# Mood Detection and Regulation using Machine Learning

**Himadri Patil[1] Makarand Kulkarni[2]**
[1]M.Tech.Scholar [2]Assistant Professor
[1,2]Department of Electronics Engineering
[1,2]K. J. Somaiya College of Engineering, Mumbai, India

*Abstract*— In today's stressful life, it is mandatory to take care of mental health. A person's day to day mood has a crucial role in mental health. This mood needs to be taken care of all the time. This paper is the state of the art work to detect mood from a person's facial expressions and to provide him / her a strategy to distract from the negative mood. We have used deep learning for facial mood detection and for generating music pieces that affect different moods. Also, voice messages have been used to distract the person and to guide a person in a negative mood. Such systems are useful in-vehicle security, classroom learning, the smart healthcare environment, etc.

***Keywords:*** Convolution Neural Network (CNN), Facial Emotion, Image Edge Detection, Long Short-Term Memory (LSTM), Mood Detection, Mood Regulation, Music Composition, Visual Geometry Group from Oxford (VGG16)

## I. INTRODUCTION

Moods or emotions play a crucial role in our day-to-day life. The mood is a state of being emotional [1]. It has a great influence on our body, perception, cognition, actions, and personality development [2]. There are multiple methods with which mood can be detected and regulated [3]. Facial expressions are the major nonverbal communication through which mood can be easily detected by a human. However, for machines, it is still a challenging task to correctly identify the mood of a person. This paper describes the two main parts of the proposed model, namely mood detection and mood regulation. In mood detection, a computer is used as an assistant tool to learn patterns of facial images captured, whereas mood regulation offers a person different musical pieces and motivational words. This model is oriented towards positive mood regulation. When the computer detects persistent negative moods like anger, fear, it activates a mechanism for interaction with the user to cheer him / her up. Recently, Deep learning is a vast field of research for pattern recognition. In deep learning, complex feature engineering is not needed. The network learns the pattern by itself. David Orozco et al. [4] have developed facial expression recognition using pre-trained models of VGG, AlexNet, and ResNet. Weights of the network were fined tuned by transfer learning. Peter Burkert et al. [5] have introduced FeatEx architecture for facial expression recognition. It consists of several convolutional layers of different sizes, as well as Max Pooling and ReLU layers. It creates a rich feature representation of the input. Antonio Fernández-Caballero et al. [6] have proposed a smart environment which detects the person's emotional state by analyzing physiological signals, facial expression, and behavior of a person. The system provides an environment to regulate these emotions towards a positive mood by changing music and lighting. Shlok Gilda et al. [7] have designed a music player that automatically generates a sentiment-aware playlist based on the emotional state of the user which uses a convolution neural network (CNN) for mood detection, audio features for music classification, and emotion-mood mapping for song recommendation. Krittrin Chankuptarat et al. [8] utilized a multimodal method to classify the emotion with the user's heart rate and facial image. According to the user's emotions and preferences, relevant songs are suggested.

Ample research is available for mood detection, however, mood regulation with the help of machines is still under research. The use of feature extraction methods with the classification of mood using a traditional machine learning algorithm is time-consuming and complex. Instead, the deep learning method which extracts features automatically is robust and less time-consuming. For mood regulation, current automatic emotion-based music recommendation systems work on a person's previous music preferences. However, preferences get changed after some time. This paper has tried to work on reducing these problems related to mood detection and regulation. In this paper, audio interaction has been chosen. Voice messages and music therapy have been used to regulate or change the mood. This paper is organized as follows: Section II describes the proposed methodology; Section III focuses on results obtained along with the discussion. Section IV concludes this paper with future scope.

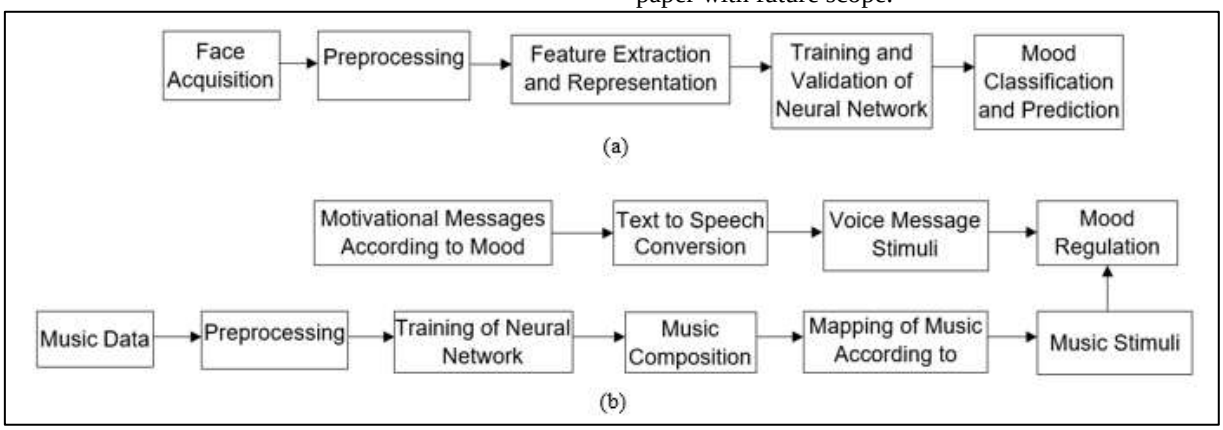


Fig. 1 (a) Flow Mood Detection (b) Flow of Mood Regulation

 

## II. METHODOLOGY

The proposed model comprises of two main parts – mood detection and mood regulation. The block diagram of mood detection and regulation has shown in Fig. 1.

### *A. Mood Detection*

Facial mood detection comprises three main steps: Face Acquisition, Feature Extraction and Representation, Mood classification and Prediction. The dataset used for training is from FER2013 [9] which has facial images of seven emotions. In this paper, 3000 images for training and 750 images for testing are used with three moods (Angry, Sad, Happy) after preprocessing. Training and testing dataset ratios have been taken as 80:20.

*1) Face Acquisition*

Faces have been acquired using the face detection method which determines the size and location of a face in digital images or video [10]. For this, Viola-Jones and Histogram of Oriented Gradients (HOG) methods are used. The Viola-Jones (haar cascade classifier) is fast in detecting faces from images and videos. However, it gives a poor performance with conditions like occlusion, non-frontal, face scaling, makeup, etc. The second method used is the HOG detector which gives a good performance in conditions like occlusion, non-frontal faces. But it is comparatively complex than viola jones and slow in detecting faces. Face detected images are as shown in Fig. 2

(a) (b) (c)

Fig. 2: (a) Original image [20] (b) Face detected and Scale Normalized image (c) Grey level equalized image

*2) Feature Extraction and Representation*

Once the facial images are detected and cropped, preprocessing is required as all the images have different sizes and contrasts. It can improve the image features to control the redundant information for adapting the feature extraction step [11]. Scale normalization with bilinear interpolation and histogram equalization method is used to make them of equal sizes and to reduce the effect of illumination, shadows, uneven distribution of light on original images resp. In deep learning models, features are extracted by hidden layers of the network, unlike traditional machine learning models. CNN is the most widely used algorithm for image classification where features are extracted automatically. To train a model from scratch is time-consuming and less accurate. In this paper, a pre-trained model of deep CNN, VGG16 (named after Visual Geometry Group from Oxford) [12] is used to solve the problem of mood classification and prediction with the help of transfer learning.

VGG16 model is 16 layers deep, 13 convolution layers, and 3 fully connected layers. VGG16 network layers are as shown in Table 1. The first five blocks of convolution layers, activation function, and pooling layer are used for feature extraction. Weights of all five convolutional blocks are frozen and features obtained from it are used to retrain the further layers. In this paper, input image dimensions have been changed to 128 × 128 pixels. The rest of the network architecture is the same as mentioned in the original paper [12]. Edge extraction is carried out to study the impact of the manual feature extraction method on the performance of the neural network. Canny edge detection and kirsh edge detection are used and their impact has been compared. Here, three types of images are applied to the network one at a time. i.e., without edge detection, with canny edge detection, with kirsh edge detection. The canny function provided by OpenCV is employed. It uses two threshold values to detect strong and weak edges. The Kirsch operator [13], a non-linear edge detector uses the convolution process in eight direction. The edge magnitude of the kirsch operator is calculated as the maximum magnitude across all eight directions. Edged detected images are as shown in Fig. 3

(a) (b)

Fig. 3: (a) Canny Edge Detection (b) Kirsh Edge Detection

| Block | Layers | | Input Image Size | Output image Size | Kernel Size | Stride |
|---|---|---|---|---|---|---|
| Input Layer | | | 128 × 128 × 3 | 128 × 128 × 3 | - | - |
| 1 | 1 | Convolution | 128 × 128 × 3 | 128 × 128 × 64 | 3 × 3 | 1 |
| | 2 | Convolution | 128 × 128 × 64 | 128 × 128 × 64 | 3 × 3 | 1 |
| | | Max Pooling | 128 × 128 × 64 | 64 × 64 × 64 | 2 × 2 | 2 |
| 2 | 3 | Convolution | 64 × 64 × 64 | 64 × 64 × 128 | 3 × 3 | 1 |
| | 4 | Convolution | 64 × 64 × 128 | 64 × 64 × 128 | 3 × 3 | 1 |
| | | Max Pooling | 64 × 64 × 128 | 32 × 32 × 128 | 2 × 2 | 2 |
| 3 | 5 | Convolution | 32 × 32 × 128 | 32 × 32 × 256 | 3 × 3 | 1 |
| | 6 | Convolution | 32 × 32 × 256 | 32 × 32 × 256 | 3 × 3 | 1 |
| | 7 | Convolution | 32 × 32 × 256 | 32 × 32 × 256 | 3 × 3 | 1 |
| | | Max Pooling | 32 × 32 × 256 | 16 × 16 × 256 | 2 × 2 | 2 |
| 4 | 8 | Convolution | 16 × 16 × 256 | 16 × 16 × 512 | 3 × 3 | 1 |
| | 9 | Convolution | 16 × 16 × 512 | 16 × 16 × 512 | 3 × 3 | 1 |
| | 10 | Convolution | 16 × 16 × 512 | 16 × 16 × 512 | 3 × 3 | 1 |
| | | Max Pooling | 16 × 16 × 512 | 8 × 8 × 512 | 2 × 2 | 2 |
| | 11 | Convolution | 8 × 8 × 512 | 8 × 8 × 512 | 3 × 3 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 | 12 | Convolution | 8 × 8 × 512 | 8 × 8 × 512 | 3 × 3 | 1 |
| | 13 | Convolution | 8 × 8 × 512 | 8 × 8 × 512 | 3 × 3 | 1 |
| | | Max Pooling | 8 × 8 × 512 | 4 × 4 × 512 | 2 × 2 | 2 |
| - | 14 | Flatten | 4 × 4 × 512 | 8192 | - | - |
| | 15 | Dense | 8192 | 1024 | - | - |
| | | Dropout | 1024 | 1024 | - | - |
| | 16 | Dense | 1024 | 3 | - | - |

Table 1: VGG16 Neural Network Layers

*3) Mood classification and Prediction*

The first 13 layers of the VGG16 model (base model) are used for feature extraction and further layers are modified to form a complete network to solve the task of mood detection. The flattening layer flattens the output of the convolutional layers to create a single long feature vector. The dense layer is used to pass all outputs from the previous layer to all its neurons, each neuron providing one output to the next layer. A regularization method called dropout is used to avoid overfitting where randomly selected neurons are ignored during training. The softmax layer is the final output layer used for mood classification.

In this paper, three moods are classified – Angry, Happy, Sad.

*B. Mood regulation*

Studies of mood management state that different types of messages affect different moods in various ways. One or more sensory modalities like vision, audition, olfaction, taste, or touch [14 - 15] can be used to induce desired emotional and behavioral effects in a person. Sound is capable to evoke emotions more effectively [16]. In this paper, audio messages and music are used to change the mood. Once the mood has been detected from facial images, an appropriate voice message or music piece gets played to change the bad mood of a person and make him / her get distracted from thoughts in the brain. The musical pieces have been created using a deep learning model.

*1) Motivational Messages*

Studies have shown that a single word is powerful to influence the expression of genes that regulate physical and emotional stress [17] Sometimes some motivational messages can change the thinking pattern which may lead to a change in the mood. In this paper, some messages are used to make a person think about those words and to provide a distraction from the root of the felt emotion for a certain period. Text messages are converted into a voice using Google's text to speech API. Some sample messages used according to detected mood are given in Table 2.

| Mood Detected | Voice Messages |
|---|---|
| Angry | You are looking angry. Please calm down. |
| | Take few deep long breaths. |
| | Anger does not solve anything. It builds Nothing. But it can destroy everything |
| Happy | You are Happy, that's Great. |
| | Just Be happy and positive always. |
| | Keep smiling and keep shining. |
| Sad | Please don't be so sad. Sit comfortably, close your eyes and take deep long breaths. |
| | Focus on the solution and not the problem. |
| | The only thing that makes you sad are your own thoughts. Change them. |

Table 2: Voice Messages according to mood

*2) Music Therapy*

Music can cause multiple feelings and emotions to fire off at the same time. The mood is an emotional state of mind, and when a powerful piece of music is heard, the mind interprets it, and codes people to feel a certain way. Different attributes of music such as genre, tempo, rhythm, pitch, mode, volume, etc. have an impact on different emotions in humans as well as animals [18]. Studies have shown that major keys and rapid tempos cause happiness, whereas minor keys and slow tempos cause sadness, and rapid tempos together with dissonance cause fear [19 - 20].

In this paper, music pieces have been created to provide music therapy if the mood detected is negative such as sad or angry. Classical music pieces are used to train the model as listening to classical music can reduce blood pressure and can help in dopamine secretion [21]. Musical Instrument Digital Interface (MIDI) files of different compositions composed by Franz Peter Schubert, Ludwig Van Beethoven, Johannes Brahms, Frederic Chopin, Franz Joseph Haydn, Johann Sebastian Bach, Wolfgang Amadeus Mozart, Sergei Rachmaninov, Robert Schumann, Dmitri Shostakovich have been used [22 - 23]. 80 Musical compositions in the major scale and 75 compositions in the minor scale have been used for training the network. A recurrent neural network (RNN) called Long Short-Term Memory (LSTM) is used to train the model to learn the continuous sequences of music. LSTM is a special kind of RNN that is capable of learning long-term dependencies. LSTM networks use memory cells which consist of three gates: input, output, and forget. Input gates control the amount of data inputted into memory, the output gate controls the data transferred to the next layer, and the forget gate controls the loss or tearing in the stored memory [24]. In this paper, the encoding of input MIDI files of music has been done via a Python toolkit named Music21. Notes and chords from the input music pieces have been extracted and represented in the form of sequence. This sequence of notes has been encoded in integers. Before applying to the LSTM network, integer values of input sequences have been reshaped and normalized. The network consists of three LSTM layers with 512 units and two fully connected layers. The dropout layer has been used for regularization. Batch Normalization layer which makes every layer of the network to learn more independently have used. The summary of the network is shown in Table 3.

| Layers | Input size | Output Size |
|---|---|---|
| Input | 200 × 1 | 200 × 1 |
| LSTM | 200 × 1 | 200 × 512 |
| LSTM | 200 × 512 | 200 × 512 |

 

| | | |
|---|---|---|
| LSTM | 200 × 512 | 512 |
| Batch Normalization | 512 | 512 |
| Dropout | 512 | 512 |
| Dense | 512 | 256 |
| ReLU | 256 | 256 |
| Batch Normalization | 256 | 256 |
| Dropout | 256 | 256 |
| Dense | 256 | 588 |
| Softmax | 588 | 588 |

Table 3: Summary of the LSTM network

For training of the network, the batch size selected is 128, the number of epochs used is 10, Optimizer used is Root Mean Square Propagation (RMS Prop). During the training of the network weights of the network are saved. These same weights are used during the decoding of music sequences. To generate the music, the decoder uses the same network architecture used for training and it has converted predicted notes into the MIDI file. A random sequence from the input has been selected as a starting point and further notes have been predicted. Two datasets are used separately for generating music in minor keys and major keys.

## III. RESULTS AND DISCUSSION

Face detection methods have been implemented using pre-trained models provided by OpenCV and Dlib. The performance of the face detection methods has been compared using 30 images with different sizes, frontal faces, non-frontal faces, and occluded facial images. Here accuracy has been calculated as the fraction of the number of faces detected and the total number of faces. Comparative analysis of Performance of face detection methods is shown in Table 4.

| Face Detection Methods | Execution Time Per Image | Complexity | Accuracy (%) | | | |
|---|---|---|---|---|---|---|
| | | | Scale | Frontal | Non frontal | Occlusion |
| Viola Jones | Approx. 0.18 Sec | Simple | 70 | 100 | 46.67 | 56.67 |
| HOG | Approx. 2.26 Sec | Moderate | 70 | 100 | 80 | 70 |

Table 4: Comparative Analysis of Performance of Face Detection Methods

Hyperparameters of models govern the entire training process. In this work, optimizers, several epochs, learning rate, and dropout are varied to compare the performance. The model has trained using three optimizers - Adam, Stochastic Gradient (SGD), and Root Mean Square Propagation (RMS Prop). The number of epochs selected is 10, 50, 100 gradually. The batch size selected is 32. The performance of the network with a dropout of 0.2, 0.5 is performed and compared. Learning rates used for training are 0.01, 0.001, 0.00001. Here, from Table 5 it can be seen that if we increase the number of epochs, the model performance increases, and a greater number of times the weights are changed in the neural network. However, at one point, the model starts overfitting. In order to reduce the problem of overfitting dropout is used which randomly drops units along with their connections from the neural network. An increase in dropout results in an increase in performance however, if we increase dropout above the threshold, the training process starts degrading. Optimizers called RMS Prop and Adam could perform with the approximately same training and validation accuracy. However, as compared to Adam, RMS Prop is more prone to overfitting. In the case of SGD, the network could not train properly with the same learning rate of 0.00001. Learning Rate is used to tell the optimizer how far to move the weights in the direction of the gradient to reduce the loss and increase accuracy. If we reduce the learning rate, RMS Prop and Adam could not train well and the network got stuck in undesirable local minima. Optimizer SGD could perform optimally with a learning rate of 0.001. Final model hyperparameters selected are: optimizer – Adam, Learning rate – 0.00001, Dropout – 0.5, epochs – 100. After training the model, total of 750 facial images were used for testing the performance of the model where 250 images were belonging to each mood class. Confusion matrix and classification report for three sets of data i.e., without edge detection, with canny edge detection, with kirsh edge detection, have shown in the Table. 5-12.

| Hyperparameters | | | Without Edge Detection | With Canny Edge Detection | With Kirsh edge Detection |
|---|---|---|---|---|---|
| For, Epochs = 50, Batch = 32, Optimizer = RMS Prop, Learning Rate = 0.00001 | | | | | |
| Dropout | 0.2 | TA | 86.77 | 73.43 | 80.87 |
| | | VA | 71.60 | 66.53 | 64.40 |
| | 0.5 | TA | 85.53 | 71.17 | 74.53 |
| | | VA | 71.73 | 59.73 | 63.20 |
| For, Dropout = 0.5, Batch = 32, Optimizer = RMS Prop, Learning Rate = 0.00001 | | | | | |
| Epochs | 10 | TA | 64.87 | 59.30 | 58.60 |
| | | VA | 65.47 | 58.93 | 57.47 |
| | 50 | TA | 85.53 | 71.17 | 74.53 |
| | | VA | 71.73 | 59.73 | 63.13 |
| | 100 | TA | 94.40 | 84.40 | 87.73 |
| | | VA | 74.13 | 57.33 | 68.00 |
| For, Dropout = 0.5, Epochs = 100, Batch = 32, Learning Rate = 0.00001 | | | | | |
| Optimizer | RMS Prop | TA | 94.40 | 84.40 | 87.73 |
| | | VA | 74.13 | 57.73 | 68.00 |
| | Adam | TA | 91.13 | 86.67 | 86.00 |

 

| | | | | | |
|---|---|---|---|---|---|
| | | V A | 73.33 | 60.93 | 65.33 |
| | SGD | T A | 42.33 | 52.03 | 44.57 |
| | | V A | 44.00 | 54.93 | 43.33 |

TA = Training Accuracy, VA = Validation accuracy

Table 5: Evaluation of the Network

| Optimizer | | Learning Rate | | |
|---|---|---|---|---|
| | | 0.01 | 0.001 | 0.00001 |
| RMS Prop | TA | 32.10 | 32.53 | 94.40 |
| | VA | 33.33 | 33.33 | 74.13 |
| Adam | TA | 31.90 | 31.90 | 91.13 |
| | VA | 33.33 | 33.33 | 73.33 |
| SGD | TA | 32.74 | 79.50 | 42.33 |
| | VA | 33.33 | 68.80 | 44.00 |

TA = Training Accuracy, VA = Validation accuracy

Table 6: Evaluation of the Network

| | | Predicted Label | | |
|---|---|---|---|---|
| | | Angry | Happy | Sad |
| True Label | Angry | 90# | 78 | 82 |
| | Happy | 90 | 79# | 81 |
| | Sad | 77 | 92 | 81# |

# - True Positive Values

Table 7: Confusion Matrix for without Edge Detection

| Mood | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| Angry | 56.4 | 35 | 36 |
| Happy | 54.53 | 31.7 | 31.6 |
| Sad | 55.73 | 33.2 | 32.4 |

Table 8: Classification Report for without Edge Detection

| | | Predicted Label | | |
|---|---|---|---|---|
| | | Angry | Happy | Sad |
| True Label | Angry | 70# | 72 | 108 |
| | Happy | 84 | 79# | 87 |
| | Sad | 68 | 79 | 103# |

# - True Positive Values

Table 9: Confusion Matrix for with Canny Edge Detection

| Mood | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| Angry | 58.4 | 32 | 28 |
| Happy | 56.31 | 34 | 32 |
| Sad | 54.4 | 35 | 41 |

Table 10: Classification Report for with Canny Edge Detection

| | | Predicted Label | | |
|---|---|---|---|---|
| | | Angry | Happy | Sad |
| True Label | Angry | 63# | 92 | 95 |
| | Happy | 57 | 77# | 116 |
| | Sad | 67 | 80 | 103# |

# - True Positive Values

Table 11: Confusion Matrix for with Kirsh Edge Detection

| Mood | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| Angry | 58.5 | 34 | 25 |
| Happy | 54 | 31 | 31 |
| Sad | 52.27 | 33 | 41 |

Table 12: Classification Report for with Kirsh Edge Detection

The accuracy of the model is high when input images to the network are processed with edge detection. However, the number of true negatives (i.e., model predicts correctly the negative class.) is high when edge detection is used as compared to true positives. Facial images with the Canny edge detection method has shown good accuracy for all the classes as compared to images with kirsh edge detection. Precision and recall are also good with canny edge detection for happy and sad moods. The overall performance of the model can be increased by tuning hyperparameters appropriately. Images with canny edge detection show significantly good performance while training the neural network and learning the features. For mood regulation, music pieces have been generated using the LSTM neural network. Some Sample of music generated is shown in Fig. 4.

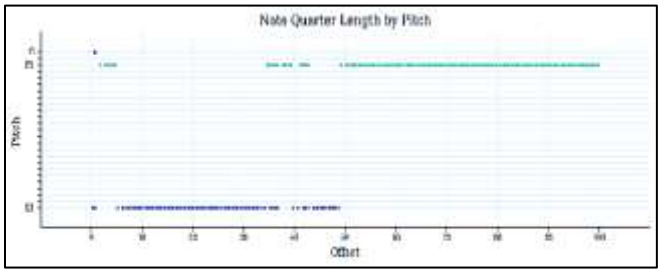


Fig. 4: a) Music Generated with Minor Scale Data

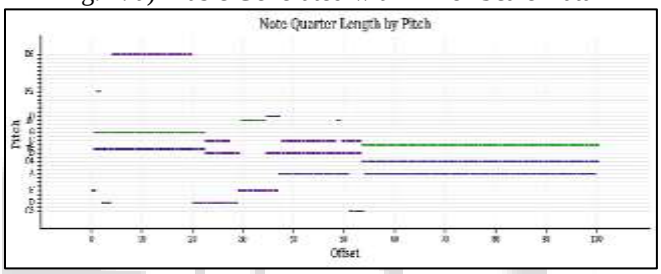


Fig. 4: b) Music Generated with Major Scale Data

## IV. CONCLUSIONS

Mood detection and regulation is a multidisciplinary field of research that includes computer science, psychology, music theory, etc. In this paper, various steps that make the mood detection complex have been explained and implemented. There are various challenges while moving towards mood detection from facial expressions namely computational complexity, lack of open-source data, the correct choice of hyperparameters. Also, mood regulation using machines is complicated yet interesting. This paper has used audio interference for mood regulation wherein voice messages and artificially generated music have used to induce or change emotions in a person. Further work can be done by using multimodalities to detect mood and therapeutic music can be generated with proper training and data.

## REFERENCES

[1] Hume, D., "Emotions & Moods", in Organizational Behavior, Pearson, pp. 258–297, 2012.
[2] Izard, C. E., "The Emotions in Life and Science" in Human Emotions, New York: Plenum Press, pp. 99–129, 1977.
[3] Patil, H., & Kulkarni, M. "A Review on Various Mood Detection and Regulation Methods", in International Journal of Engineering Research in Computer Science

 

and Engineering (IJERCSE), Vol. 7, Issue. 9, pp. 37–46, 2020.
[4] D. Orozco, C. Lee, Y. Arabadzhi, and D. Gupta, "Transfer learning for Facial Expression Recognition.", 2018
[5] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression Recognition," pp. 1–8, 2015.
[6] Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, M. T., Zangróniz, R., Latorre, J. M., & Fernández-Sotos, A. "Smart environment architecture for emotion detection and regulation", in Journal of Biomedical Informatics, pp. 55–73, 2016.
[7] Gilda, S., Zafar, H., Soni, C., & Waghurdekar, K., "Smart music player integrating facial emotion recognition and music mood recommendation". in Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET, pp. 154–158, 2017.
[8] Chankuptarat, K., Sriwatanaworachai, R., & Chotipant, S., "Emotion-based music player", in Proceeding of 5th International Conference on Engineering, Applied Sciences and Technology, ICEAST, pp. 1–4, 2019.
[9] www.kaggle.com/datasets
[10] Kumar, A., Kaur, A., & Kumar, M. "Face detection techniques: a review", in Artificial Intelligence Review, Vol. 52, No. 2, pp. 927–948, 2019.
[11] N. P. Nirmala Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis and R. Boothalingam R., "Grey Wolf optimisation-based feature selection and classification for facial emotion recognition," in IET Biometrics, Vol. 7, No. 5, pp. 490-499, 2018.
[12] Simonyan, K., & Zisserman, A. "Very deep convolutional networks for large-scale image recognition", in Proceeding of 3rd International Conference on Learning Representations, ICLR , pp. 1–14, 2015
[13] Zhang, H., Jolfaei, A., & Alazab, M. "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing", in IEEE Access, 2019.
[14] Schreuder, E., van Erp, J., Toet, A., & Kallen, V. L. "Emotional Responses to Multisensory Environmental Stimuli: A Conceptual Framework and Literature Review" in SAGE Open, Vol. 6, No. 1, 2016
[15] Zillmann, D. "Mood management through communication choices", Vol. 31, pp. 327–340, 1988.
[16] Tajadura-Jiménez, A., & Västfjäll, D. "Auditory-induced emotion: A neglected channel for communication in human-computer interaction", in Lecture Notes in Computer Science, pp. 63–74, 2008
[17] Stephanie Reeds, "The Power of Positivity: This Is How Your Words Can Restructure Your Brain", in Curious Mind Magazine.
[18] Laura Bezbatchenko, Sohinee Dutt, M.J. Juergens, Mary Lowery, Cassidy Pierce, Blythe Ramsay, David Ream Dan Talpas "What Emotions are Elicited from Different Genres of Music?", in Music and Emotions.
[19] Geetanjali Vaidya, "Music, Emotion and the Brain", in Serendip, 2004.
[20] Bakker, D. R., & Martin, F. H. "Musical chords and emotion: Major and minor triads are processed for emotion", in Cognitive, Affective and Behavioral Neuroscience, Vol. 15, No. 1, pp. 15–31, 2014.
[21] Brooke Neuman,"10 Shocking Benefits of Listening to Classical Music", in takelessons, 2016
[22] http://piano-midi.de
[23] http://piano-e-competition.com
[24] Hewahi, Nabil & AlSaigal, Salman & AlJanahi, Sulaiman. "Generation of music pieces using machine learning: long short-term memory neural networks approach", in Arab Journal of Basic and Applied Sciences, pp. 397-413, 2019.

 