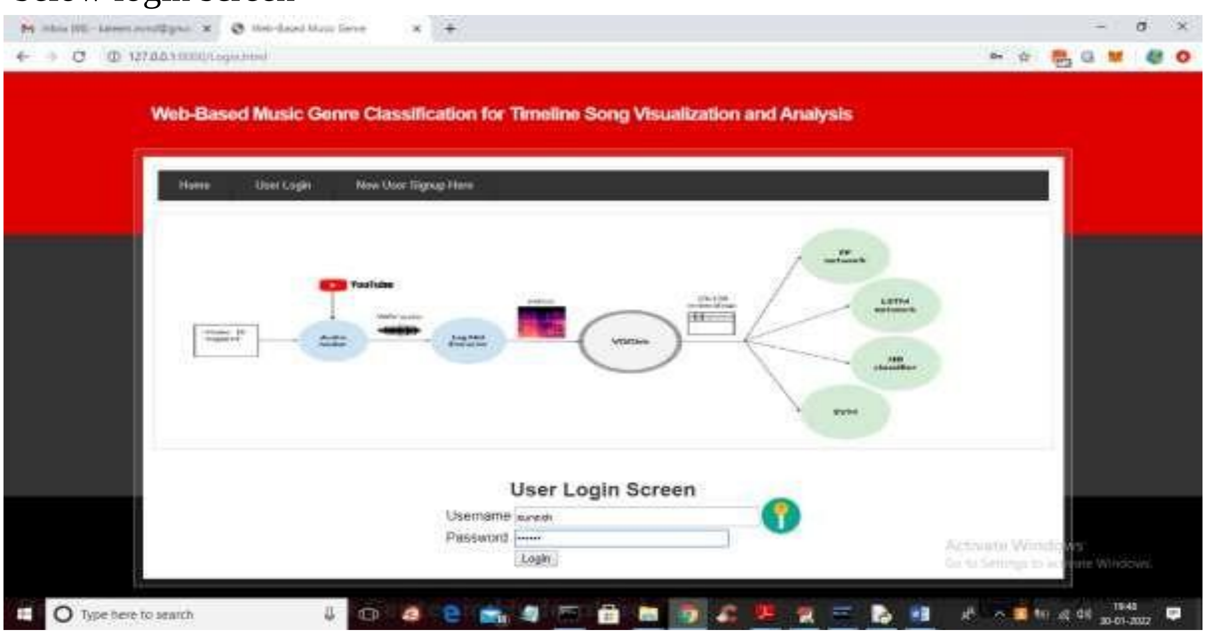

Figure.5 User login Screen

In above screen user is login and after login will get below screen

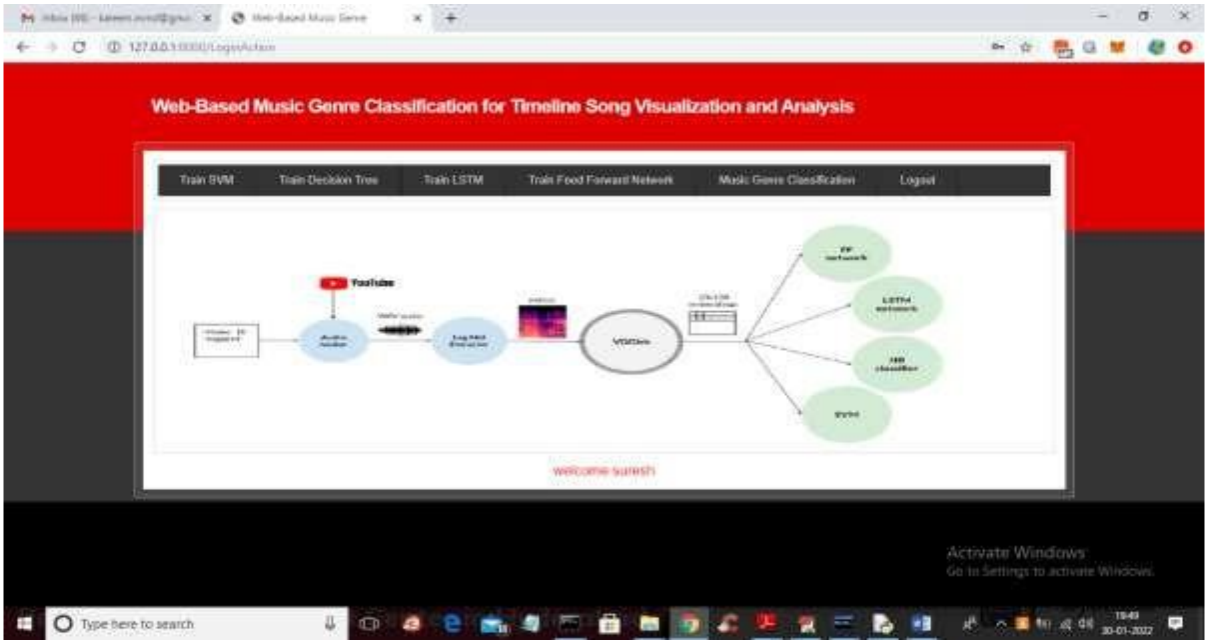

Figure.6 User Home Screen

In above screen user can click on 'Train SVM' link to train SVM algorithm and get below classification result on test data using SVM

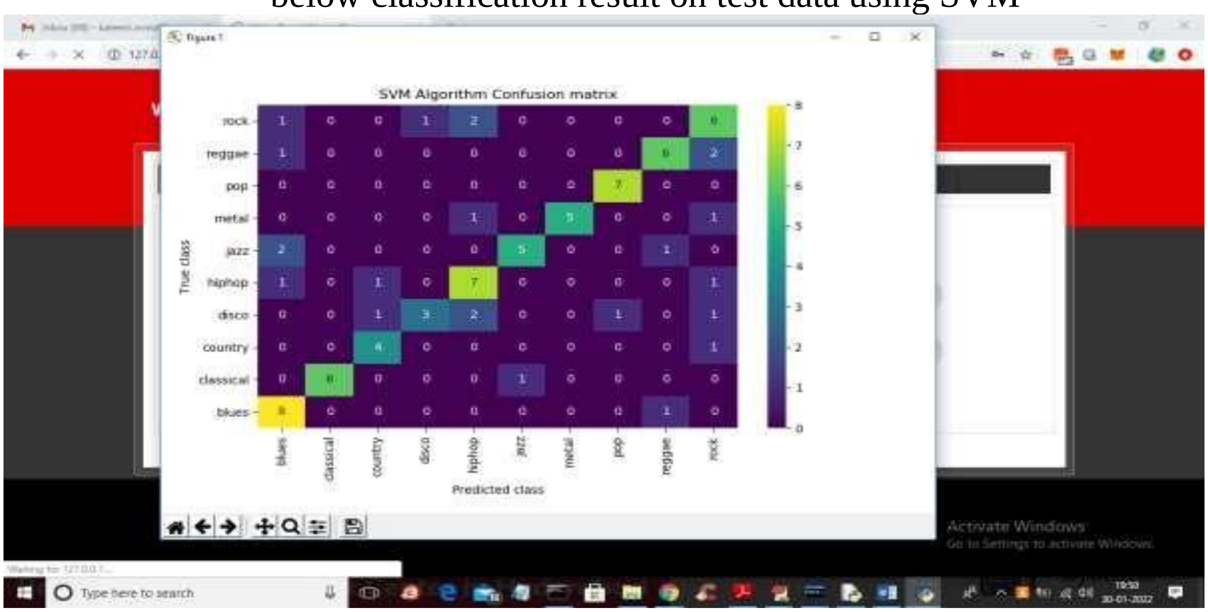

Figure.7 Train SVM Screen

In above SVM confusion matrix graph x-axis represents predicted music genre classes and y-axis represented TRUE test classes and all values in horizontal part are correct prediction by SVM remaining values greater than 0 in other boxes are the wrong prediction and we can see SVM has predicted so many wrong classes and now close above graph to get below SVM precision value

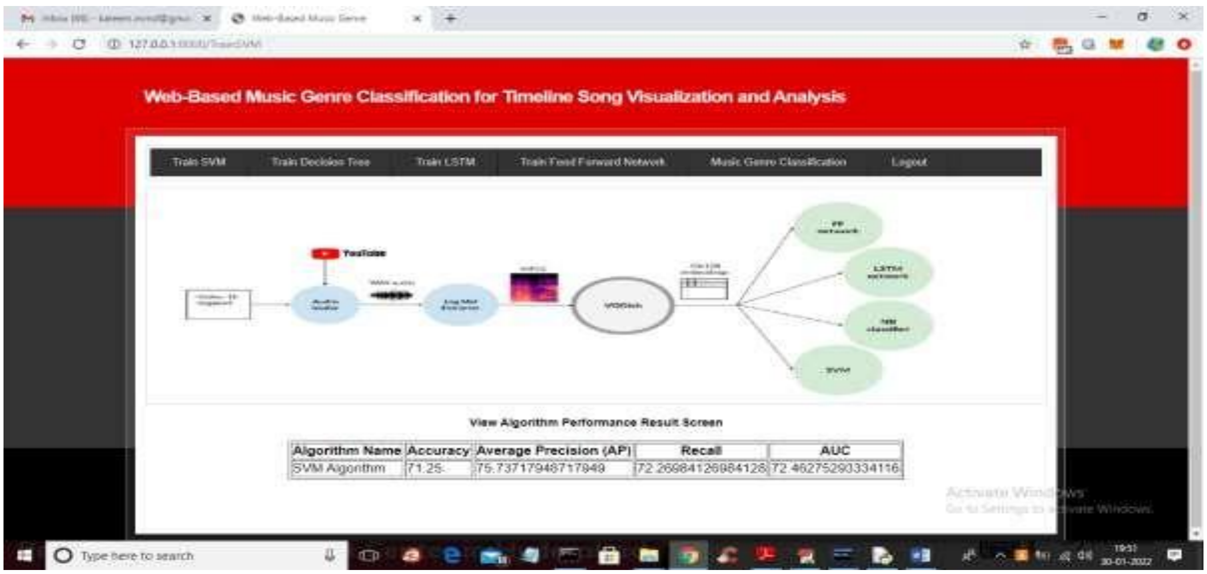

Figure.8 SVM Precision Screen

In above screen with SVM we got precision value as 75% and now click on 'Train Decision Tree' link to train decision algorithm and get below graph

Figure.9 Train Decision Tree Screen

In above screen with decision tree also so many wrong classes are predicted and now close above graph to get decision tree precision

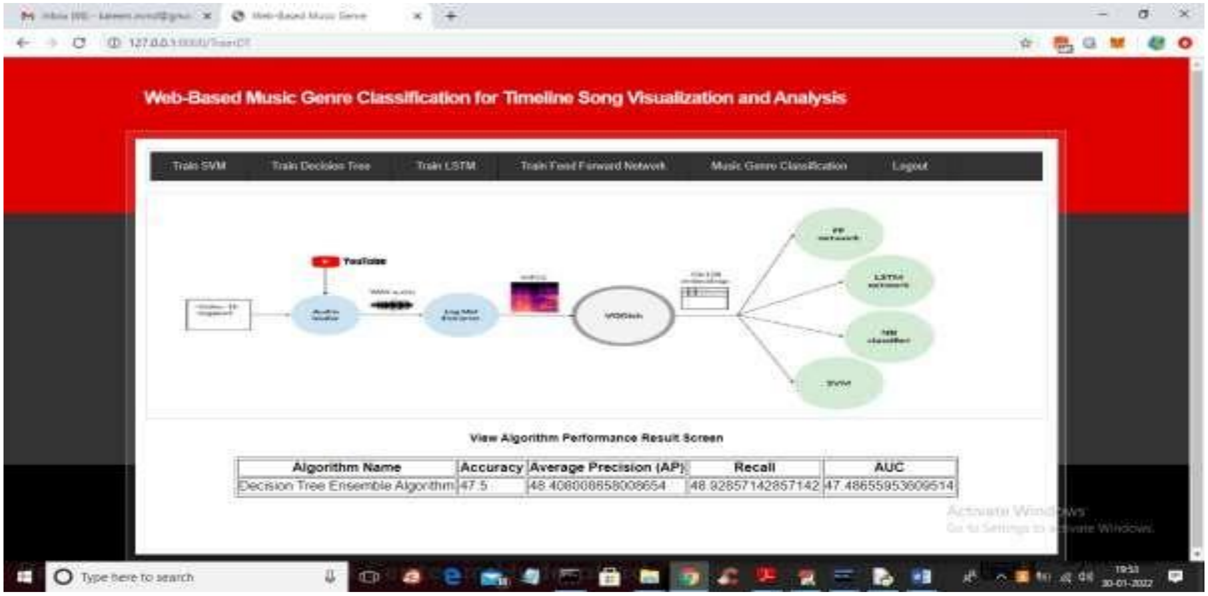

Figure.10 Train Decision Tree Screen

In above screen with decision tree algorithm we got 48% precision so its performance is not good and now click on 'Train LSTM' to train LSTM and get below output

Figure.11 LSTM Precision Screen

In above LSTM confusion matrix in diagnol boxes all classes are correctly predicted and only 1 class in other boxes is wrongly predicted so LSTM is good in performance and now close above graph to get below LSTM precision

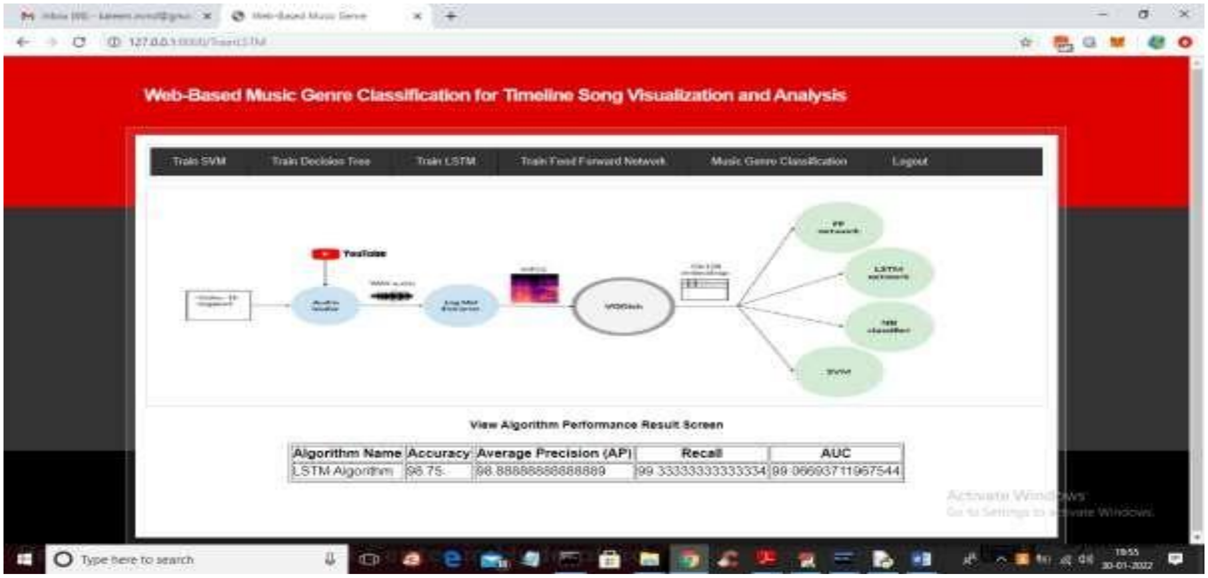

Figure.12 LSTM Precision Screen

In above screen with LSTM we got 98% precision so its performance is best compare to other algorithm and now click on 'Train Feed Forward Network' link to get below output

Figure.13 Train Feed Forward Network Screen

In above screen with feed forward neural network we can see in diagnol only few classes are correctly predicted so its performance also not good and now close above graph to get feed forward output

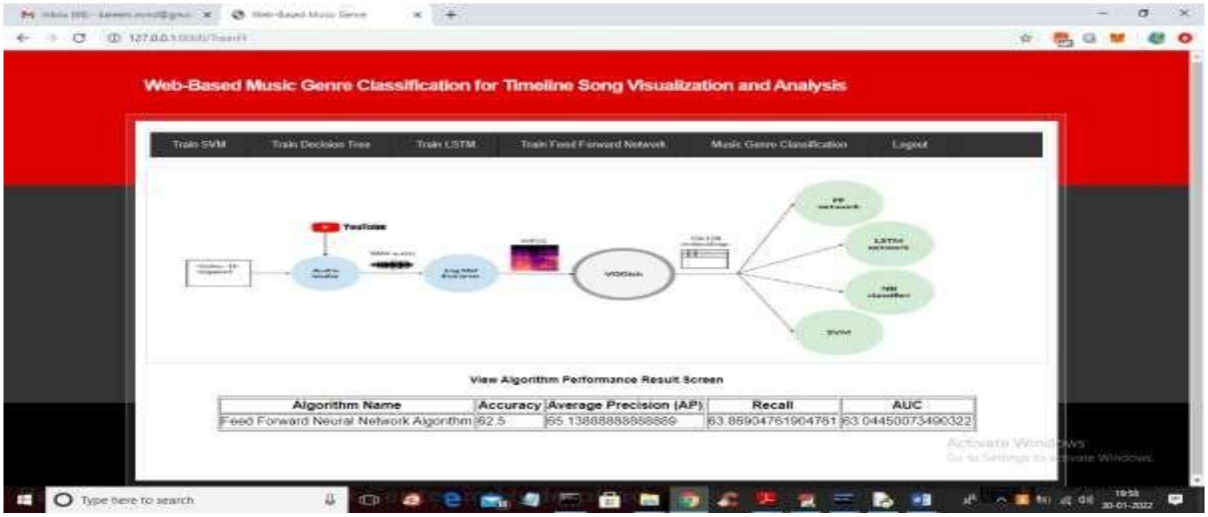

Figure.14 Feed Forward Network Precision Screen

In above screen with Feed Forward we got precision as 65% and we can see in all algorithms LSTM got better performance and in paper also author saying LSTM is better in performance and now click on 'Music Genre Classification' link to get below screen

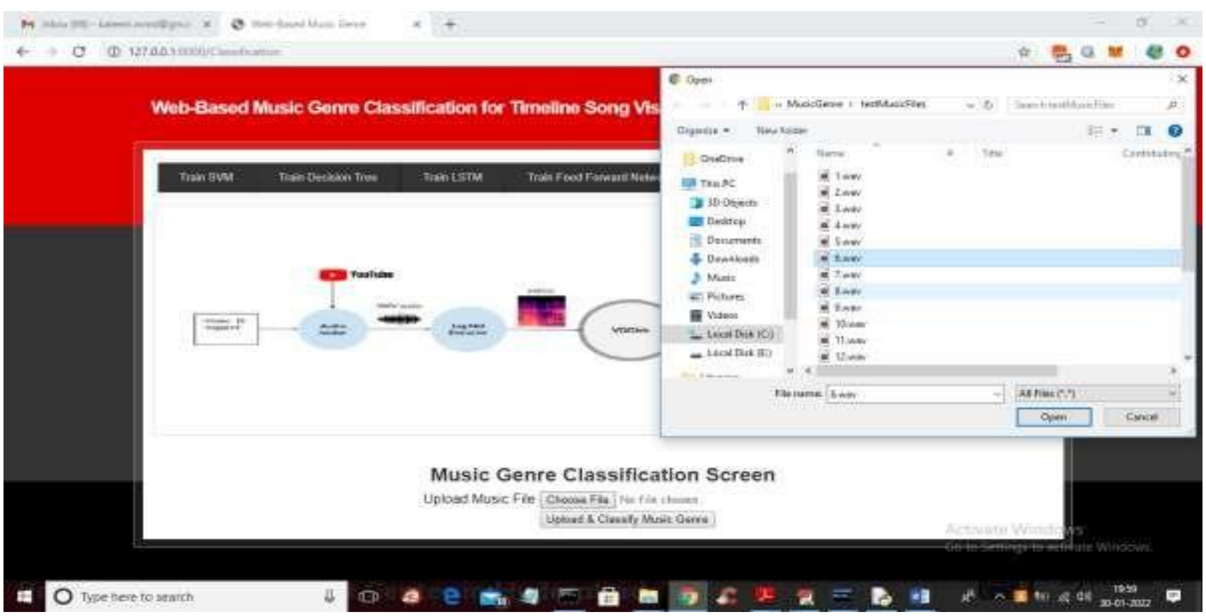

Figure.15 Music Genre Classification Screen

In above screen browsing and uploading '6.wav' file and then click on 'Open' button to load audio file and then click on 'Upload & Classify Music Genre' button so LSTM can predict or classify music Genre from uploaded audio like below screen

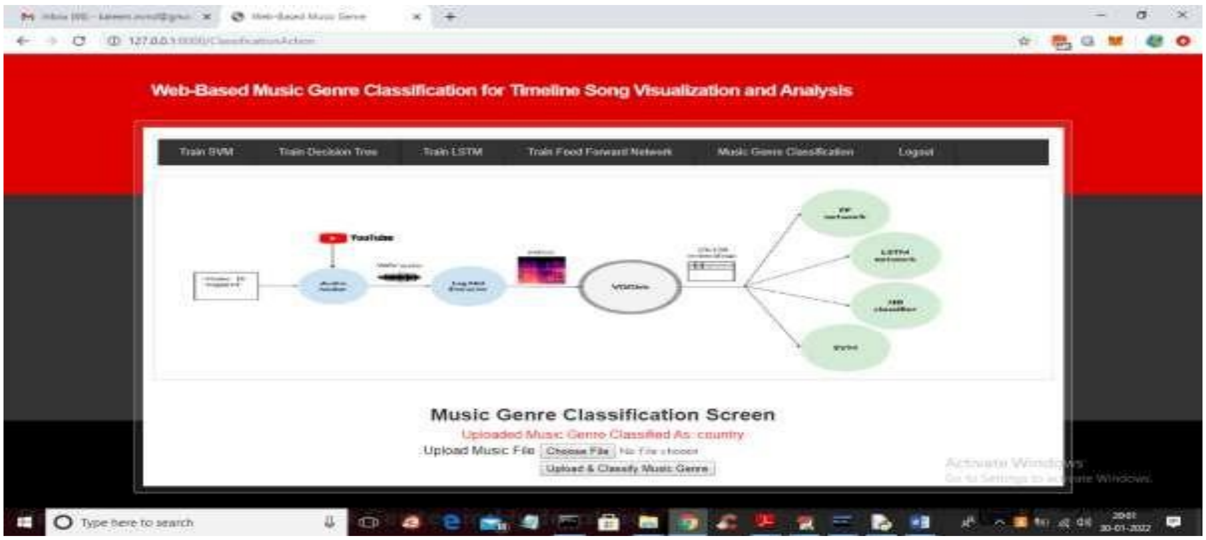

Figure.16 Upload & Classify Music Genre Screen

In above screen in red colour text we can see uploaded music genre classified as 'Country' and now test other files

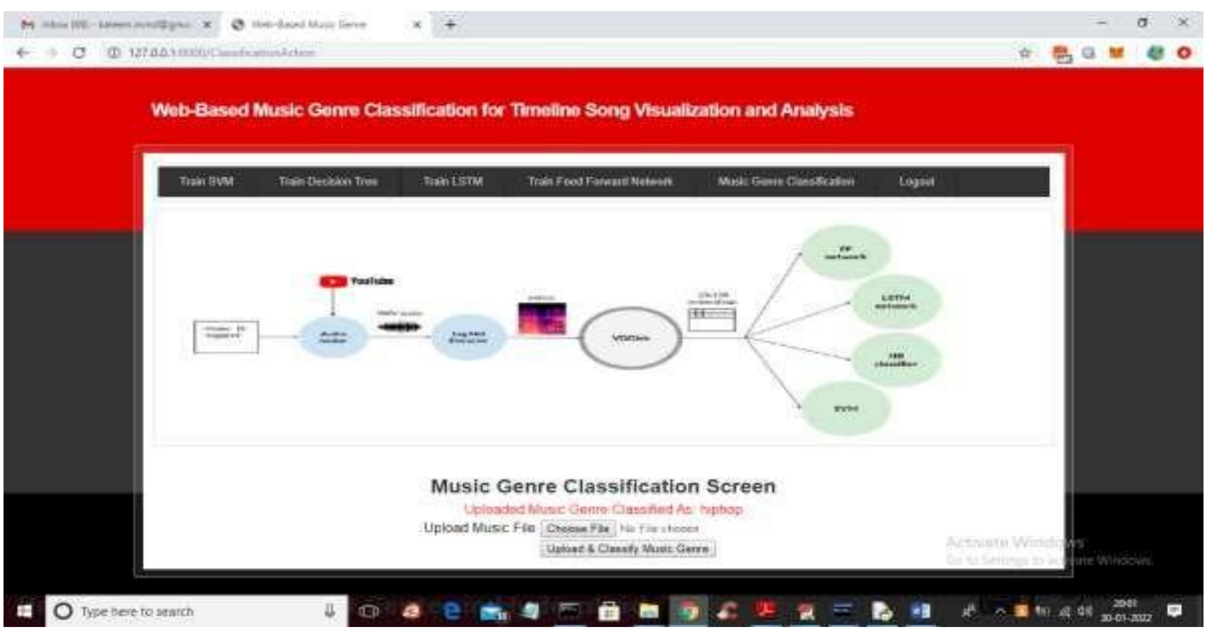

Figure.17 Upload & Classify Music Genre Screen

In above screen in red colour text we can see uploaded music genre classified as 'Country' and now test other files

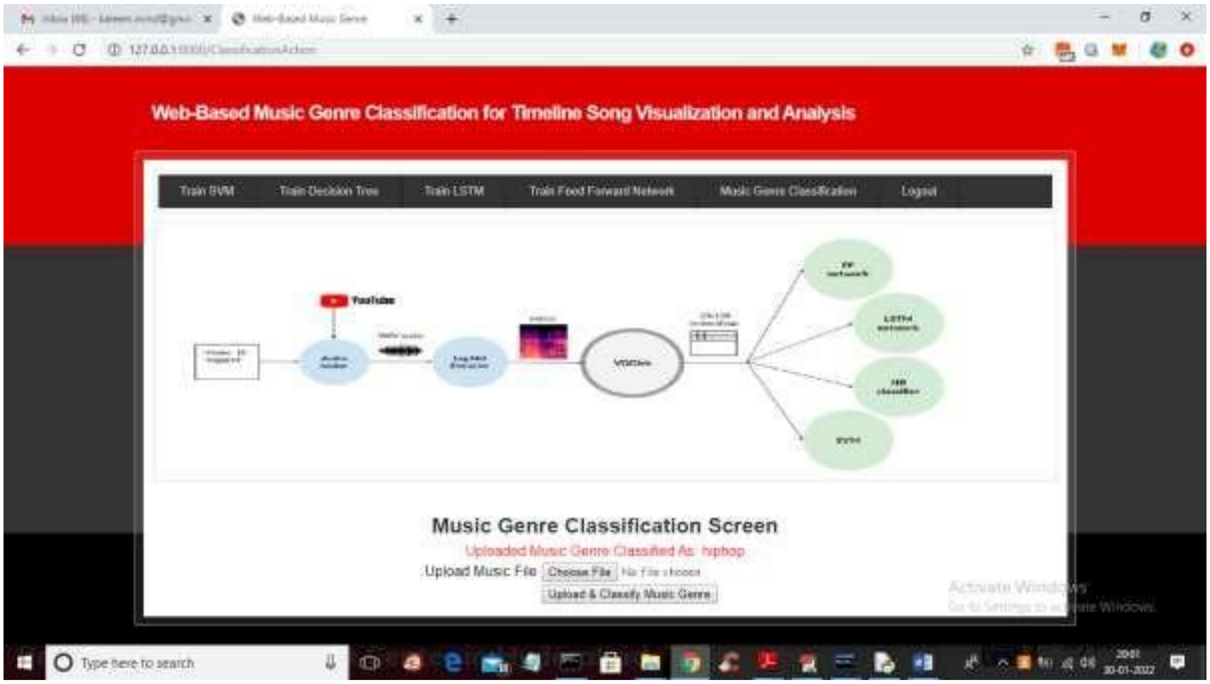

Figure.18 Music Genre Screen

In above screen another audio genre classified as 'hiphop' and similarly you can upload other files and classified them.

8.CONCLUSION

The article presents a web application to discover music genres present in a song, along its timeline, based on a previous experimentation with different machine learning models [6]. By identifying genres in each 10-second fragment, we can get an idea of how each model perceives each part of a song. Moreover, by presenting those data in a stacked area timeline graph, the application is also able to quickly show the behavior of the models, which at the same time, is an interesting way to detect undesired or rare predictions.

We believe that this application could be a supporting tool for the traditional evaluation metrics in MGC, especially when manual introspection of questionable results is required beyond classic performance metrics, such as average precision or AUC.

It is, in any case, a challenge to establish a formal way to validate genre predictions, particularly when trying to compare them with categorizations from other sources, such as online music platforms, because there is no standard or formal way of dening genres. Last.fm, to name an example, has a completely different set of tags, which, in many cases, do not correspond or exist in the Audioset ontology.

The application is also a rst step towards an eventual user-centered MGC tool, in which the users can submit feedback about the correctness of the predictions. To our knowledge, there is no visual tool that provides this level of verication on genre classication results for different fragments of the song.

The design of the precision/sensitivity metric, and its use for comparing the models' results, is an additional contribution of this paper. The incorporation of available tags from public and online services enabled the proposed evaluation method.We believe that the extension and renement of these metrics and matching algorithms is a promising future line of work and deserves attention. As mentioned throughout the paper, a consensus for a standardized taxonomy for music genre categorization is an open challenge for MGC. We plan to open a research line approaching this issue, and we feel we should incorporate semantic elements and ontology-based information to properly tackle the genre-mapping problem across different taxonomies.

9.REFERENCES

[1] J. S. Downie, ``Music information retrieval," Annu. Rev. Inf. Sci. Technol., vol. 37, no. 1, pp. 295340, 2003.

[2] C.-Z. A. Huang, C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft, ``The bach doodle: Approachable music composition with machine learning at scale," 2019, arXiv:1907.06637. [Online]. Available: http://arxiv.org/abs/1907.06637

[3] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, ``Deep learning techniques for music generationA survey," 2017, arXiv:1709.01620. [Online]. Available: http://arxiv.org/abs/1709.01620

[4] H. Li, ``Piano automatic computer composition by deep learning and blockchain technology," IEEE Access, vol. 8, pp. 188951188958, 2020.

[5] G. Tzanetakis and P. Cook, ``Musical genre classication of audio signals," IEEE Trans. Speech Audio Process., vol. 10, no. 5, pp. 293302, Jul. 2002.

[6] J. Ramírez and M. J. Flores, ``Machine learning for music genre: Multifaceted review and experimentation with audioset," J. Intell. Inf. Syst., vol. 59, pp. 469499, Nov. 2019.

[7] K. He, X. Zhang, S. Ren, and J. Sun, ``Delving deep into rectiers: Surpassing human-level performance on ImageNet classication," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 10261034.

[8] R. Basili, A. Serani, and A. Stellato, ``Classication of musical genre: A machine learning approach," in Proc. 5th ISMIR Conf., Barcelona, Spain, 2004.

[9] J.-J. Aucouturier, F. Pachet, P. Roy, and A. Beurivé, ``Signal C context= better classication," in Proc. 8th ISMIR Conf., Vienna, Austria, 2007, pp. 425430.

[10] T. D. Nielsen and F. V. Jensen, Bayesian Networks and Decision Graphs. New York, NY, USA: Springer, 2009.

[11] M. J. Flores, J. A. Gámez, and A. M. Martínez, ``Supervised classication with Bayesian networks: A review on models and applications," in Intelligent Data Analysis for Real-Life Applications: Theory and Practice. Hershey, PA, USA: IGI Global, 2012, pp. 72102.

[12] D. Temperley, ``A unied probabilistic model for polyphonic music analysis," J. New Music Res., vol. 38, no. 1, pp. 318, Mar. 2009.

[13] J. Pickens, ``A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval," in Proc. 1st ISMIR Conf., Plymouth, MA, USA, 2000, pp. 111.

[14] H.-S. Park, J.-O. Yoo, and S.-B. Cho, ``A context-aware music recommendation system using fuzzy Bayesian networks with utility theory," in Proc. Int. Conf. Fuzzy Syst. Knowl. Discovery. Berlin, Germany: Springer, 2006, pp. 970979.

[15] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, ``An efcient hybrid music recommender system using an incrementally trainable probabilistic generative model," IEEE Trans. Audio, Speech, Language Process., vol. 16, no. 2, pp. 435447, Feb. 2008.

[16] S. A. Abdallah, ``Towards music perception by redundancy reduction and unsupervised learning in probabilistic models," Ph.D. dissertation, Dept. Electron. Eng., Queen Mary Univ. London, London, U.K., 2002.

[17] C. J. C. Burges, ``A tutorial on support vector machines for pattern recognition," Data Mining Knowl. Discovery, vol. 2, no. 2, pp. 121167, 1998.

[18] T. Li, M. Ogihara, and Q. Li, ``A comparative study on content-based music genre classication," in Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2003, pp. 282289.

[19] C. N. Silla, A. L. Koerich, and C. A. A. Kaestner, ``Improving automatic music genre classication with hybrid content-based feature vectors," in Proc. ACM Symp. Appl. Comput. (SAC), 2010, pp. 17021707.

[20] R. Dechter, ``Learning while searching in constraint-satisfactionproblems," in Proc. 5th Nat. Conf. Artif. Intell. Philadelphia, PA, USA: Morgan Kaufmann, 1986, pp. 178185.

[21] G. E. Hinton, S. Osindero, and Y.-W. Teh, ``A fast learning algorithm for deep belief nets," Neural Comput., vol. 18, no. 7, pp. 15271554, Jul. 2006.

[22] L. Deng and D. Yu, ``Deep learning: Methods and applications," Found. Trends Signal Process., vol. 7, nos. 34, pp. 197387, Jun. 2014.

[23] E. J. Humphrey, J. P. Bello, and Y. LeCun, ``Feature learning and deep architectures: New directions for music informatics," J.

Intell. Inf. Syst., vol. 41, no. 3, pp. 461481, Dec. 2013.

[24] W. W. Y. Ng, W. Zeng, and T. Wang, ``Multi-level local feature coding fusion for music genre recognition," IEEE Access, vol. 8, pp. 152713152727, 2020.

[25] S. Böck, F. Krebs, and G.Widmer, ``Joint beat and downbeat tracking with recurrent neural networks," in Proc. 17th ISMIR Conf., New York City, NY, USA, 2016, pp. 255261.

[26] Y. M. G. Costa, L. S. Oliveira, and C. N. Silla, ``An evaluation of convolutional neural networks for music classication using spectrograms," Appl. Soft Comput., vol. 52, pp. 2838, Mar. 2017.

[27] R. Yang, L. Feng, H. Wang, J. Yao, and S. Luo, ``Parallel recurrent convolutional neural networks-based music genre classication method for mobile devices," IEEE Access, vol. 8, pp. 1962919637, 2020.

[28] D. H. Hubel and T. N.Wiesel, ``Receptive elds, binocular interaction and functional architecture in the cat's visual cortex," J. Physiol., vol. 160, no. 1, pp. 106154, Jan. 1962.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ``ImageNet classication with deep convolutional neural networks," in Proc. Adv. Neural Inf. Pro- cess. Syst., 2012, pp. 10971105.

[30] K. Simonyan and A. Zisserman, ``Very deep convolutional networks for large-scale image recognition," in Proc. 3rd Int. Conf. Learn. Represent., San Diego, CA, USA, 2015, pp. 114.

[31] K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770778.

[32] K. W. Cheuk, H. Anderson, K. Agres, and D. Herremans, ``NnAudio: An on-the-Fly GPU audio to spectrogram conversion toolbox using 1D convolutional neural networks," IEEE Access, vol. 8, pp. 161981162003, 2020.

[33] G. Korvel, P. Treigys, G. Tamulevicus, J. Bernataviciene, and B. Kostek, ``Analysis of 2D feature spaces for deep learning-based speech recognition," J. Audio Eng. Soc., vol. 66, no. 12, pp. 10721081, Dec. 2018.

[34] S. Gururani, C. Summers, and A. Lerch, ``Instrument activity detection in polyphonic music using deep neural networks," in Proc. 19th ISMIR Conf., Paris, France, 2018, pp. 569576.

[35] J. S. Gómez, J. Abeÿer, and E. Cano, ``Jazz solo instrument classication with convolutional neural networks, source separation, and transfer learning," in Proc. 19th ISMIR Conf., Paris, France, 2018, pp. 577584.

[36] J. Pons, O. Nieto, M. Prockup, E. M. Schmidt, A. F. Ehmann, and X. Serra, ``End-to- end learning for music audio tagging at scale," in Proc. 19th ISMIR Conf., Paris, France, 2018, pp. 637644.

[37] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, ``Empirical evaluation of gated recurrent neural networks on sequence modeling," in Proc. Deep Learn. Represent. Learn. Workshop, 2014, pp. 110.

[38] S. Hochreiter and J. Schmidhuber, ``Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 17351780, 1997.

[39] C.-C.-J. Chen and R. Miikkulainen, ``Creating melodies with evolving recurrent neural networks," in Proc. Int. Joint Conf. Neural Netw., vol. 3, Jul. 2001, pp. 22412246.

[40] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, ``Audio chord recognition with recurrent neural networks," in Proc. 14th ISMIR Conf., 2013, pp. 335340.

[41] M. Längkvist, L. Karlsson, and A. Lout, ``A review of unsupervised feature learning and deep learning for time-series modeling," Pattern Recognit. Lett., vol. 42, pp. 1124, Jun. 2014.

[42] B. L. Sturm, ``A survey of evaluation in music genre recognition," in Proc. Int. Workshop Adapt. Multimedia Retr. Copenhagen, Denmark: Springer, 2012, pp. 2966.

[43] J.-J. Aucouturier and F. Pachet, ``Representing musical genre: A state of the art," J. New Music Res., vol. 32, no. 1, pp. 8393, Mar. 2003.

[44] H. Pálmason, B. T. Jónsson, L. Amsaleg, M. Schedl, and P. Knees, ``On competitiveness of nearest-neighbor-based music classication: A methodological critique," in Proc. Int. Conf. Similarity Search Appl. Munich, Germany: Springer, 2017, pp. 275283.