Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

# MELODIC MOOD: EMOTION-BASED MUSIC RECOMMENDATION SYSTEM USING DEEP LEARNING TECHNIQUES

**B.Venkata Varma**, Assistant Professor, CSE(Data Science),Sreyas Institute of Engineering and Technology, India, venkatavarma@sreyas.ac.in
**M.Jayaram,** Professor, CSE(Data Science),Sreyas Institute of Engineering and Technology, India
**Billapati Meghana, Chintalapati Tharun Sai, Gandepalli Karthikeya, Macherla Anugna,** UG Scholar, CSE(Data Science),Sreyas Institute of Engineering and Technology, Hyderabad, India

***Abstract***— In recent frenzied world full of tension and anxiety, mental and emotional well-being is becoming important factors in maintaining a person's health, but with the ever accumulating work and social burden, people don't have enough time to look after of themselves, especially their mental well-being. There are extensive number of studies that demonstrate music have a paramount impact on the emotional state. So, our elucidation to is to track the persons emotions using cameras and set music in line with the mood of person and then constantly moving them to the more positive moods to try to enhance the mood of the person. The field of music recommendation systems has seen remarkable expansion, particularly with the combination of Convolutional Neural Networks (CNNs). This study prospects the application of CNNs in the domain of music recommendation, aiming to improve the reliability and manifestation of music suggestions. The proposed system leverages the fundamental ability of CNNs to capture complex hierarchical features in image data, thereby providing a more refined interpretation of musical content. We also verify the best algorithm among different algorithms like Resnet, Densenet, VGG-16 and CNN. This abstract outline the possibility of different algorithms in transforming music recommendation systems, promising a more appealing and personalized user experience in the realm of music.
***Keywords***— *Deep learning, Convolutional Neural Networks, Emotional well-being, music recommendation, Resnet, Densenet, VGG-16.*

## I. INTRODUCTION

Music has an extensively important role in enhancing an individual's life because it is a major source of delightment for both music fans and listeners. Today, with persistently escalating advances in digital media and technology, various music players have been developed with attributes such as fast forward, backward, differentiated playback speed (seek and time compression), local playback, streaming playback with multicast streams, volume adjustment, category, and so on. Although these options meet the user's basic needs, the user must dynamically browse the playlist of music and pick tracks that goes with his current mood and attitude[1].

This method excludes the additional attempt required to select music that portrays the user's emotions by forecasting the user's emotions and suggesting songs based on them. Face capture and emotion recognition provide users an advantage by foreseeing how they are feeling using deep learning. Examining a person's facial expression is the most efficacious way to determine his or her emotional state. Because each human face has unique demographics, the model can quickly analyze and distinguish any emotion, reducing calculation time. Thus, face expression is the most successful technique to anticipate emotion, allowing us to offer suitable music [2].

With the progress of digital music technology, it is requisite to develop a personalized music recommendation system that recommends songs to users. Providing recommendations based on the vast amount of data available on the internet is a remarkable undertaking. E-commerce Leviathans such as Amazon and eBay make personalized suggestions to customers based on their tastes and history, whereas Spotify and Pandora make use of Machine Learning and Deep Learning algorithms to make

pertinent recommendations. Some work has been done on personalized music recommendation to recommend songs based on the user's predilection [3].
The motive of this paper is to lift the user's spirits by playing music that meets the user's stipulation and taking the user's snapshot. Facial expression recognition has long been regarded as the best form of expression estimation known to humans. Facial grimaces are the most effective way for people to elucidate or regulate the emotion, frame of mind, or thoughts that another person is pursuing to suggest. In some situations, mood shifts can help patients overcome despondency and melancholy. Many health risks can be prevented by expression analysis, and efforts can be made to enhance the user's mood [4]. Conforming to recent studies, people reciprocate and react to music, and it has a considerable impact on brain activity. In one study of why people listen to music, scientists ascertained that music had a huge incline on the link between thrill and mood. Music has been authenticated to have two significant effects on participants: it can lift their spirits and boost their self-consciousness. It has been demonstrated that musical preferences are instantly related to emotions and personality [5].

## II. LITERATURE SURVEY

Ms. P. V. Shitole introduced Emo Player as an ingenious solution that allows users to automatically play songs based on their moods. It analyses the user's facial expressions and plays music that matches those emotions. The emotions are identified using a machine learning technology called the SVM algorithm. The outer component may be a necessary component of an individual's body, and it plays a significant role in bringing out an individual's behaviors and feelings. The camera captures an image of the user. It then extracts the user's countenance from the captured image. Facial expressions are divided into two types: smiling and not smiling. According to the impression, music square measure competes with prearranged directories [1].
According to another study, "Melomaniac-Emotion Based Music Recommendation System," the most demanding aspect of listening to music depending on our mood is selecting the accurate tune, which can be solved by using pragmatic CNN techniques that reliably distinguish users' moods. The Facial Expression Recognition system should identify issues such as face detection and positioning in chaotic photos, facial feature extraction, and expression assortment. Successive to training, the model accurately differentiates emotions as angry, happy, sad, or neutral [2]. A lot of research is being done in the field of Computer Vision and Machine Learning (ML), which trains machines to various determine human emotions or moods. Machine learning propose a number of advances to acknowledge human emotions. One approach is to use the Mobile-Net model with Keras, which results in a tiny trained model and promotes Android-ML integration [3].
The dataset applied to train the classifier is Cohn Kanade Extended (CK+). The dataset includes 593 facial action coded sequences from 123 individuals. The categorization is affianced to convey information about the subject's expression. The HELEN dataset includes approximately 200 photos that are used to train the classifier. A.txt file contains 164 landmark points for each image in the cluster, in addition to the images [4]. There are various technologies that can identify facial emotions. However, some systems suggest music. The overall purpose of the article is to develop a system that will recommend music based on the user's mood as assessed by facial emotions. In the future, emotion recognition could be operated in robotics to analyze sentiments without the need for human interference [5].
Another study by M. V. Manoj Kumar, "Expression X: Emotion Based Music Recommendation System" [6], stipulated that the user's emotional state is recognized using the Google Mobile Vision SDK. The discovered emotion state is fed into the Expression-X algorithm, which assort the music (based on the emotion value entered) and generates a playlist modified to the user's emotion state. Since emotions are forecasted based on a user's facial expression, reaching 100% accuracy is surely tough because everyone asseverates emotions differently. However, after repeated testing, it obtained a 70-75% success rate in detecting the user's correct mood state and delivering the proper set of music recommendations.

The paper [7]describes a novel technique to computerized music selection based on face emotion recognition, with the aim of escalating efficiency and user satisfaction. Unlike preceding methods, the system recognizes emotions from facial expressions in real time using CNNs. The Pygame and Tkinter packages provide ideal music recommendations based on observed emotions. Eminent advantages include less processing times and lower system costs, which increase accuracy. The FER2013 dataset verify the method's caliber to recognize emotions like happiness, sadness, and neutrality. The technique surpasses current algorithms in terms of refining efficacy by automatically designing music playlists based on the user's emotional state. This work is a significant step in employing face expression recognition for customized music experiences, promising higher user engagement and contentment.

The proposed system determines emotions using a Support Vector Machine method and a supervised learning model. The training dataset utilized here is Olivetti faces, which incorporates of 400 faces and their corresponding values or parameters. The camera captures the user's facial attributes or extracts them from a formerly captured image. The training phase commence with random values and proceeds until the projected output values blend to the model's predictions. A non-training model is accoutered for testing to assess the performance and efficiency when used in factual circumstances [8]. This study will make significant bestowal to our understanding of machine learning technologies. The EMO player classifies music by the user's emotions, such as happiness or melancholy. Our goal is to produce a player that appease the requirements of users and may be utilized in their leisure time to listen to music relevant to their current situation [9].

Using facial images and the Haar cascade algorithm, an emotion-based music recommendation system attained approximately 70% accuracy. This manifests how facial expressions can effectively predict a user's emotion and pick out suitable music. The technology provides music experiences, which is an important aspect in today's personalized world. The allusion matches songs to user’s emotions, amplifying their mood and experience [10].

Facial emotion-based music recommendation systems are an exhilarant technology that has the capability to revolutionize the music recommendation market. Facial emotion detection technology accredits personalized music recommendations based on the listener's current emotional state. While this system has demerits and limitations, more research and development can steer to more accurate and effective systems capable of providing a more personalized music listening experience [11].

An emotion-based music recommendation framework that learns a user's emotions from signals composed by wearable physiological sensors. A wearable computer device accoutered with galvanic skin response (GSR) and photo plethysmography (PPG) physiological sensors can categorize a user's emotion. This emotion data provides as supplementary data for any collaborative or content-based recommendation engine. The emotional outcome of earlier recommendations on the user are reserved in the system's database and used in future recommendations, because the consequences of the invariable musical composition can diverge among various users [12].

## III. METHODOLOGY

### *A.* Proposed System

The suggested music recommendation system is a pioneering exertion of the CNN architecture, specifically built for evaluating and elucidating auditory information contained inside music. The system's crucial function is to handle raw image input using a series of convolutional layers. This method accredits the extraction of hierarchical representations of musical patterns intrinsic in audio files. Using convolutional layers, the system can ascertain complicated features like happy, sad, surprised, fear, neutral etc which can classify different moods of an individual. The learned elements are then transformed into rich embeddings that successfully capture the intricate characteristics of the human emotions.

Furthermore, the system employs a variety of approaches, including ResNet, DenseNet, VGG-16, and CNN, to select the most efficient model for its conveyance. This is performed through a stringent review of performance metrics such as accuracy, precision, recall, and F1 scores. By juxtaposing the performance of different algorithms, the system can decide the ideal strategy for achieving its objectives. The sophisticated design of the proposed music recommendation system, as well as the algorithm selection approach, aim to provide users with highly accurate and customized recommendations.

### B. *System Architecture*

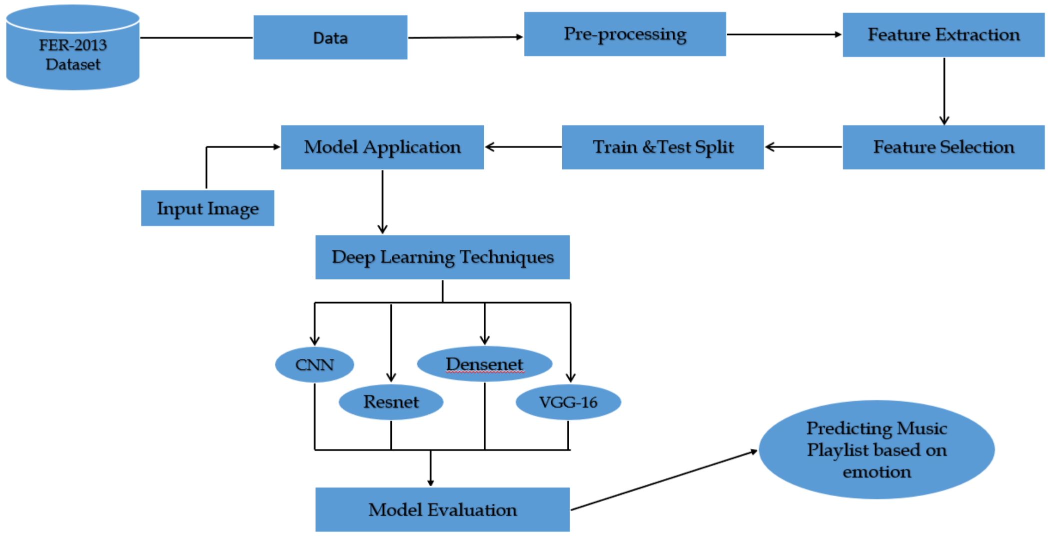


***Fig.1 System Architecture***

System Architecture gives the overall design of the model that is being generated. It has flow of steps throughout the system.

***Crafting Emotional Playlist using deep learning techniques (CNN, VGG-16, Resnet, Densenet)***

***INPUT:*** Face Captured through webcam

***OUTPUT:*** Predicts the emotion from input image and recommend song corresponding to that emotion sad, happy or neutral.

***Step-1: Load the Dataset***

Load the dataset i.e FER-2013, which consists of facial expressions in pixels (csv format) that have been labeled with corresponding emotions.

***Step-2: Pre-Processing***.

Pre-processing is performed by importing 'label encoder' from sklearn . This may involve reshaping the pixel values into image arrays and normalizing them.

***Step-3: Feature Extraction***

Extract required features from the dataset like different emotions and their corresponding pixels.

***Step-4: Feature Selection***

Select the related features from the dataset that are required by the model like only sad, happy and neutral emotions.

***Step-5: Train & Test Split***

Divide the dataset into training and testing data. Training data is 80% and test data is 20%.

***Step-6: Model Application***

Utilize the trained model to predict the emotion from user image .The model outputs the predicted emotion label

***Step-7: Input Image***

User Captures image through a webcam integration to predict emotion and recommend song corresponding to the emotion predicted.

***Step-8: Deep Learning Techniques***

This encompasses various types of deep learning models like Resnet, Densenet, VGG-16, and CNN techniques

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

***Step-9: Model Evaluation***

Analyze the performance of the emotion detection model and the overall music recommendation system.

Accuracy: Percentage of correct emotion predictions.

Precision, recall, F1-score: Metrics for imbalanced datasets.

***Step-10: Predicting Music***

Based on the predicted emotion, the system searches the music database. The database associates music tracks with emotional tags or analyses their audio features to categorize them by mood.

*1)* ***Dataset:*** The FER2013.csv dataset is a prominent resource in the field of facial emotion recognition since it furnish a large number of grayscale photos of facial expressions in disparate emotional states. Each image in the collection depicts one of seven emotions: anger, contempt, fear, happiness, sorrow, surprise, or neutral. With 35,887 images, the dataset contains a wide range of facial expressions collected in a variation of settings and lighting conditions. Its organized style, together with labeled emotion categories, facilitates the development and evaluation of machine learning models, namely convolutional neural networks (CNNs), for accurate emotion identification. Researchers and practitioners frequently use the FER2013 dataset to evaluate the effectiveness of their algorithms and evolve the area of facial expression acknowledgment.

| | emotion | pixels | Usage |
|---|---|---|---|
| 0 | 0 | 70 80 82 72 58 58 60 63 54 58 60 48 89 115 121... | Training |
| 1 | 0 | 151 150 147 155 148 133 111 140 170 174 182 15... | Training |
| 2 | 2 | 231 212 156 164 174 138 161 173 182 200 106 38... | Training |
| 3 | 4 | 24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1... | Training |
| 4 | 6 | 4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84... | Training |
| ... | ... | ... | ... |
| 35882 | 6 | 50 36 17 22 23 29 33 39 34 37 37 37 39 43 48 5... | PrivateTest |
| 35883 | 3 | 178 174 172 173 181 188 191 194 196 199 200 20... | PrivateTest |
| 35884 | 0 | 17 17 16 23 28 22 19 17 25 26 20 24 31 19 27 9... | PrivateTest |
| 35885 | 3 | 30 28 28 29 31 30 42 68 79 81 77 67 67 71 63 6... | PrivateTest |
| 35886 | 2 | 19 13 14 12 13 16 21 33 50 57 71 84 97 108 122... | PrivateTest |

35887 rows × 3 columns

***Fig.2 Sample Dataset***

### *C. Deep Learning Techniques*

Deep learning models make use of neural network topologies. A neural network, swayed by the human brain, is made up of coordinated nodes, or neurons, that form a layered structure and tag inputs and outputs. A neural network's hidden layers are neurons orchestrated between its input and output layers. "Deep" often indicate to the number of hidden layers in a neural network. Deep learning models can contain hundreds or even thousands of hidden layers. Deep learning models are trained on comprehensive amounts of labeled data and can often learn features directly from the data, eradicate the requirement for human feature extraction[13].

**1)** ***CNN (Convolutional Neural Network):***

Convolutional Neural Networks (CNNs) are a sort of deep neural network that itemize in processing visual data such as images. They revamp computer vision by providing ingenious performance in a variety of applications such as picture categorization, object recognition, and image contrasting.

CNNs are esteemed by a unique architecture that includes convolutional, pooling, and fully linked layers.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

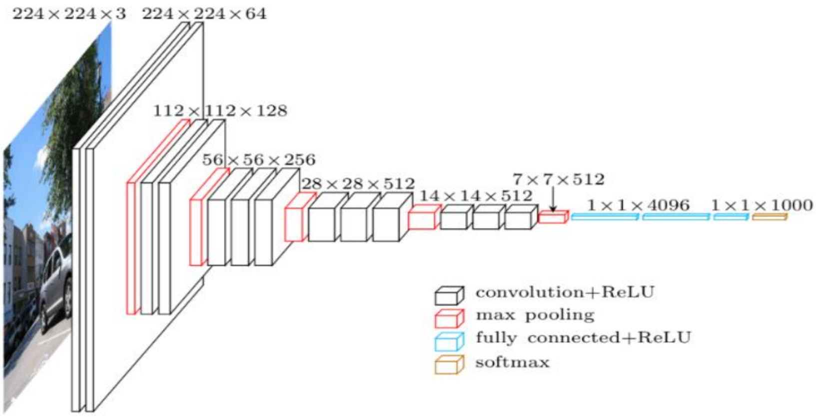


***Fig.3 CNN Architecture***

**2) *RESNET 50:***

ResNet50 is an intricated picture classification model that can be trained on large datasets. One of its most important innovations is the application of residual connections, which allow the network to learn a set of residual functions that construes input into desired output. These prevailing connections accredit the network to learn much profound structures than preliminarily possible, while circumventing the issue of vanishing gradients. ResNet-50 is incorporated by 50 layers divided into five blocks, each having a set of residual blocks. The residual blocks preserve information from previous levels, helping the network to evolve more precise representations of the input data.

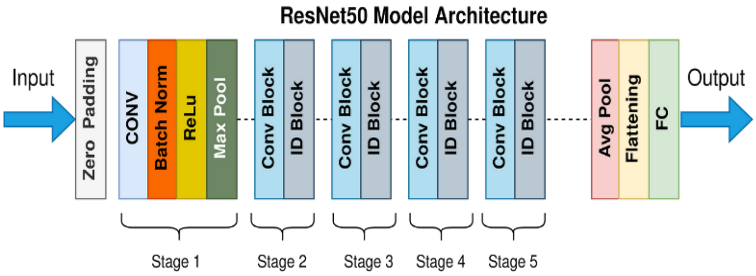


***Fig.4 Resnet Architecture***

***3) DENSENET 121:***

DenseNet-121 consists of 121 layers, each with an enunciated dense connection structure. Unlike traditional topologies, which connect one layer to the next, DenseNet provides dense connections betwixt all layers inside a thick block. This dense connectivity elevates feature restate and gradient flow across the network, leading in better feature propagation and elevates depreciation in the vanishing gradient problem.

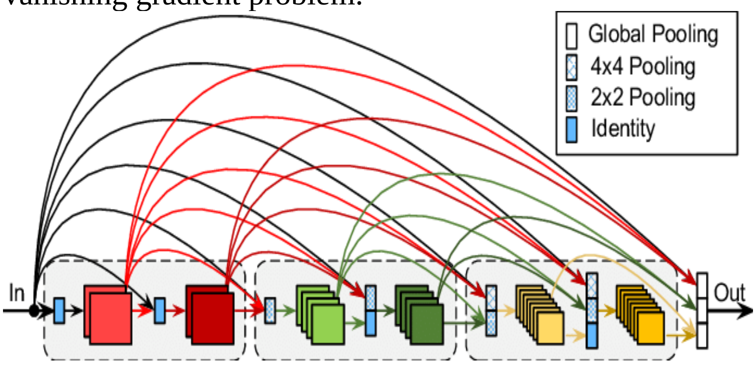


***Fig.5 Densenet Architecture***

***4) VGG-16:***

VGG16 is a deep convolutional neural network model flourished for image classification. The network is made up of 16 layers of artificial neurons, each working sequentially to elucidate image data and ameliorate prediction accuracy.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

Alternately a large number of hyperparameters, VGG16 uses convolution layers with a 3x3 filter and stride 1 that have the same padding, as well as a maxpool layer with a 2x2 filter and stride 2. This arrangement of convolution and max pool layers is abetted throughout the architecture. Finally, it contains two completely connected layers, followed by a softmax output.

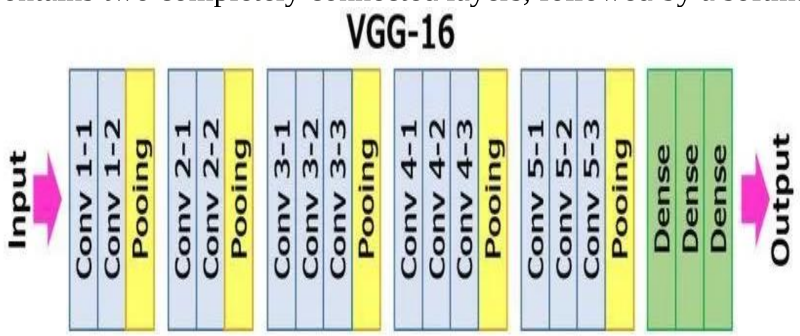


***Fig.6 VGG-16 Architecture***

## IV. RESULTS AND DISCUSSION

**Accuracy:** Accuracy is a metric that stipulates how frequent a model accurately predicts an outcome. To calculate accuracy, divide the number of correct guesses by the total number of forecasts. The accuracy can be measured on a 0–1 scale or as a percentage. The more accurate, the better. A perfect accuracy of 1.0 is accomplished when every prediction made by the model is correct.

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{All predictions}}$$

Given below is the accuracy of different models that we used in our paper.

| Notebook | Accuracy |
|---|---|
| CNN | 0.914474 |
| DENSENET | 0.967105 |
| RESNET | 0.914474 |
| VGG16 | 0.947368 |

***Fig.7 Evaluation Table***

**Classification Report:** A classification report summarizes a model's performance over many measures, comprising as accuracy, recall, and F1-score, to provide acumen into the model's behavior across classes.

Given below are the classification report of both existing and proposed system. We can clearly see the differences in their accuracy. In Fig.8 for existing system accuracy was around 69-70%. In this paper we have accuracy of 97% in Fig.9.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.60 | 0.62 | 0.61 | 991 |
| 1 | 0.68 | 0.62 | 0.65 | 109 |
| 2 | 0.58 | 0.47 | 0.52 | 1024 |
| accuracy | | | 0.69 | 7178 |
| macro avg | 0.67 | 0.66 | 0.67 | 7178 |
| weighted avg | 0.69 | 0.69 | 0.68 | 7178 |

***Fig.8 Classification Report of Existing System***

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| happy | 0.98 | 0.96 | 0.97 | 45 |
| normal | 0.97 | 0.95 | 0.96 | 59 |
| sad | 0.96 | 1.00 | 0.98 | 48 |
| accuracy | | | 0.97 | 152 |
| macro avg | 0.97 | 0.97 | 0.97 | 152 |
| weighted avg | 0.97 | 0.97 | 0.97 | 152 |

***Fig. 9 Classification Report of Proposed System***

**Confusion Matrix:** A confusion matrix reprises a machine learning model's performance on a set of test data. It is a method for staging the number of correct and incorrect instances based on the model's predictions

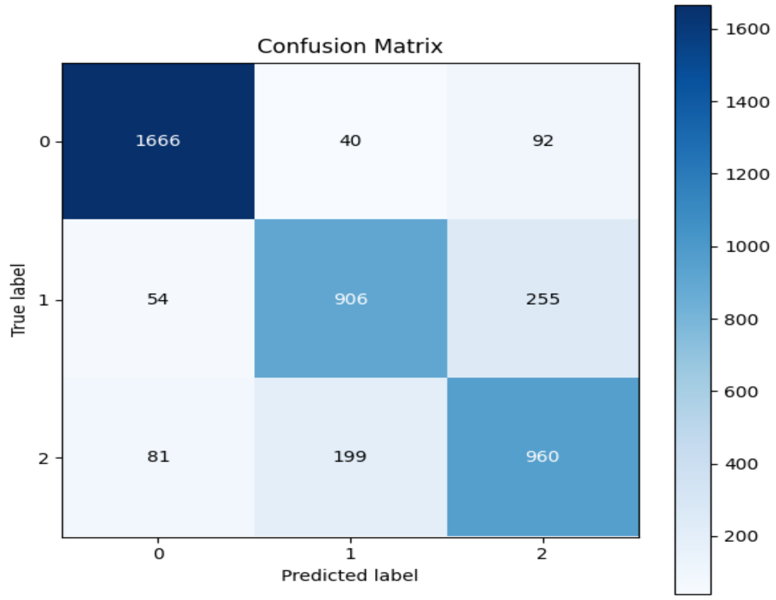


***Fig.10 Confusion Matrix***

**Visualization:** The technique of establishing the trends and correlations in our data by articulate it graphically. It can be achieved using matplotlib.

The graph shows the training and validation loss over epochs, providing insight into model performance and exploring overfitting or underfitting trends. Fig.11 shows the visualization of existing system with training accuracy 75% and validation accuracy 68%. In Fig.12 training accuracy of proposed model is 97% and validation accuracy is 94%.

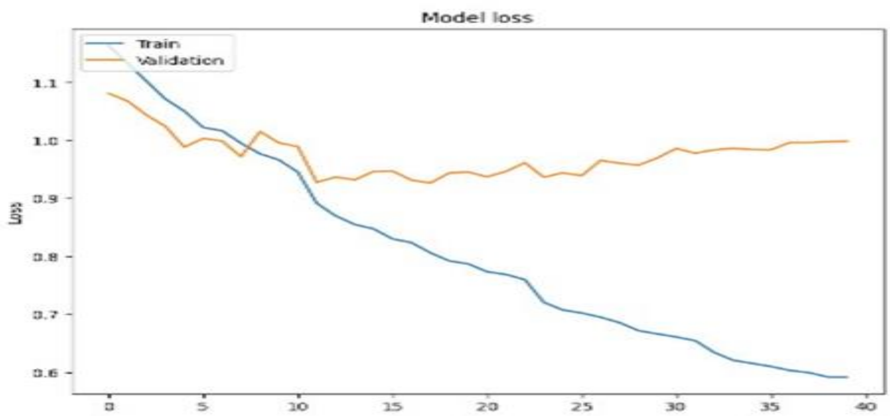


***Fig. 11 Training and validation Loss of Existing System***

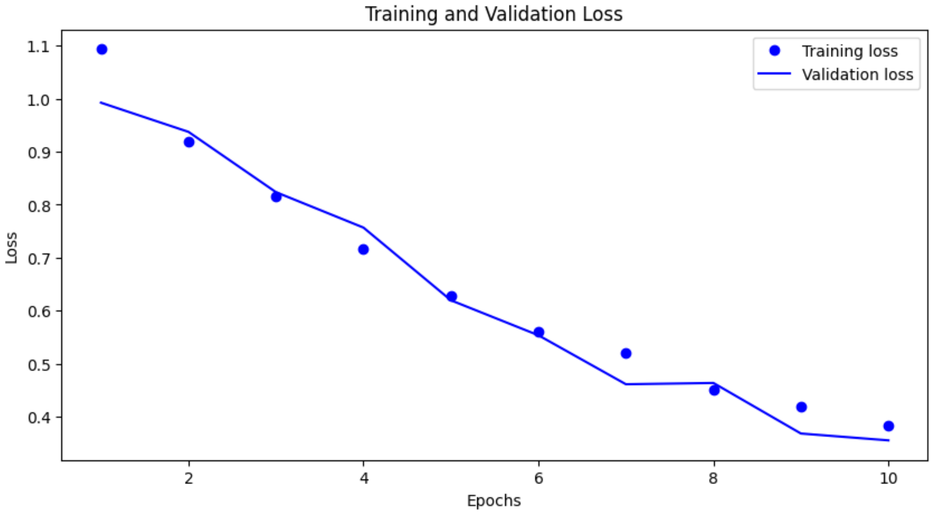


***Fig. 12 Training and validation Loss of Proposed System***

## V. CONCLUSION

Throughout this project, we've acquired vital lessons about the complicacies of human emotion and the various ways music can impact and reflect these emotions. Our efforts have been inveigled by the concept that music has an unrivaled ability to extort and intensify emotions, making it a perfect tool for self-expression, mood control, and overall wellbeing. We aimed to provide users with an instinctual and seamless experience by merging powerful machine learning algorithms and data analysis approaches, allowing them to find music that articulates to their present emotional needs. For seeking relaxation our platform endeavor to help users explore their emotional prospects via the power of music. As we end this project, we understand that our efforts are only the beginning. There is still much to learn and explore in the field of emotion-based music recommendation, such as integrating real-time physiological data, improving user feedback mechanisms, and incorporating cultural and contextual aspects.

## REFERENCES

[1] P. V Shitole, O. Kawade, P. Kamble, H. Chavan, and A. Athavale, “Emotion based Music Player Department of Computer Engineering.” [Online]. Available: www.ijsrd.com

[2] I. More, V. Shirpurkar, Y. Gautam, and N. Singh, “Melomaniac-Emotion Based Music Recommendation System.” [Online]. Available: www.ijariie.com

[3] A. Mahadik, V. Bharathi Jagan, S. Milgir, V. Kavathekar, and J. Patel, “Mood based music recommendation system,” 2021. [Online]. Available: https://www.researchgate.net/publication/352780489

[4] S. Metilda Florence and M. Uma, “Emotional Detection and Music Recommendation System based on User Facial Expression,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1757-899X/912/6/062007.

[5] D. J. Samuvel, B. Perumal, and M. Elangovan, “Music recommendation system based on facial emotion recognition,” *3C Tecnología_Glosas de innovación aplicadas a la pyme*, pp. 261–271, Mar. 2020, doi: 10.17993/3ctecno.2020.specialissue4.261-271.

[6] M. V. Manoj Kumar, B. S. Prashanth, S. Sarah, Md. Kashif, and H. R. Sneha, “Expression X: Emotion Based Music Recommendation System,” *J Comput Theor Nanosci*, vol. 17, no. 9, pp. 3958–3963, Dec. 2020, doi: 10.1166/jctn.2020.8996.

[7] M. Athavle, D. Mudale, U. Shrivastav, and M. Gupta, “Music Recommendation Based on Face Emo-tion Recognition,” *Journal of Informatics Elec-trical and Electronics Engineering*, vol. 02, no. 018, pp. 1–11, 2021, doi: 10.54060/JIEEE/002.02.018.

[8] T. J. Carolin, "EMUSIC USING SUPPORT VECTOR MACHINE LEARNING ALGORITHM," 2022, doi: 10.22159/ijet.2022v10i1.46865.

[9] T. Patanwala, N. Khan, A. Choudhary, N. Shaikh, and V. Shah, "Issue 4 www.jetir.org (ISSN-2349-5162)," 2022. [Online]. Available: www.jetir.org

[10] S. Katkuri, M. Chegoor, K. C. Sreedhar, and M. Sathyanarayana, "Emotion Based Music Recommendation System." [Online]. Available: http://www.ijert.org

[11] S. Kulkarni, S. Lingawar, S. Patil, and D. Bhadane, "FACIAL EMOTION BASED MUSIC RECOMMENDATION SYSTEM," 2023. [Online]. Available: www.ijnrd.org

[12] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion Based Music Recommendation System Using Wearable Physiological Sensors," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 196–203, May 2018, doi: 10.1109/TCE.2018.2844736.

[13] Thotakura Venkata Sai Krishna, Tummalapalli Siva Rama Krishna, Srinivas Kalime, Chinta Venkata Murali Krishna, Sadineni Neelima, Raja Rao PBV, "A novel ensemble approach for Twitter sentiment classification with ML and LSTM algorithms for real-time tweets analysis", Indonesian Journal of Electrical Engineering and Computer Science Vol. 34, No. 3, June 2024, ISSN: 2502-4752, DOI: 10.11591/ijeecs. v34.i3. pp1904-1914.

[14] Srinivas Kalime, Naresh Boddula,"A Study on Detection of Distributed Denial of Services Attacks Using Machine

Learning Techniques". International journal of research (IJR), Volume 05 Issue 12 April 2018, e-ISSN: 2348-6848 p-ISSN: 2348-795X.