[1]Jiang Qi
[2] Fangfang Li*

Design of Music Emotion Analysis and Creation Aid System Based on Machine Learning

Journal of Electrical Systems

Abstract: - In human-centric contexts, visual vision plays an increasingly important role in decision making, learning, communication, and situation awareness. Music is a language that conveys emotion to all creatures, including plants and animals. These emotional characteristics are included in the emotion found in music, along with other feelings conveyed via melodies, rhythms, and composer-highlighted moments. Human interaction, communication, decision-making, and cognitive processes are all heavily influenced by emotion. However, Subjective music emotion classification is made more difficult by pure machine classification due to its objectivity; additionally, emotion classification has not reached the same level of accuracy as style classification due to the complex and poorly understood mechanisms underlying how humans perceive music and generate emotions. This manuscript proposes a Multi-Scale Adaptive Graph Neural Network (MSAGNN)optimized with the Remora Optimization Algorithm (ROA) for predicting music emotion MEA-MSAGNN-ROA. Initially data is taken from Music Video Emotion (MVE) dataset Afterward the data is fed to Attribute-based Neural Collaborative Filtering (AbNCF) based pre-processing process. The outcome from the pre-processing data is transferred to the MSAGNN. The music emotions are successfully classified by using MSAGNN are exciting, fear, sad and relaxation. The ROA is used to optimize the weight parameter of MSAGNN. The proposed MEA-MSAGNN-ROA is applied in Python working platform. Performance parameters, like accuracy, precision, sensitivity, F1-score, ROC and recall analyzed to compute proposed approach. The proposed MEA-MSAGNN-ROA approach yields improved results in terms of accuracy (22.46%, 38.58%, 21.74%), sensitivity (21.97%, 33.88%, 25.52%), and precision. reduced computation times of 21.86%, 36.76%, 28.95%, and 81.46%, 95.97%, 86.77%.The proposed MEA-MSAGNN-ROA method is compared with the existing methods such as MEA-CNN, MEA-BPNN, and MEA-FNN, respectively. From the result it is concludes that the proposed MEA-MSAGNN-ROA method based accuracy is higher than the existing methods.

Keywords: Multi-Scale Adaptive Graph Neural Network, Remora Optimization Algorithm, Attribute-based Neural Collaborative Filtering, Music Emotion Analysis.

I. INTRODUCTION

In human-centric situations, both music and visual perception are important for communication, learning, situation awareness, and decision-making. Everyone, including plants and animals, can understand the emotions expressed in music [1-3].Though listeners may experience diverse emotions, musicians typically use dynamics or articulation to generate certain sensations in their work [4]. Human emotion is typically ill-defined and subjective, depending on human perception as well as changes in the surrounding environment [5]. These ambiguities typically show up in analyses of the emotions in music videos [6].

The difficulty of classifying subjective music emotions is increased by the objectivity of pure machine classification. Emotion classification is still not as accurate as style categorization due to complex and poorly understood mechanisms behind how people hear music and how emotions are created [7-9]. It is important to conduct studies on the specific features that are most successful at expressing musical emotion, how to evaluate the recovered features automatically, and which classification method might yield higher accuracy [10]. Even if they are unable to "hear" music, people with hearing impairments nevertheless want to play, learn about music, and feel its passion [11].

People's expectations for a spiritual life have increased as their material living standards have increased every [12]. Through the "listening-visual tactile" multisensory channels, normal-hearing individuals can participate and enjoy music synchronously, which can enhance the musical immersion [13–15]. Current psychology research has demonstrated that the perception of music is not just confined to hearing, but rather is a cognitive process synchronized by other sensory systems [16]. When given the proper stimulation, synaesthesia allows senses such as touch and vision to detect music [17]. Because normal people use the same brain regions to process tactile information as they do to interpret auditory information, people with hearing problems are nevertheless able to experience and enjoy music because of the tactile sense that their skin provides [18, 19]. To get the best regression result, a customized study needs to be done for certain problems [20].

[1]Professor, School of Music, Jining University, Qufu, Shandong, 273155, China
[2*]Lecturer, School of Art, Beijing Union University, Beijing, 100101, China
*****Corresponding author e-mail: ldtfangfang@buu.edu.cn

Major contribution to this work summarized here;

- The Music Video Emotion dataset is initially used to collect the data.
- Using a Data Pre-Processing Using Attribute-based Neural Collaborative Filtering to eliminate the noise of Music Video Emotion dataset in the pre-processing segment.
- The pre-processed data's are fed into the MSAGNN in order to effectively classifies the Exciting, Fear, Sad and Relaxation of the music data.
- The proposed MEA-MSAGNN-ROA technique is implemented and performance metrics such as ROC, accuracy, precision, sensitivity, F1-score, and recall are examined.

The remaining portions of this paper are organized as follows: segment 2 examines a survey of the literature; segment 3 explains the suggested approach; segment 4 presents the results and discussion; and segment 5 concludes.

II. LITERATURE SURVEY

Among the frequent research work on design of music emotion analysis and creation aid system; some of the latest investigations were assessed in this part.

Pandeya and Lee [21] have presented that design of music emotion analysis and creation aid system based on machine learning First, create a balanced collection of emotions from music videos that incorporates a variety of geographical, linguistic, cultural, and musical instrument diversity. Next, apply this dataset to four music and video convolutional neural networks (CNNs)—four unimodal and four multimodal. Next, each pre-trained unimodal convolutional neural network was individually adjusted, and its performance was evaluated with unidentified data. Additionally, train a 1-dimensional music emotion classifier based on convolutional neural networks using raw waveform data. The optimal model for integration into a multimodal structure was determined by comparing each unimodal classifier across many optimizers. For the multimodal classifier, the best unimodal modality was combined with complimentary characteristics from the music and video networks. Through the use of a late feature fusion technique, the multimodal structure combines all of the features from the music video and uses the Soft Max classifier to produce the final classification.

Wang [22] have presented that design of music emotion analysis and creation aid system based on machine learning. Examine the topic of emotion analysis in art. In comparison to a conventional weighted art sentiment analysis technique, when sentiment orientation was examined using the program in numerous investigations, it was more effective and may yield better findings. It provides a method for sentiment analysis of art based on the extraction of core words for long-form artworks using a conditional random field. An method for the emotional polarity weight synthesis of sentiment phrases was devised, and the assessment items from which the core sentences were taken were found using the conditional random field.

Xiang [23] have presented that design of music emotion analysis and creation aid system based on machine learning. A thorough examination of the computer's automated recognition technology in conjunction with its potent subcapacity allowed for the quick development of study on musical emotion. In addition to doing a thorough analysis of music emotion through the research of computer-related automatic recognition technology, the study looks at the technical research of automatic recognition and analysis of music emotion in the computer.

Xia and Xu [24] have discovered that design of music emotion analysis and creation aid system based on machine learning. The geographic range of the five different emotion types on the emotional plane was discovered in order to transform the music emotion classification problem into a regression problem. The relationship between the musical elements and the VA value was then determined using the regression technique. The training and testing phase includes the key components that were created and implemented for a regression-based music emotion categorization system. Three distinct algorithms were used to obtain the regression model during the training phase: k-plane piecewise regression, support vector regression, and polynomial regression.

Seo and Huh [25] have presented that design of music emotion analysis and creation aid system based on machine learning. People's emotional spectrum was used to accurately categorize music. In particular, it was important to develop techniques to automatically classify the new (unlearned) songs based on human emotion when they were added to an Internet of Things application related to music. One of the practical concerns for creating the applications was this point. Based on the emotional model, a survey was carried out to gather emotional data. Furthermore, a small and medium-sized business's working group was consulted when developing the musical features. Support vector machines and multiple regression analysis were used to classify emotions.

Dang et al. [26] have presented that design of music emotion analysis and creation aid system based on machine learning. It has many uses and has grown to be a potent way to find out what people think on social networks like Facebook and Twitter. However, sentiment analysis's efficacy and accuracy were being hampered by issues with NLP.A recent study suggests that deep learning models offer a potential remedy for the problems associated with natural language processing.

S. Hizlisoy et al. [27] have discovered that design of music emotion analysis and creation aid system based on machine learning. Furthermore, a novel Turkish emotional music database was developed, comprising 124 30-second-long Turkish traditional music samples. The effectiveness of the proposed methodology was assessed using this database. We employ features produced by feeding log-Mel filter bank energies and MFCCs to CNN layers in addition to conventional acoustic characteristics.

III. PROPOSED METHODOLOGY

In this segment, MEA-MSAGNN-ROA is discussed. As shown in Fig 1, the block diagram for the proposed methodology. There are three phases to it: classification, pre-processing, and data gathering. Consequently, a thorough explanation of each step is provided below.

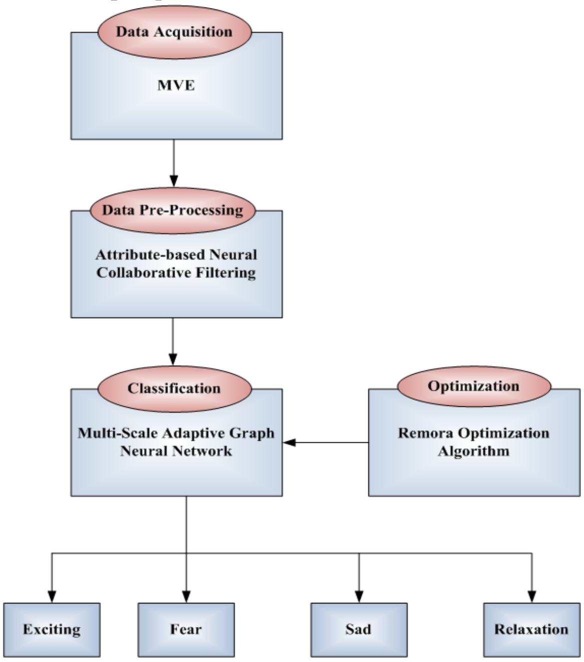

Fig 1: Block diagram for the proposed methodology

A. Data Collection

Our dataset has substantially more diversity in terms of languages, cultures, musical instruments, and geographical locations because the majority of the music videos are posted online. The emotion dataset for music videos is where the data was collected. Table 1music video emotion dataset shown below

Table 1: Music video emotion dataset

Emotion Class	Emotion Adjectives	No. of Samples
Excitation	Happy, Fun, Love, Joy, Pleasure, Exciting, Adorable, Cheerful, Surprising, Interest.	720
Fear	Horror, Fear, Scary, Disgust, Terror.	519
Neutral	Little (Sad, Fearful, Exciting, Relax) Ecstasy, Mellow.	599
Relaxation	Calm, Chill, Relaxing.	574
Sad	Hate, Depressing, Melancholic, Sentimental, Shameful, Distress, Anguish.	498
Tension	Anger, Hate, Rage.	528

B. Data Pre-Processing UsingAttribute-based Neural Collaborative Filtering

In this section, Attribute-based Neural Collaborative Filteringis proposed for pre-processing [28]. In pre-processing segment, it confiscates the reduction of the dynamics error using in the adaptive Attribute-based NCF.A deep recommendation technique for neural collaborative filtering that incorporates attribute information is proposed, along with its basic workflow. It is necessary to explain the extended matrix factorization model and the multi-layer perceptron model. Which make of attribute data. The specifics of how to incorporate the attribute data into the appropriate model will be covered in this procedure.

The interaction function f can be redefined as follows since it is defined by the multilayer neural network:

$$f\left(P^T V_{ui}^U, Q^T V_{vi}^V\right) = \phi\left(\varphi_X\left(\ldots\varphi_2\left(\varphi_1\left(P^T V_{ui}^U, Q^T V_{vi}^V\right)\right)\ldots\right)\right) \tag{1}$$

In this case, x^{th} NCF layers total and ϕ and φ_X stand for the output layers' and the xNCF layer's respective mapping functions. To have a better understanding of the multi-layer neural network, $P^T V_{ui}^U$ and $Q^T V_{vi}^V$ V could be renamed as v and$k$, correspondingly, to denote the representation vector that is produced when the entire user/item potential representation is subjected to the weight matrix calculation. Examine concurrently the interactions between users and items that are low-order linear and high-order nonlinear. Consequently, the mapping function $\varphi 1$ has two expressions.

$$\varphi_1'(k,v) = k \bullet v \tag{2}$$

$$\varphi_1''(k,v) = \begin{bmatrix} k \\ v \end{bmatrix} \tag{3}$$

The primary method by which the φ_1' the low-order linear relationship between the user and the object is found using the dot product operation routines. The φ_1' dot product operation is largely used by to determine the low-order linear relationship between the user and the item..The concatenation procedure is used by the function. At this point, the multilayer neural network transforms into:

$$f\left(P^T V_{ui}^U, Q^T V_{vi}^V\right) = \phi\left(\varphi_1'\left(P^T V_{ui}^U, Q^T V_{vi}^V\right)\right) \tag{4}$$

In order to mine the high-order nonlinear link among people and things, ANCF can additionally employ multi-layer neural networks:

$$f\left(P^T V_{ui}^U, Q^T V_{vi}^V\right) = \phi\left(\varphi_x\left(\ldots\varphi_2\left(\varphi_1''\left(P^T V_{ui}^U, Q^T V_{vi}^V\right)\right)\ldots\right)\right) \tag{5}$$

For the network output to provide a probabilistic interpretation of the ANCF, it must be restricted to the interval [0, 1]. Here, the sigmoid activation function on the output layer achieves our objective. As a result, the likelihood function has the following definition:

$$p\left(R^+, R^- \mid P, Q, \theta_f\right) = \prod_{(u,v)\in R^+} \hat{y}_{uv} \prod_{(u,k)\in R^-}\left(1-\hat{y}_{uk}\right) \tag{6}$$

Furthermore, taking the NLL, obtain:

$$L = -\sum_{(u,v)\in R^+} \log \hat{y}_{uv} - \sum_{(u,v)\in R^-} \log\left(1-\hat{y}_{uv}\right) \tag{7}$$

$$= -\sum_{(u,v)\in R^+ \cup R^-} y_{uv} \log \hat{y}_{uv} + \left(1-y_{uv}\right)\log\left(1-\hat{y}_{uv}\right) \tag{8}$$

Therefore, an objective function finally, utilizing the previously outlined methodologya suitable for gaining knowledge from data on implicit feedback is created. After the noise reduction of data, then the data are transferred to MSAGNN for classification.

C. Classification using Multi-Scale Adaptive Graph Neural Network (MSAGNN)

Classification of automatic generative fashion style utilizing MSAGNN [29] consisting of four primary components: a) a graph learning module that adapts to the graph to automatically infer relationships between variables; b)a pyramid network with multiple scales that preserves the temporal hierarchy at various levels; c) a pyramid network with multiple scales that preserves the temporal hierarchy at various levels; d) a scale-wise fusion module that efficiently encourages cooperation throughout a variety of time periods.

$$x *_G \theta = \sigma\left(\theta\left(\tilde{D}^{\frac{-1}{2}} \tilde{A} \tilde{D}^{\frac{-1}{2}}\right) x\right) \tag{9}$$

Here $\tilde{D}$ denoted diagonal degree matrixof $\tilde{A}$, $\tilde{A} = I_n + A$ indicates the self-connected adjacency matrix., and θ indicates the learnable parameter matrix. x indicates node representations, σ indicates activation function, and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ indicates graph with weighted adjacency matrix $G = (V, E, A)$

$$X_{rec}^{k} = \mathrm{Re}\,LU\left(W_{rec}^{k} \otimes X^{k-1} + b_{rec}^{k}\right) \tag{10}$$

The bias vector and convolution kernel are denoted by the letters W_*^k and b_*^k, respectively, in the pyramid k^{th} layer.

$$X_{norm}^{k} = Pooling \quad \left(\mathrm{Re}\,LU\left(W_{norm}^{k} \otimes X^{k-1} + b_{norm}^{k}\right)\right) \tag{11}$$

These two convolutional neural networks' outputs are then added point-wise at each scale.

$$X^{k} = X_{rec}^{k} + X_{norm}^{k} \tag{12}$$

Then, a broad range of temporal dependencies are retained in the learnt multi-scale feature representations due to their flexibility and breadth. Using a method to produce A_{full}^{k} full sparse will improve the model's robustness, reduce the graph convolution's computation cost, and minimize the impact of noise.

$$A^{k} = Sparse \quad \left(Soft\max\left(A_{full}^{k}\right)\right) \tag{13}$$

where the last neighboring matrix of the k^{th} layer denoted $A^k \in R^{N\times N}$. To capture scale-specific temporal patterns spanning time steps and variables, a multi-scale temporal graph neural network is presented. The scale-specific representations are combined by means of a weighted aggregation layer:

$$h_m = \mathrm{Re}LU \quad \left(\sum_{k-1}^{K} \alpha[k] \times h^k\right) \tag{14}$$

Here, the multi-scale representation that is finalized is denoted by h_m.

This convolutional neural network has a kernel size of $1\times d_s$ is incorporated into the output module to convert $h_m \in R^{N\times d_s}$ into the appropriate output dimension. . Then, a convolutional neural network with a kernel size of 1×1..is added to produce the anticipated values of $\hat{x} \in R^{N\times1}$.

As stated below is the objective function:

$$L = \frac{1}{T_{train}} \sum_{i=1}^{T_{train}} \sum_{j=1}^{N} \left(\hat{x}_{ij} - x_{ij}\right)^2 \tag{15}$$

Here N denotes the count of variables and T_{train}denotes the count of training samples. In the i^{th} sample, the j^{th} variable has a ground-truth of x_{ij} and a forecasted value of $\hat{x}_{ij}$ correspondingly.

To obtain the multi-scale matrix $H \in R^{K\times N\times d_s}$ the first step is to concatenate the scale-specific representations $\{h^1,..., h^k,..., h^K\}$.

$$H = Concat\left(h^1,..., h^k,..., h^K\right) \tag{16}$$

Here, the act of concatenation is indicated by the word. Next, we utilize a scale dimension average pooling layer.

$$h_{pool} = \frac{\sum_{k-1}^{K} H^k}{K} \tag{17}$$

$h_{pool} \in R^{1\times N\times d_s}$ where. Subsequently, Using a refining module, we flattened the $h_{pool}$and compressed the fine-grained data over a number of time scales. Two completely linked layers make up this module:

$$\alpha_1 = \mathrm{Re}\,LU \quad \left(W_1 h_{pool} + b_1\right), \tag{18}$$

$$\alpha = Sigmoid\left(W_2\alpha_1 + b_2\right), \tag{19}$$

where the weight matrices are W_1 and W_2 . b_1 And b_2 are vectors of bias.

Activation functions are often used in the second layer. The significance score vector, or $\alpha \in R^k$, is defined as the weight assigned to various scale-specific representations. Finally, the MSAGNN processing that divides the data into four categories: excited, afraid, sad, and relaxed. Typically, MSAGNN lacks the optimization technique needed to determine the best variables to validate a precise detection. The MSAGNN weight parameters W_1, W_2, are therefore crucial for the optimization method to work.

D. Remora Optimization Algorithm

This ROA is a brand-new meta-heuristic algorithm based on bionics and nature. The primary source of inspiration for ROA is the parasitic nature of remora. Different hosts updates different locations: Remotra avoids natural adversaries by feeding on the ecoparasites or wrecks of some large hosts, such as enormous whales. Remogra pursues its host to the bait-rich area where it finds prey in some small hosts; fast-moving swordfish are one example of this. Remora additionally renders certain judgments based on experience in the case of these two update techniques [30]. It notifies the host and updates everyone globally if it decides to pursue prey on its own. Remora keeps updating locally and does not modify the host if it consumes around the host. In Fig 2, the ROA flow chart is displayed.

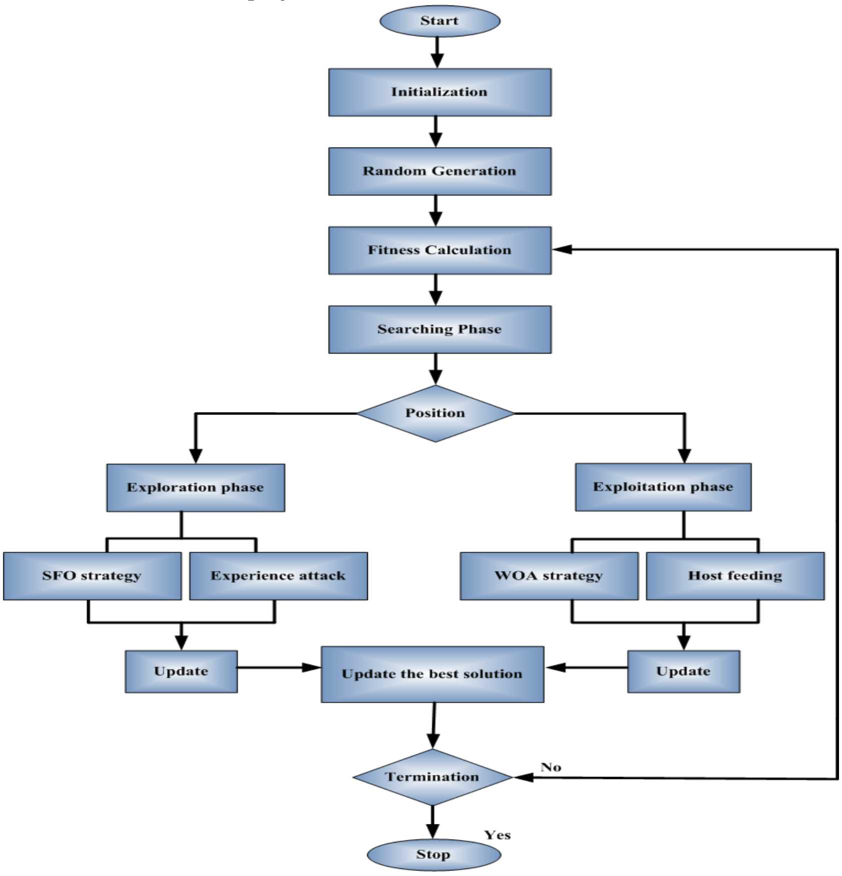

Fig 2: Flow chart of ROA

***Step 1*: Initialization**

Set the input parameters to their initial values. The MSAGNN weight parameters, designated as W_1 and W_2, are the input parameters in this case.

***Step 2*: Random Generation**

After setup, the random vectors produce the input parameters at random.

$$R = \begin{bmatrix} R_{1.1} & R_{1.2} & \dots & R_{1.x} \\ R_{2.1} & R_{2.2} & \dots & R_{2.x} \\ \dots & \dots & \dots & \dots \\ R_{y.1} & R_{y.2} & \dots & R_{y.x} \end{bmatrix} \tag{20}$$

Here, y indicates the count of variables and x indicates the count of dimensions in the solution space.

***Step 3*: Fitness Function**

Fitness is based on the objective function

$$\mathit{Fitness\ Function} = \mathit{Optimizing}\left[W_1 + W_2\right] \tag{21}$$

Step 4: Free travel (Exploration phase)

SFO Strategy

The swordfish's position can be seen of as being updated simultaneously with the remora's attachment. The position update formula for this method was enhanced based on its superior concept, yielding the subsequent formula:

$$R_i^{t+1} = R_{Best}^t - \left(rand(0,1) * \left(\frac{R_{Best}^t + R_{rand}^t}{2} \right) - R_{rand}^t \right) \tag{22}$$

Here, is the number of iterations that are presently underway, and T is the highest possible number of retries. R_{rand} is a point of arbitrariness. The primary criterion used to select remora for various hosts is whether or not it has consumed prey, or whether the current fitness function value obtained is superior to that of the preceding generation.

Experience Attack

It must continually take tiny steps around the host, much like the accumulation of experience, in order to assess whether changing the host is essential. The following formula is used to model the aforementioned concepts:

$$R_{att} = R_i^t + \left(R_i^t - R_{pre} \right) * randn \tag{23}$$

where R_{pre}, which can be viewed as a sort of experiment, represents the position of the preceding generation. R_{att} is a provisional measure. Remora's action qualifies as a "small global" movement when it occurs. This mechanism has been evolved across a larger spectrum of development and can take predictability into consideration while efficiently jumping out of the local optimum.

Step 5: Eat thoughtfully (Exploitation phase***)***

WOA Strategy

The location update formula of the remora attached to the whale may be retrieved by utilizing the basis of the original ROA technique. As demonstrated below:

$$R_{i+1} = D * e^a * \cos(2\pi a) + R_i \tag{24}$$

When a remora is on a whale, their positions might be considered as equal in a larger solution space. D indicates distance that separates the target and the hunter.

Host Feeding

An additional part of the exploitation phase is host feeding., the host's location space can be used to represent the solution space. One could consider moving past or around the host to be a tiny step.

$$R_i^t = R_i^t + A \tag{25}$$

Here, the letter A was utilized to indicate a tiny movement that was connected to the host and remora's volume spaces. To determine where the host and remora are located in the solution space

Step 6: Update the Best Solution

By taking into account the optimal solution, the algorithm modifies music emotion. The optimal solution has been chosen. A emotion is selected as random. After updating the search position with data from Music Video Emotion, the ROA iteration is deemed to be finished.

Step 7: Termination

Verify the termination criteria; if it is met, the best possible solution has been found; if not, repeat the procedure. Utilize the generator's optimized weight parameter values that you obtained from the MEA by using the Remora Optimization Algorithm. Assess music emotion more accurately.

IV. RESULT AND DISCUSSION

In this segment, the experimental outcome of the MEA-MSAGNN-ROA method assessment music emotion analysis approach is discussed. The Python working platform is used for the simulations. Python is used to simulate the suggested approach under various performance criteria. Results of MEA-MSAGNN-ROA were examined using the MEA-CNN, MEA-BPNN, and MEA-FNN techniques, among others.

A. Performance measures

In order to choose the optimal classifier, this is an important task. Performance measures including recall, F1-score, accuracy, precision, and ROC are analyzed in order to assess performance. The confusion matrix is used to determine how to scale the performance measures. The True Positive, True Negative, False Positive, and False Negative, values are needed to scale the confusion matrix.

- True Positive (TP): Samples in which the real class label is exactly the same as the predicted class label when the count is positive.
- True Negative (TN): Number of samples where the actual class label is exact and the predicted class label implies a negative value.
- False Positive (FP): The number of samples in which the real class label is imprecise and the predicted class label implies a positive value.
- False Negative (FN): Number of examples where the real class label is imprecise and the predicted class label implies a negative value.

1) Accuracy

It is the proportion of the total number of predictions to the total number of items for a given dataset that were properly predicted. To quantify it, use equation (26).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{26}$$

2) F1-Score

The proposed MEA-MSAGNN-ROA technique's performance is examined using the F-score metric. It's calculated in Equation (27),

$$F\ score = F1score = \frac{TP}{TP + \frac{1}{2}[FP + FN]} \tag{27}$$

3) Precision (P)

A statistic called precision counts the number of correctly predicted favourable outcomes. This is computed via following Eqn (28)

$$P = \frac{TP}{TP + FP} \tag{28}$$

4) Recall (R)

Recall is a measure that determines how accurate a forecast is based on the total number of accurate forecasts made. To measure it, use Equation (29),

$$R = \frac{TP}{TP + FN} \tag{29}$$

B. Performance Analysis

Fig 3 to 8 shows the simulation outcomes of MEA-MSAGNN-ROA. Then the outcomes are analyzed with existing MEA-CNN, MEA-BPNN, and MEA-FNN methods.

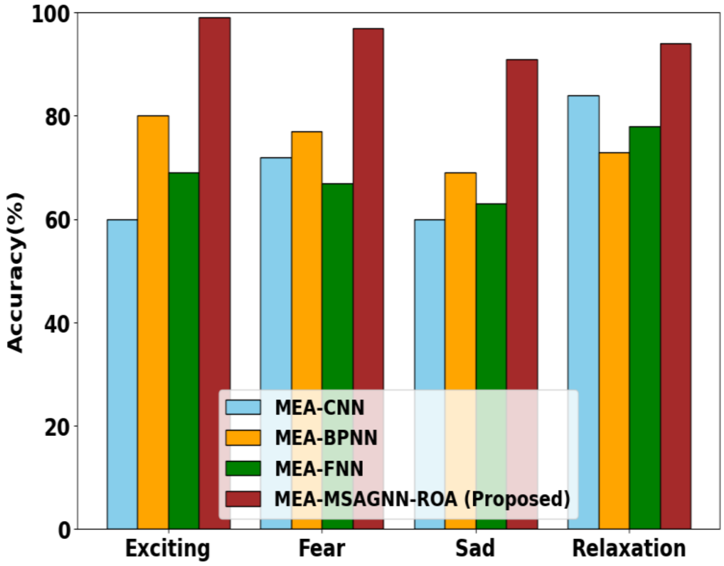

Fig 3: Performance analysis of accuracy

The performance analyses of accuracy is displays in Fig 3.The performance of the proposed MEA-MSAGNN-ROA techniques results in accuracy that are 37.58%, 58.56%, 47.27%, higher for the classification of Excitement, 22.46%, 38.58%, 21.74%, higher for the classification of Fear, 21.44%, 30.86%, 25.43%, higher for the classification of Sadness and 30.95%, 15.13%, 18.53%, higher for the classification of Relaxation when evaluated to the existing MEA-CNN, MEA-BPNN, and MEA-FNN models respectively.

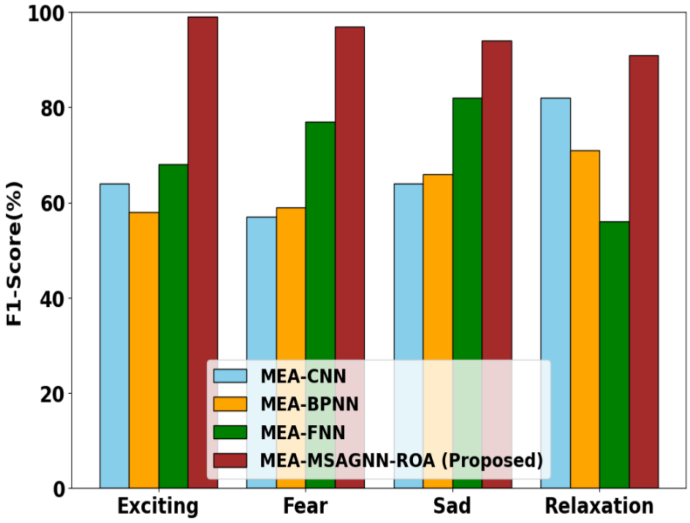

Fig 4: Analysis the performance of F1-score value

Analysis the performance of F1-score value with proposed method is displays in Fig 4.The performance of the proposed MEA-MSAGNN-ROA technique results in F1-score that are 32.41%, 26.59%, 35.94%,%, higher for the classification of Excitement, 22.66%, 36.52%, 23.97%, higher for the classification of Fear 21.45%, 23.96%, 28.83% higher for the classification of Sadness, and 40.48%, 32.86%, 19.48%higher for the classification of Relaxation when evaluated to the existing MEA-CNN, MEA-BPNN, and MEA-FNN models correspondingly.

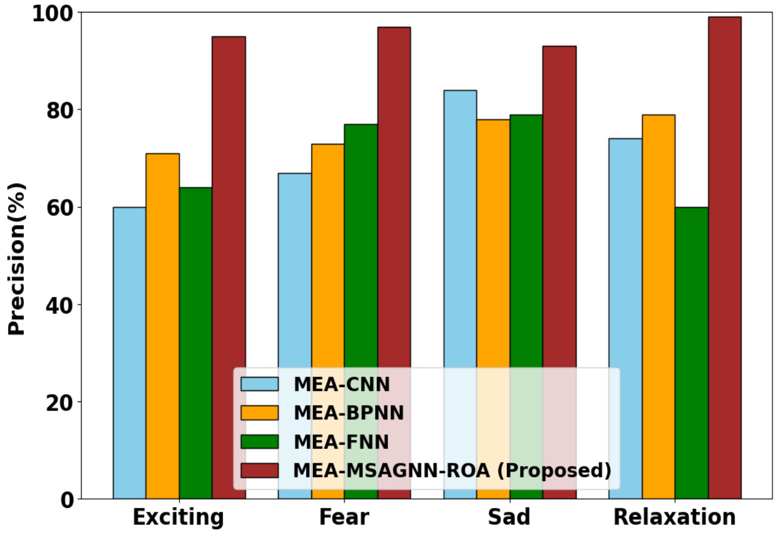

Fig 5: Comparison of the precision value between the existing and proposed techniques.

A comparison of the precision value between the proposed and current systems is displays in Fig 5. Here, a direct comparison with proposed approaches is offered to show how the suggested method's precision is higher. The proposed method provides for a more extensive analysis of a proposed and has higher precision than existing methods due to its wider consideration of factors. The performance of the proposed MEA-MSAGNN-ROA technique results in precision that are 21.86%, 36.76%, 28.95%, higher for the classification of Excitement, 21.86%, 27.56%, 31.94% higher for the classification of Fear and 38.95%, 18.56%, 23.47% higher for the classification of Sadness, 22.84%, 30.96%, 16.47% higher for the classification of Relaxation when evaluated to the existing MEA-CNN, MEA-BPNN, and MEA-FNN models correspondingly.

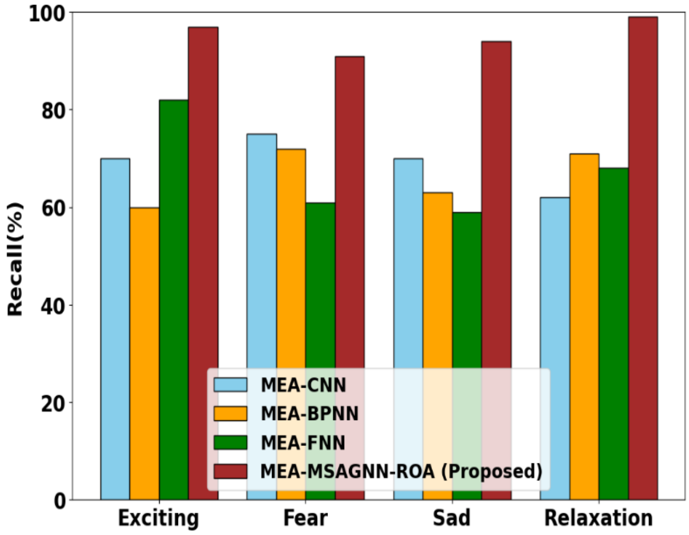

Fig 6: Comparing the recall value using the existing and proposed approaches.

The Comparison of the precision value between the existing and proposed techniques is displays in Fig 6. The performance of the proposed MEA-MSAGNN-ROA technique results in recall that are 29.86%, 21.71%, 35.47%higher for the classification of Excitement, 31.96%, 26.56%, 21.12% higher for the classification of Fear and 29.43%, 24.97%, 19.48% higher for the classification of Sadness, 19.47%, 28.97%, 22.45% higher for the classification of Relaxation when evaluated to the existing MEA-CNN, MEA-BPNN, and MEA-FNN models correspondingly.

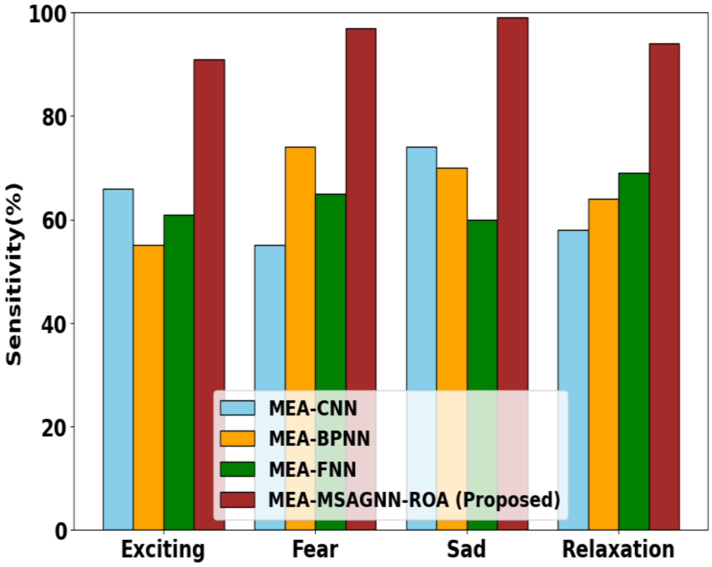

Fig 7: Comparing the proposed and current methods with the sensitivity value

Comparing the recall value using the existing and proposed approaches .with the sensitivity value is displays in Fig 7. The performance of the proposed MEA-MSAGNN-ROA technique results in sensitivity that are 35.34%, 21.59%, 30.27%, higher for the classification of Excitement, 21.97%, 33.88%, 25.52% higher for the classification of Fear and31.55%, 27.79%, 18.99% higher for the classification of Sadness, 20.46%, 23.57%, 29.14% higher for the classification of Relaxation when evaluated to the existing MEA-CNN, MEA-BPNN, and MEA-FNN models respectively.

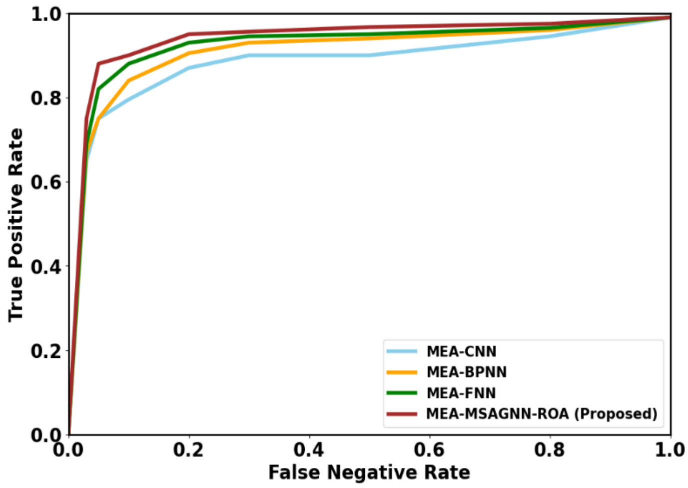

Fig 8: Performance analysis of ROC

The performance analysis of ROC is shown in fig 8. Each point on the curve represents a different threshold value, and the curve is created by connecting these points. The prediction system performs better the closer the curve is to the top-left corner of the graph. Better performance in differentiating between positive and negative examples is indicated by a greater ROC..The proposed MEA-MSAGNN-ROA methods the ROC provides high music emotion compared with other existing methods. The existing methods like MEA-CNN, MEA-BPNN, and MEA-FNN the ROC become lower compared with the proposed MEA-MSAGNN-ROA technique.

V. CONCLUSION

In conclusion, this research harnesses the power of MSAGNN to identify the music emotion and analyse it. During pre-processing, the music data is processed using attribute-based neural collaborative filtering. The MSAGNN receives the pre-processing output and uses it to effectively categorize the feelings of excitement, fear, sadness, and relaxation through music. The proposed method is assessed using the Python working platform and contrasted with other existing methods. The proposed method is examined in several scenarios, including those involving recall, ROC, F1-score, accuracy, precision, and sensitivity. The MSAGNN classifier is used in the system's last stage. Which has increased its accuracy to 98% due to its precise identification of the suggested optimum region expansion technique. This shows that the system is capable of accurately detecting the emotion evoked by music.

REFERENCES

[1] Cunningham, S., Ridley, H., Weinel, J., & Picking, R. (2021). Supervised machine learning for audio emotion recognition: Enhancing film sound design using audio features, regression models and artificial neural networks. *Personal and Ubiquitous Computing, 25*, 637-650.

[2] Xu, L., Wen, X., Shi, J., Li, S., Xiao, Y., Wan, Q., &Qian, X. (2021). Effects of individual factors on perceived emotion and felt emotion of music: based on machine learning methods. *Psychology of Music, 49*(5), 1069-1087.

[3] Rahman, J. S., Gedeon, T., Caldwell, S., Jones, R., & Jin, Z. (2021). Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres. *Journal of Artificial Intelligence and Soft Computing Research, 11*(1), 5-20.

[4] Zhang, J., Yin, Z., Chen, P., &Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion, 59*, 103-126.

[5] Florence, S. M., & Uma, M. (2020, August). Emotional detection and music recommendation system based on user facial expression. *In IOP conference series: Materials science and engineering* (Vol. 912, No. 6, p. 062007). IOP Publishing.

[6] Wen, X. (2021). Using deep learning approach and IoT architecture to build the intelligent music recommendation system. *Soft Computing, 25*, 3087-3096.

[7] Huang, I. S., Lu, Y. H., Shafiq, M., Ali Laghari, A., &Yadav, R. (2021). A generative adversarial network model based on intelligent data analytics for music emotion recognition under IoT. *Mobile Information Systems, 2021*, 1-8.

[8] Gómez-Cañón, J. S., Cano, E., Eerola, T., Herrera, P., Hu, X., Yang, Y. H., & Gómez, E. (2021). Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications. *IEEE Signal Processing Magazine, 38*(6), 106-114.

[9] Panda, R., Malheiro, R. M., &Paiva, R. P. (2020). Audio features for music emotion recognition: a survey. *IEEE Transactions on Affective Computing*.

[10] Pandeya, Y. R., Bhattarai, B., & Lee, J. (2021). Deep-learning-based multimodal emotion classification for music videos. *Sensors*, *21*(14), 4927.

[11] Hassouneh, A., Mutawa, A. M., &Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, *20*, 100372.

[12] Khan, M. A., Abbas, S., Raza, A., Khan, F., &Whangbo, T. (2022). Emotion Based Signal Enhancement Through Multisensory Integration Using Machine Learning. *Computers,Materials & Continua, 71*(3).

[13] Chan, K. Y., Kwong, C. K., Wongthongtham, P., Jiang, H., Fung, C. K., Abu-Salih, B., ... & Jain, P. (2020). Affective design using machine learning: a survey and its prospect of conjoining big data. *International Journal of Computer Integrated Manufacturing, 33*(7), 645-669.

[14] Alhalaseh, R., &Alasasfeh, S. (2020). Machine-learning-based emotion recognition system using EEG signals. *Computers*, *9*(4), 95.

[15] Kim, T. Y., Ko, H., Kim, S. H., & Kim, H. D. (2021). Modeling of recommendation system based on emotional information and collaborative filtering. *Sensors, 21*(6), 1997.

[16] Doma, V., &Pirouz, M. (2020). A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *Journal of Big Data, 7*(1), 1-21.

[17] Chaturvedi, V., Kaur, A. B., Varshney, V., Garg, A., Chhabra, G. S., & Kumar, M. (2022). Music mood and human emotion recognition based on physiological signals: asystematic review. *Multimedia Systems, 28*(1), 21-44.

[18] Yadav, S. P., Zaidi, S., Mishra, A., &Yadav, V. (2022). Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (RNN). *Archives of Computational Methods in Engineering, 29*(3), 1753-1770.

[19] Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., &Valtonen, T. (2021). Machine learning for middle schoolers: Learning through data-driven design. *International Journal of Child-Computer Interaction, 29*, 100281.

[20] Rajesh, S., &Nalini, N. J. (2020). Musical instrument emotion recognition using deep recurrent neural network. *Procedia Computer Science, 167*, 16-25.

[21] Pandeya, Y. R., & Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. *Multimedia Tools and Applications, 80*, 2887-2905.

[22] Wang, D. (2022). Research on the art value and application of art creation based on the emotion analysis of art. *Wireless Communications and Mobile Computing*, *2022*, 1-10.

[23] Xiang, Y. (2022). Computer analysis and automatic recognition technology of music emotion. *Mathematical Problems in Engineering*, *2022*, 1-9.

[24] Xia, Y., &Xu, F. (2022). Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm. *Mathematical Problems in Engineering*, 2022.

[25] Seo, Y. S., & Huh, J. H. (2019). Automatic emotion-based music classification for supporting intelligent IoT applications.Electronics, *8(*2), 164.

[26] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, *9*(3), 483.

[27] Hizlisoy, S., Yildirim, S., &Tufekci, Z. (2021). Music emotion recognition using convolutional long short term memory deep neural networks. *Engineering Science and Technology, an International Journal*, *24*(3), 760-767.

[28] Chen, H., Qian, F., Chen, J., Zhao, S., & Zhang, Y. (2021). Attribute-based neural collaborative filtering.*Expert Systems with Applications, 185*, 115539.

[29] Chen, L., Chen, D., Shang, Z., Wu, B., Zheng, C., Wen, B., & Zhang, W. (2023). Multi-scale adaptive graph neural network for multivariate time series forecasting.*IEEE Transactions on Knowledge and Data Engineering.*

[30] Jia, H., Peng, X., & Lang, C. (2021). Remora optimization algorithm. *Expert Systems with Applications*, *185*, 115665.