Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 10
October 2023

Automated Mood Categorization of Indian Melodies Using Random Forest Approach

Neha Gupta[1], Rosey Chauhan[2*], Shikha Verma[3]
*Corresponding author: Rosey Chauhan
[1]CS Department, ABES Engineering College, Ghaziabad, UP, India
[2*]CSE Department, Sharda University, Greater Noida
[3]CSE Department, Sharda University, Greater Noida

Abstract

Most frequently, we decide to play a song or piece of music that best suits our current state of mind. Despite this significant link, the majority of the music software programmes available today do not yet offer the capability of mood-aware play-list building. By associating songs with the appropriate emotion category they communicate, music listeners might avoid spending more time manually compiling a list of tracks that suit a given mood or occasion. The challenge is to automatically and intelligently recognise this element because it is time-consuming to manually add lyrics to music with the appropriate emotion. Machine learning & data mining techniques have made significant contributions to analysing and determining how music and emotion relate to one another, which has given rise to a lot of interest in the research of mood detection in the field of music in recent years. By attempting to develop a system for automatically identifying the mood underlying the audio tracks by mining their spectral, temporal audio properties, we carry the same inspiration forward and make a contribution. Hindi songs that are popular in India are the main focus. In order to educate, prepare, and evaluate the model that captures the emotions of these audio songs, several data classification techniques were evaluated and created an open source platform for it. By using the grouping (ensemble) of random-forest-approach tested on a set of 2700 audio samples and determined the mood underlying Indian popular music with an adequate precision of 70% to 75%.

Keywords: Machine learning, Data Mining, temporal audio properties, classification techniques, random forest approach.

1. Introduction

Data mining is a relatively new, multidisciplinary area of computer science that focuses on examining and identifying patterns in huge data sets that are both intriguing and practical. The most significant work pertinent to the research is classification, one of many tasks involved in this discipline for assessing data. The generalisation of a known function or pattern among some of the available data that has already been given some special class or label is the classification job. The class of a novel and unknown data can thus be predicted using this generalised pattern. The identification and classification of musical emotions has been examined and investigated before. At first, the majority of them used a pattern recognition strategy. [1] used an SVM to classify music into six categories: joyful, robust, restless, lyrical, serious, and sad. SVMs extract features from MIDI files. Good classification accuracy was reported, however it is difficult to represent real music symbolically, as is done with MIDI files. [2] grouped emotions into six classes after dividing them into thirteen categories. They then utilised SVM to learn and recognise music emotion, and MARSYAS [3] to retrieve music features from acoustic data. A Bayesian classifier is used to categorise music recordings, and a Gaussian mixture model (GMM) is used to describe the feature collection [4] in Hierarchical mood identification system. Support vector regression and the Thayers emotion model were used in Byeong-jun Han's proposal for a music emotion detection system [5]. Features that are temporary, permanent, semantic, and compositional are the four subcategories into which [6] have divided the audio features.

Scaringella [7] divided the audio elements utilised based on timbre, melody, and intonation information, enabling genre classification into three groups, adhering to a more conventional taxonomy. Each taxonomy makes an effort to record auditory features from a particular angle. Zhouyu Fu and team [8] combine the two categories and propose a hierarchical categorization that describes audio properties from many angles and levels as opposed to a single-level taxonomy. We can categorise auditory aspects into low-level, mid-level, and top-level labels from the view of music comprehension. Two types of timbre and spatial features can be used to further categorise low-level features. Whereas temporal characteristics record timbre fluctuation and evolution across time, timbre collection the melodic quality of sound that is associated to various instrumentation.

The foundational characteristics of the audio data, such as pace, beats per minute, and so on, are called low-level features. Contrarily, mid-level elements [9] are created by combining these fundamental qualities in order to produce music-related technical understanding, such as rhythm and pitch, which are then experienced by people as genre and mood, respectively, and constitute the top level of the taxonomy. Over the course of many years, specialists and experts have studied a wide variety of audio features. Many of these features have even been standardised, such as those in the MPEG7 standards [10], which offer a list of low level audio identifiers (features) as well as methods and resources to retrieve them. To transform the digital audio data into useful features represented by numbers, the audio feature extraction procedure uses a great deal of intricate mathematics and signal processing. The primary objectives of this research are to develop an automatic system for identifying moods in popular Indian music and to create an open-source framework for analysing and testing different machine learning and data mining methods on music data.

2. Literature Review

Models representing human emotions have been developed by numerous professionals in the domains of psychology and musicology. Hevner's [11] oldest experiment helped classify diverse adjectives into 8 groups, each of which stood for a particular mood category. The model was primarily a categorical one where the same emotions were represented by a list of adjectives that were grouped together. Later, Russell [12] proposed the circumplex model, which plots mood categories on a circle, separating them from those other categories along polar axes. This model represents human feelings on a circle. An Arousal-Valence (A-V) based mood categorization model for music recommender systems was put forth by JungHyun Kim and colleagues [13]. 20 individuals' A-V values and collected music mood tags were examined, and the k-means clustering algorithm was used to divide the A-V plane into 8 regions that represented different moods. Their research demonstrates that while some locations on the A-V plane can be distinguished by representative mood tags, some mood labels are overlapped in nearly all regions. An strategy for extracting features for audio mood categorization is discussed by Akase and group [14]. In their study, they suggested extracting the rhythm and bass-line patterns for mood assessment, combining statistical feature extraction with unit pattern analysis. The efficacy of the features was tested empirically in conjunction with statistical features like MFCCs and musical grade features.

In their study, Z. Fu and colleagues [15] present a thorough analysis of audio-based classification. It provides an organised summary of the most recent developments in this area as well as the most advanced methods for classifying music. The survey has a focus on recent advancements in the approaches and outlines a number of unresolved problems for future study. The survey has included a present conversation of attributes and literature-based classification methods. The specific activities for music categorization and annotation were also examined, and both task-specific problems were found. In their work, which includes a user study on the utility of the "PANAS-X" emotion descriptions as mood labels for music, Doris Baum and company introduce EmoMusic [16]. It outlines an effort to classify and organise music according to emotions using a variety of machine learning techniques, including Naive Bayes, Self-organizing Maps, Random Forest, and Support Vector Machine classifiers.

Classification of song moods based on the lyrics and meta-data is presented [17] and also proposed a number of techniques for the supervised learning of classifiers. The LiveJournal blog site where the training data came from tags and every blog post has an aura and a music. Next SVM, Nave Bayes, and graph-based approaches, three different types of machine learning algorithms, were used to train classifiers. The findings demonstrated that the accuracy of mood categorization techniques is insufficient for use in a genuine music search engine system. There are two basic explanations: mood is a personal meta-data; and the lyric is brief and full of metaphors that only people can comprehend. For further development, the authors therefore planned to incorporate audio information with lyrics.

SVM-based multi-label classification methods for two problems—classification into the 13 adjective groups and classification into the six super-groups—were discussed by T. Li and M. Ogihara [18]. There were many borderline examples in the studies that made it challenging for the labeler to make a choice, which is why the experiments overall performed poorly. Studies reveal that emotion recognition is a challenging task, and that the most pressing need is for performance enhancement. Extending the sound data sets and gathering labelling in many rounds can also help to alleviate this problem.

Almost 72% of India's music sales are made up of Indian popular music, demonstrating the country's enormous appeal with the populace. We use this opportunity to create an auto mood recognition system for Indian popular music by analysing current classification mining techniques and creating a novel approach to automatically categorise the songs that are part of Indian popular music, according to their underlying mood, in light of the dearth of mood-based categorizers and the increasing use and acceptance of Indian popular soundtracks**.**

2.1 Melody Mood Relation

There are several broad generalisations from musical psychological studies on song mood that are relevant to MIR research, as listed below:

• Studies have demonstrated that music has the power to affect people's moods and that this impact does occur [19]. Also, listeners' natural tendency is to assign mood labels to the music they hear [20]

• Not all emotions are likely to be affected in the same way by music being played. For instance, compared to anger or contempt, feelings of melancholy, happiness, and calm are much more likely to be created by music. 12 3.2 Models of Mood (Emotion)

• Different persons do experience uniform mood affects. According to Sloboda and Juslin, listeners frequently make consistent decisions on the expression of emotion of music.

• There is unquestionably a relationship between the listener's assessment of mood and musical elements including pace, tempo, intensity, tone, mode, beats, harmony, etc. Humans can relate to a song's melody or rhythm, and they frequently sing along to the music.

2.2 Approaches to Study Emotion Model

There are two main ways used to study emotion models:

• Categorical approach: This establishes discrete types of moods that serve as the foundation for all other potential emotional variations.

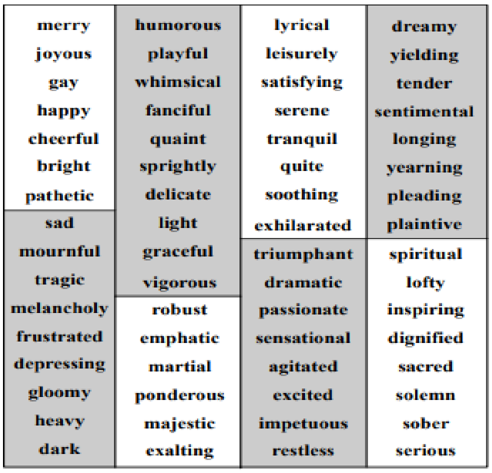
• Dimensional approach: This divides feelings into categories based on valence (pleasure), arousal (activity), potency (dominance), and other factors. This method is typically the one that music applications employ the most.

Human psychologists have put a lot of effort into studying human emotions and have put forth a number of models. Some of the important models that we will be travelling through have been extended and adopted by musicologists as well. Psychology is well aware of the six Emotions that are universal identified [21]: rage, disapproval, anxiety, joy, grief, and surprise. Nevertheless, several of them might not be appropriate for music (such as contempt) and certain frequent music moods (such as quiet or calming) are lacking because they were created for encoding facial emotions.

3. Emotion Models

3.1 Hevner's Model

Hevner explored the relationship between mood and six musical elements, including tempo, mode, rhythm, pitch, harmonisation, and melody. According to the research, 67 adjectives were divided into eight separate emotional groups that share a common sentiment. The emotional groupings are depicted in Figure 3.1 along with the appropriate adjectives for each group.



merry	humorous	lyrical	dreamy
joyous	playful	leisurely	yielding
gay	whimsical	satisfying	tender
happy	fanciful	serene	sentimental
cheerful	quaint	tranquil	longing
bright	sprightly	quite	yearning
pathetic	delicate	soothing	pleading
sad	light	exhilarated	plaintive
mournful	graceful	triumphant	spiritual
tragic	vigorous	dramatic	lofty
melancholy	robust	passionate	inspiring
frustrated	emphatic	sensational	dignified
depressing	martial	agitated	sacred
gloomy	ponderous	excited	solemn
heavy	majestic	impetuous	sober
dark	exalting	restless	serious

Fig 3.1 Hevner's Emotion Model [Source: researchgate.net]

3.2 Russell's Emotion Model

Because the mood spaces are made up of a number of distinct mood categories, both the Ekmans and Hevners models fall within the category of the categorical model. Fig 3.2 depicts Russell's Emotion Model.

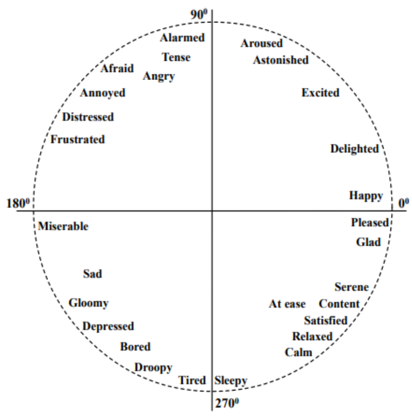


Fig 3.2 Russell's Emotion Model [Source: researchgate.net]

3.3 Thayer's model

Thayer [22] put forth another well-known dimensional model. Happiness, depression, exuberance, and anxiety are the four groups used to organise musical mood in line with the four corners of a two-dimensional area. It uses two elements to describe the melancholy: Energy dimension (composed / dynamic) and Distress dimension (joyful/apprehensive).

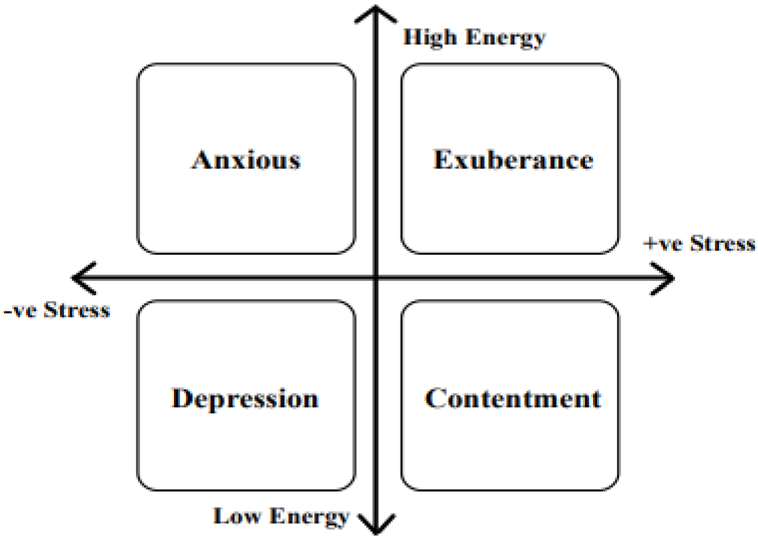


Fig 3.3 Thayer's Mood Model [Source: researchgate.net]

4. Data Collection

Personal music collections of well-known Hindi songs from India were used in the data collection. Only songs that are well-known and widely popular among the public were chosen, and effort was made to ensure that the 38 songs in the collection represent a good representation of each one of the five melancholy classes. According to the research parameters, only MP3 or WAV-format songs were chosen for the shortlist.

5. Data Pre-Processing

Three steps were taken in the production of the dataset. The first step included 490 songs, the second stage 2200 songs, and the third stage processed a total of 2300 popular audio songs from Indian Hindi films to create the dataset. The songs were all cut down to 30-second cuts. These low-level traits were gathered into an ARFF (Attribute-Relation File Format) file-dataset after being extracted. In order to establish a realistic setting for supervised training, each entry was marked with its own most likely mood based on the information gathered by consulting the group of five persons.

6. Training and Testing

The data-sets for each stage were submitted to several runs and folds of various current classification methods. The data-set underwent a 65%-35% training-testing divide training and evaluation for all the algorithms as those algorithms displaying a bias towards only certain class labels or functioning very poorly were removed. The top 11 algorithms that yielded the most comparable results all through this experiment are - Support Vector Machines, Naive Bayes, and J48 (an implementation of the ID3 method), Bagging of random trees, random forests, REPTree, simple CART (Classification and Regression Trees), simple CART, and REPTree are all examples of classification and regression trees.

7. Results

For each of the created datasets, the 11 algorithms for classification were assessed using the following four evaluation measures:-

ROC (Receiver Operating Characteristic) - The balance between the actual positivity rate and the rate of false positives is demonstrated. It is a figure with two dimensions where the genuine positive rate is represented by the vertical axis and the rate of false positives is represented by the horizontal axis. The area of a representation with perfect precision will be "1". The precision of the model is determined by the area over the receiver operator (ROC) curve. The test tuples are ranked in decreasing order, with the one most likely to fall under the positive category at the peak of the list. The model is less accurate the closer it is to the diagonal line (i.e., the area is to 0.5). The graph of sensitivity versus 1-specificity, which is an identical chart to that previously defined, was believed to be the area under the ROC curve, which was mostly employed in the theory of signal detection and the medical field. Area under ROC is determined for each of the five classes of the mood model, and the closer the value is to "1," the more accurate the classification.

Confusion Matrix: The rows of the confusion matrix reflect the actual-class, and the columns of the confusion-matrix represent the forecasts. Correct predictions are always located on the matrix's diagonal. The general structure of the confusion matrix is seen in Equation.

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

Where True-Positives (TP) represent the proportion of instances of a group that were accurately predicted, and True-Negatives (TN) represent the proportion of instances of a class that were correctly anticipated to not be a member of that class. False-Negatives (FN) represent the number of examples that were wrongly predicted to belong to another class whereas False-Positives (FP) represent the number of examples NOT belonging to a class that were incorrectly predicted to belong to that class.

Recall: This indicator indicates the proportion of actual class members that the classifier properly identified. (FN + TP) is the total number of minority members.

$$RECALL = \frac{TP}{TP + FN}$$

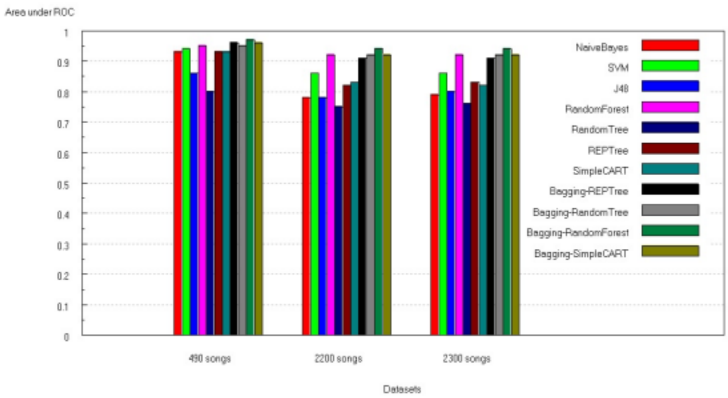
Precision: It provides us with the overall percentage of instances of each class that, according to the model or classifier, genuinely belong to that class. The sum of the classifier's optimistic predictions is denoted by (TP + FP). Equation 7.3 provides precision.

$$PRECISION = \frac{TP}{TP + FP}$$

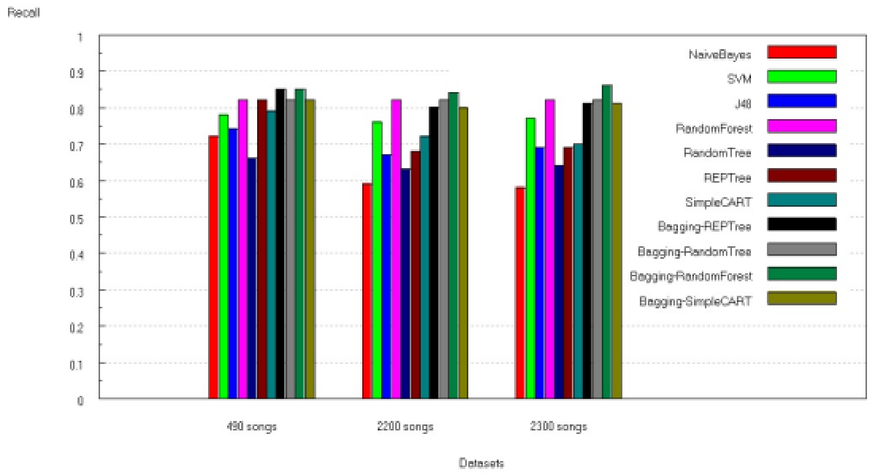
F-Measure: The F-Measure is a natural average of recall and precision. We can infer that it effectively represents the average of the two percentages. It makes contrasting the classifiers much easier. Equation 7.4 provides the solution.

$$F - Measure = \frac{2}{\left(\frac{1}{Recall}\right) + \left(\frac{1}{Precision}\right)}$$

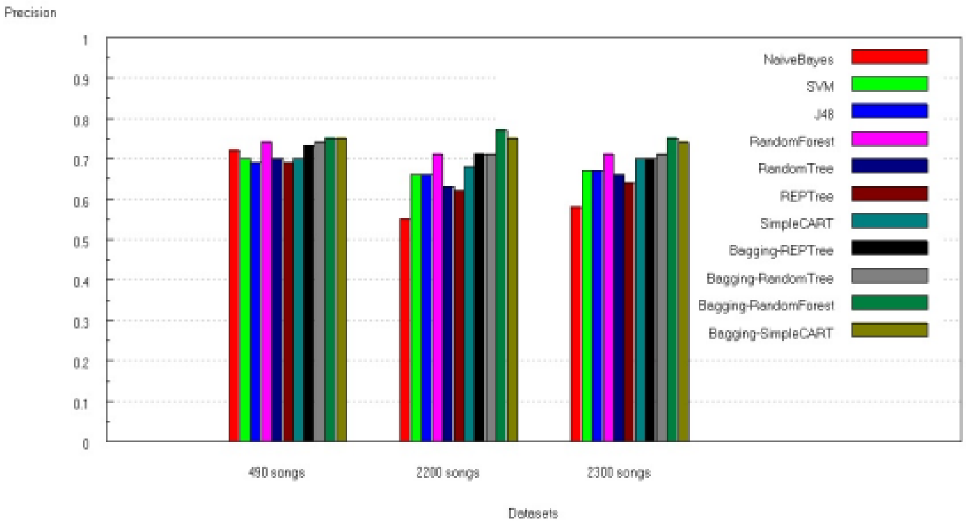
Figures 7.1, 7.2, 7.3, and 7.4 show the algorithms' performance in relation to the four metrics AUROC, Recall, Precision, and F-measure. It is clear from each set of findings that classification tree algorithms like Random-Forest, RandomTree, & SimpleCART performed better when used in an ensemble setting than when used singly, and that Bagging of Random Forest consistently outperformed all other ensembles.



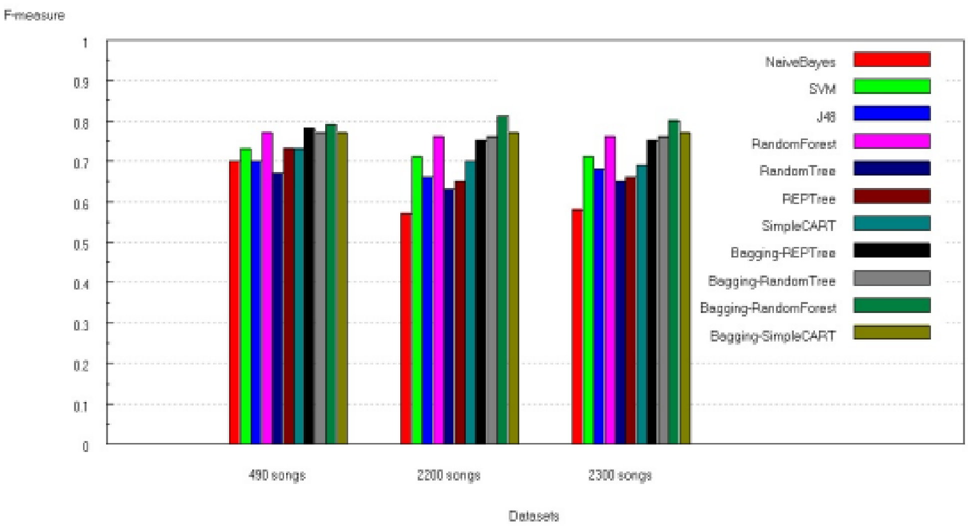
7.1 Area Under ROC statistics



7.2 Recall Statistics



7.3 Precision Statistics



7.4 F- Measure Statistics

Table 7.1 Results of Experiments using a test dataset of 2700 music clips

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Mood Class
0.946	0.117	0.759	0.979	0.845	0.991	Excited
0.805	0.021	0.914	0.805	0.856	0.978	Happy
0.77	0.006	0.971	0.77	0.859	0.967	Romantic
0.822	0.019	0.867	0.822	0.844	0.977	Sad
0.871	0.038	0.849	0.871	0.86	0.983	Silent
0.82	0.03	0.768	0.833	0.833	0.96	Avg

Table 7.2 Results of Experiments using a test dataset of 2700 music clips

Classified As	a	b	C	d	E
a = excited	704	16	1	0	9
b = happy	69	511	7	15	33
c = romantic	94	16	470	10	20
d = sad	34	5	1	314	28
e = silent	36	11	5	23	506

8. Conclusion & Future Work

It has been effectively experimented with the job of mapping auditory elements of popular Indian music with associated moods, with the highest precision measuring between 70% and 75% with respect to Fmeasure. The best accuracy in terms of ROC area was found in the range of 0.91 to 0.94, which looks fairly acceptable. Thus, the Carrying of Random Forest outscored other decision tree-based algorithms as well as other classification algorithms by a wide margin. In contrast to the analysis of western music, where SVM and neural network techniques dominated the classifier accuracy, this was a novel finding in the case of Indian popular music. As of now, the classification performance appears to be satisfactory, making it suitable for usage in practical applications. In terms of an end-to-end solution, the open source framework created as a part of the project also acts as an established structure for music data mining analysis. Although the results of the current approach have been satisfactory, we see this as only the beginning of the exploration of Indian popular music and believe that future study and advancements will lead to even more effective outcomes.

The current technology can detect the emotional impact of songs that last 30 seconds. By jointly evaluating the moods identified for the song's 30-second edited snippets, this can be further extended to determine the melancholy of the full song. With modifications to audio elements and classification algorithms, this approach can be expanded in the future to include other Indian song genres including Hindustani classical and Carnatic music. After extensive testing, it couldn't be ruled out that this technique might also be tailored to non-Indian tunes. Since lyrics analysis combined with audio elements can make the system much stronger and more accurate, this is because some of the emotions portrayed in Indian popular music are very much regulated by gestures, which are very well articulated through lyrics.

References

[1] Wang Muyuan, Zhang Naiyao, and Zhu Hancheng, "User-adaptive music emotion recognition," IEEE Xplore, Aug. 01, 2004. https://ieeexplore.ieee.org/abstract/document/1441576/

[2] T. Li and M. Ogihara, Detecting Emotion in Music. Johns Hopkins University, 2003. Accessed: Sep. 25, 2023. [Online]. Available: https://jscholarship.library.jhu.edu/items/e452670e-0342-4de1-8796-605764442469.

[3] N. Gupta, R. Chauhan, and S. Chadha, "Unmanned Aerial Vehicle (UAV) for parcel delivery," Int. J. Eng. Res. Technol., vol. 13, no. 10, pp. 2824–2830, 2020, doi: 10.37624/IJERT/13.10.2020.2824-2830.

[4] P. S. Yadav, S. Khan, Y. V. Singh, P. Garg, and R. S. Singh, "A Lightweight Deep Learning-Based Approach for Jazz Music Generation in MIDI Format," Computational Intelligence & Neuroscience, pp. 1–7, Aug. 2022, doi: https://doi.org/10.1155/2022/2140895.

[5] R. E. Thayer and R. J. McNally, "The Biopsychology of Mood and Arousal," Cognitive and Behavioral Neurology, vol. 5, no. 1, p. 65, Jan. 1992, Accessed: Sep. 25, 2023. [Online]. Available: https://journals.lww.com/cogbehavneurol/Citation/1992/01000/The_Biopsychology_of_Mood_and_Arousal.13.aspx.

[6] C. Weihs, U. Ligges, F. Mörchen, and D. Müllensiefen, "Classification in music research," Advances in Data Analysis and Classification, vol. 1, no. 3, pp. 255–291, Nov. 2007, doi: https://doi.org/10.1007/s11634-007-0016-x.

[7] Dalibor Mitrovic, Matthias Zeppelzauer, Horst Eidenberger, (2007), "Analysis of the Data Quality of Audio Descriptions of Environmental Sounds", Journal of Digital Information Management, 5(2):48.

[8] Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang, (2011), "A Survey of Audio-Based Music Classification and Annotation", IEEE Transactions on multimedia, Vol. 13, No. 2.

[9] S. Chadha, M. R. Chauhan, and M. N. Gupta, "Flood Prediction And Rainfall Analysis Using Light Gradient Boosted Machine," vol. 20, no. 6, pp. 1–6, 2022.

[10] [10] N. Gupta and K. K. Rana, "Disaster Prediction And Post Disaster Management Using Machine Learning And Bluetooth," vol. 18, no. 5, pp. 274–292, 2021.

[11] Kate Hevner, (1936), "Experimental studies of the elements of expression in music, American Journal of Psychology, 48:246268

[12] Russell J. A., (1980), "A circumplex model of affect", Journal of Personality and Social Psychology, 39: 1161-1178

[13] JungHyun Kim, Seungjae Lee, SungMin Kim, WonYoung Yoo, (2011), "Mu sic Mood Classification Model Based on Arousal-Valence Values", ICACT2011, ISBN 978-89-5519-155-4.

[14] Tsunoo, E., Akase, T., Ono, N., Sagayama S., (2010), "Musical mood classification by rhythm and bass-line unit pattern analysis", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[15] Fu, Z., Lu, G., Ting, K. M., Zhang D., (2010), "A survey of audio-based music classification and annotation", IEEE Trans. Multimedia.

[16] Chia-Chu Liu, Yi-Hsuan Yang, Ping-Hao Wu, Homer H. Chen, (2006), "De tecting and classifying emotions in popular music", JCIS Proceedings'.

[17] S. Khurana, S. Saxena, S. Jain, and A. Dixit, "Predictive modeling of engine emissions using machine learning: A review," Materials Today: Proceedings, Aug. 2020, doi: https://doi.org/10.1016/j.matpr.2020.07.204.

[18] Kumar, S., Srivastava, A. K., Soni, Y., Tyaghi, U., & Singh, N. K. (2020). Twitter sentiment analysis: A novel machine learning approach. International Journal of Control and Automation, 13(4), 412-421.

[19] S. Gupta, A. Singh, and J. Ranjan, "Sentiment Analysis: Usage of Text and Emoji for Expressing Sentiments," pp. 477–486, Jan. 2020, doi: https://doi.org/10.1007/978-981-15-0694-9_45.

[20] Juslin, P. N., & Sloboda, J. A. (Eds.). (2001). *Music and emotion: Theory and research.* Oxford University Press..

[21] "Emotion in the Human Face | ScienceDirect," www.sciencedirect.com. https://www.sciencedirect.com/book/9780080166438/emotion-in-the-human-face

[22] R. Rastogi, P. Yadav, and J. R. S. Yadav, "Statistical Analysis of Human Emotions to Suggest Suitable Music as per Individual's Mood: An Application of AI and ML for NextGen Smart Cities," International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), vol. 11, no. 3, pp. 34–67, Jul. 2021, doi: https://doi.org/10.4018/IJCBPL.2021070104.