Emotion Recognition Using Human GAIT with Chiropteran Mahi Metaheuristic Algorithm

Prachi Jain[1], Vinod Maan[2]

School of Engineering and Technology, Mody University of Sc. & Tech., Lakshmangarh, India

School of Engineering and Technology, Mody University of Sc. & Tech., Lakshmangarh, India

Abstract

A biometric identification system that analyzes body movements can increase interactive-modeling, robotics, virtual-authenticity, and biometric-identity through automatic recognition of emotions. A computer network, that deduces human emotions from body language, significantly alters how people interact with machines. Identification of emotion-specific features in numerous descriptions of human body movements is challenging. In this study, Deep BiLSTM system created using CMO method successfully detected human feelings. Gaits, movies, or feelings that were tagged serve as initial input for suggested emotion recognition-based method. For training, BiLSTM classifier is fed the most relevant features that were extracted and concatenated. Proposed Chiropteran Mahi optimization identifies better results for identifying revealed emotions of the classifier by tuning hyper parameters effectively. Learned videos and testing database are used to determine validity of proposed technique and educate it. Suggested approach has increased accuracy by 12.951% (depending on training%) and 7.793% based on K-Fold value compared with current approaches.

Keywords: BiLSTM, Optimization, Emotion recognition, Feature Selection, Skeletonization, Facial expressions, and Grid-based features.

1. Introduction

In recent years, identifying emotions in human-computer communication has gained significant attention as a research avenue. The key step in emotion recognition is analyzing computer-related information to extract features that reflect emotions. The mapping between these features and emotions is then established to categorize emotional states. This is crucial in understanding how people express their feelings and attitudes through emotions, which enables partners to react appropriately to their emotional state. Over the past few years, several studies have focused on identifying emotions based on body language, particularly through gestures, posture, and movement from a distance [1]. Movement, in particular, is considered a more effective means of conveying emotions than verbal language [7]. However, in noisy or long-distance environments where emotions are expressed through spoken language, the deaf may struggle to fully comprehend the speaker's intended message. Face emotions, body posture, or gestures are the primary components of body movements that can be seen & comprehended [4]. Body movement is a means of conveying information with a high level of emotional expression in human communication [5].

Although it can be challenging to understand people's emotions using non-verbal cues, humans tend to rely on them, particularly body movements and facial expressions, to express themselves [6]. Recent research on emotion recognition has predominantly focused on facial expressions due to the abundance of available data [9][6]. The identification of emotions has been the subject of numerous publications in various fields, including healthcare, performance monitoring, biometric security, and gaming [10][1]. Referential expressions or mock facial expressions can be unreliable in certain scenarios, as they may depend on the presence or absence of an audience [6]. Similarly, when considering whole-body expressions, even repeated movements can vary based on factors such as energy, expression, emotion, and intensity. Developing universal algorithms for accurately discerning human emotions is a complex task that cannot be achieved easily [4]. Psychological research suggests that to evaluate human performance, it is important to identify which kinetic characteristics are associated with expressing a particular type of emotion [8]. Pleasure is always characterized by serene and tranquil acts Contrary to melancholy, which is associated with greater & quicker movements[5].

Most of the conventional algorithms used for emotion recognition, such as K-means, Support Vector Machine, and apriori algorithm, primarily rely on worsening analysis [3]. While these algorithms can process data quickly and efficiently, they may not be effective for identifying emotions when dealing with vast amounts of data. In contrast, [1] used the Laban Movement Analysis (LMA) method to capture qualitative and quantitative features and understand the movement of the human body. The LMA technique considers four key elements - Shape, Space, Effort, and Body - to establish a connection between the qualities of motion and the corresponding emotion being expressed [1]. Recently, the graph neural network has been incorporated into the Convolutional Neural Network architecture for arbitrary structure graphs, and this approach is increasingly being used for recognizing actions [12], recommender systems [14], traffic forecasting [13], and other applications [2]. The ANOVA algorithm is considered to be the most flexible feature selection method due to its reliance on two stages - reliability and monotonicity. Genetic algorithms are used to extract features that maximize the rate of emotion identification, primarily by relying on binary chromosomes. As a result, supervised machine learning algorithms are used to recognize human emotions, and this approach has been proven to be more efficient in the biometric domain [1]. This research proposes a method called Chiropteran Mahi optimization to overcome challenges in identifying emotions from moving gestures. The method uses relevant texture and grid-based skeleton features to accurately classify three basic emotions (sad, angry, and happy) using a BiLSTM classifier. The key step in identifying emotions from walking data is skeletonization, which separates features based on the shape of the human body movement. The proposed approach's major contribution is its ability to accurately classify emotions using only relevant features extracted from input video image data, which is explains as follows:

- ✐ The concatenated characteristics are sent to the CMO, which improves the classifier's effectiveness in order to allow the BiLSTM classifier to accurately classify mood.
- ✐ The facial expressions of humans in the database give rise to the textural pattern attributes LTP, LOOP, and LGP, while the grid-based characteristics are obtained through the process of Image skeletonization.

Section 2 presents various current methods for emotion recognition, along with their associated challenges. Section 3 presents the proposed optimization, while the system model for emotion recognition is detailed in Section 4. The outcomes and discussion of the suggested device are presented in Section 5, and Section 6 concludes the paper.

2. Motivation

This section reviews the most effective and widely used methods for emotion recognition, taking into account their contributions, advantages, and disadvantages. Additionally, the issues addressed by various researchers in this area are discussed in this section.

2.1 Literature review

Mona Algarni *et al*.[1] enabled the emotion recognition model operate better and produce more accurate findings, which in turn helps doctors make the right decisions. The databases of Emotion Analysis through Physiological Signaling (DEAP) was employed as per usual procedure. The Binary Gray Wolf Optimizer was used to carry out the feature selection task. The stacking bi-directional Long Short-Term Memory (Bi-LSTM) Model was employed to identify human emotions during the categorization step. Arousal, valence, and liking are the three broad categories that make up emotions. The proposed methodology outperformed the methods employed in previous studies, with an average accuracy of 99.45%, 96.87%, and 99.68% for valence, arousal, and liking, respectively. This is considered a strong performance for the emotion detection model. Zihao Ye *et al.*[2] SVM-NB was suggested as a classification method to get more emotional polarity. The classifier is then used to extract the text's emotional polarities, comprising both positive & negative ones. The results of the studies demonstrated that the suggested emotion recognition approach is more accurate & robust than the standard model recognition approach. Zheng Liu *et al*.[3] One of the issues with speech emotion recognition is trying to achieve higher recognition accuracy with a small corpus. To improve the generalization of speech characteristics, authors combined acoustic & pre-trained speech characteristics. They also presented a brand-new feature fusion model based on Transformer and BiLSTM. Merged the acoustic properties of the voice with the speech pre-trained features extracted by Tera, Audio Albert, and Npc, and carried out experiments on the CASIA Chinese voice emotion corpus. The outcomes demonstrated that our approach & model had a 94% accuracy rate in the Tera model. Om Prakash Yadav *et al*.[4] To acquire emotional features from scaled spectrograms of spoken utterances, the Convolutional Recurrent Neural Network (CRNN), which consists of CNN and BiLSTM, is developed. Two commonly utilized datasets, Berlin EMO-DB in German and RAVDESS in North American English, are utilized to test the model that is suggested. The created model's performance is seen to increase to 85.76% for RAVDESS and 91.59% for Berlin EMO-DB, and the model is noted to be language independent. Mustaqeem *et al.* [5 proposed a unique structure for SER that selects critical sequence segments based on clustered comparisons of redial-based function networks (RBFN). The STFT method is used to turn the chosen sequence into a spectrogram, which is then fed into the CNN model to derive the prominent and discriminative characteristics from the spoken spectrogram. When tested against state-of-the-art SER approaches, the experiments show that the recommended SER model is both reliable and effective. It can attain accuracy levels of up to 72.25%, 85.57%, & 77.02% over the IEMOCAP, EMO-DB, and RAVDESS datasets, respectively. Fachang *et al.*[6] introduced the DE-CNN-BiLSTM innovative approach, which integrates the complexity of EEG signals, the spatial organization of the brain, and the temporal settings of emotion creation. The approach obtains an average accuracy of 94% for the DEAP database and 94.82% for the SEED dataset, According to the simulation outcomes indicated its excellent accuracy and solid robustness. Chen *et al.*[7] Applied CNN-LSTM combined networks, suggested a multifeature mixed network classifier based on CNN-LSTM that combines 1D (single-dimensional) feature input through DNN and 2D feature input through CNN-LSTM to address the shortcomings of the original single feature prototypes. The multifeature mixed network classifier in this research achieves an audio classification accuracy of 68% and a lyric classification accuracy of 74%, according to experiments on a dataset of one million songs. The multimodal's average classification accuracy is 78%, a significant improvement when contrasted to the single-modal. T. B. Alakus *et al.*[8] Positive & negative emotions were predicted via motion analysis using EEG signals. The GAMEEMO data set provided EEG signals. A ROC value of 90% and an accuracy of 76.91% were attained using the suggested strategy.

2.2 Challenges

This section discusses the challenges that have arisen during research into emotion recognition through body movement.

- One of the primary difficulties in recognizing emotions is presenting feature movement in high dimensionality. Additionally, the analysis of appropriate features for emotional categorization is challenging due to limited guidance and a restricted number of features. Early research, which is frequently biased by focusing on a small number of traits, occasionally ignores crucial elements [1]. In the context of social robots, the communication flow is highly unpredictable and constantly changing. This is especially crucial in child-robot interaction (CRI), as the interaction with adults tends to be more controlled and limited. As a result, children may struggle in these situations, presenting a significant challenge [7].
- The use of video data in emotion recognition encounters several technical issues, such as the complexity of the background, diverse behaviors of people, and variations in camera perspective and scale of individuals within the scene [8]. Human emotional recognition is inaccurate if the captured 3D images or postures are noisy. Whole-body joint locations are required for successful analysis, but there may be an absence of a whole-body pose if there are any obstructions in the video [6].
- Although it is a difficult job, training the model to identify emotion from gait is very effective. The important problem for gait skeleton sequences is achieving complex spatial and temporal dependence [2].

3. Simulation Model for Emotion Detection using the CMO-based BiLSTM Classifiers,

The "EWalk" database, which consists of gaits, videos, as well as the identified human feelings, is the first input for the suggested emotion recognizing algorithm. Before being fed to the deep learning network, the input video database undergoes preprocessing to improve the quality of the dance or motion emotion data. The preprocessing stage generates novel features based on textural patterns and the skeleton of the human body. The textural pattern features comprise facial features like LTP, LOOP, and LGP, while the skeleton-based features involve grid-based characteristics. The method of feature concatenation, which combines the different features produced from preprocessing or image segmentation, is used to integrate the obtained characteristics. The concatenated characteristics are sent into the CMO, which improves the classifier's effectiveness in order to allow the BiLSTM classifier to accurately classify feeling.

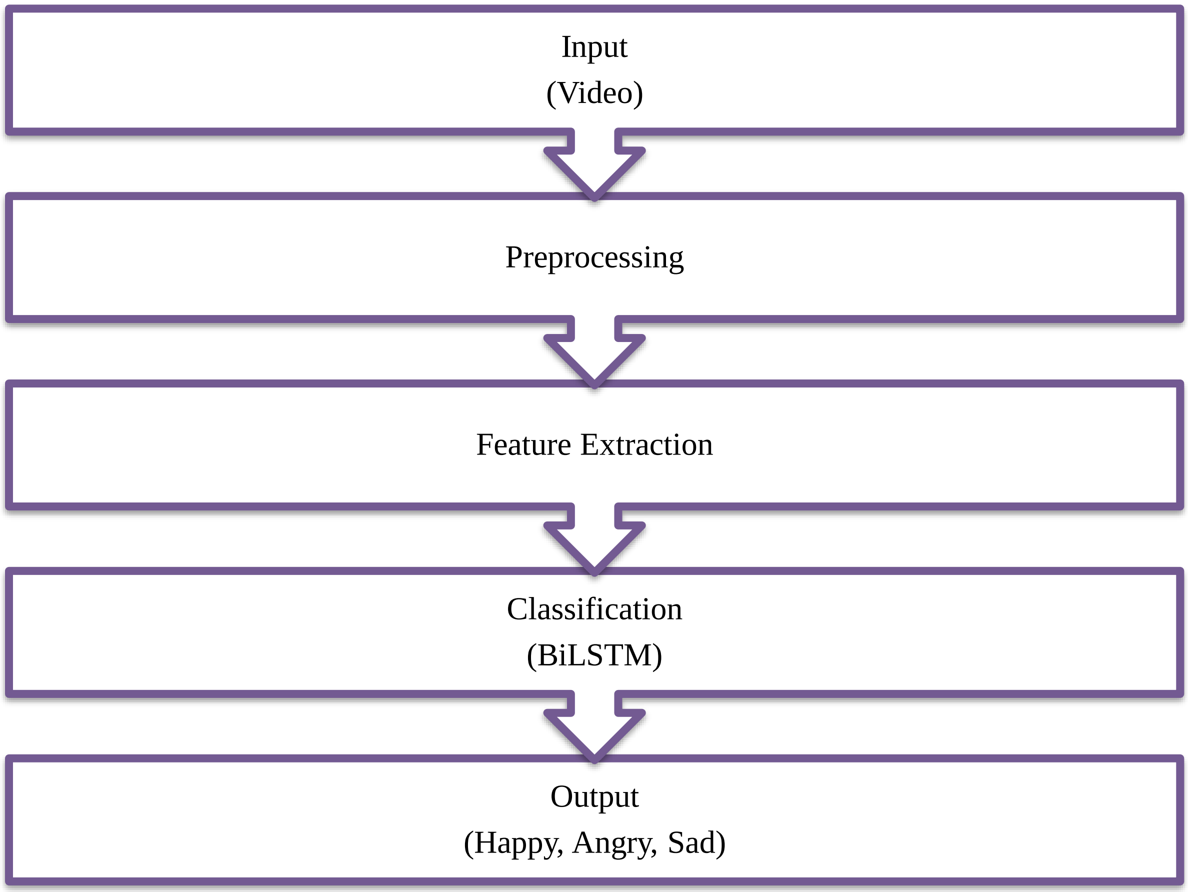


Figure 1. Emotion Recognition System Model

3.1 Preprocessing

Preprocessing helps the classifier obtain appropriate features for better emotion recognition by increasing accuracy and reducing overfitting. High-quality images have high frequency and error approximation. Unwanted features and noises are eliminated in order to obtain spotless data with increased size based on the datasets used. Colored images are transformed into grey and resized for uniformity. Eliminating noisy and inexpressive features enhances performance and contrast. Low pass filtering screens and structures processed images for appropriate preprocessed image derivation.

3.2 Image Skeletonization

The proposed method for predicting human emotions involves the utilization of skeletonization and novel textural pattern features. The skeletonization process is employed to extract the object based on the central region, reducing the foreground region from the original image, and organize the body movement based on the order of joint locations of humans. The grid-based features are then obtained from the image skeletonization process. On the other hand, the novel textural pattern features are based on the facial features of humans to detect the region of interest in the provided input image of the video database. Texture holds significant characteristics and vital data related to the physical arrangement of the surrounding atmosphere. The proposed method utilizes novel texture features to determinethe distance and space between adjacent cells present in the image.

3.2.1 LTP: The developed method employs Local Ternary Pattern to recognize the facial expression and extract specific features from the input image of the video database. The upper nose, brows, mouth, and eyes are characteristics connected to feelings, as well as the LTP feature descriptor encodes this data in these characteristics The majority of the current algorithms for feature analysis do not consider the center pixel in their final calculations, but with the LTP used in the proposed method, the center pixel is taken into account. Even in the presence of noise in the input image, the LTP only affects a single feature at a time, making it superior to both LBP and LDP.

3.2.2 LGP: "In every image, the pixels are organized in a two-dimensional space using squares, resulting in the least amount of information."The LGPH approach calculates the number of image pixels that have feature elements in every phase. "The Local Gradient pattern extracts the LGP feature by describing the central pixel of the region of interest with two-bit binary patterns, and this is repeated for each pixel."

*3.2.3 LOOP:*The execution time of the Local Optimal-Oriented Pattern is lower and the formulation of the feature descriptor is simple and easy toimplement.For eachpixel in the picture, the LOOP value $\left(y_i, z_i\right)$ is formulated as,

$$LOOP\left(y_i, z_i\right) = \sum_{j=0}^{k} m\left(n_r - n_i\right).2^{x_r} \tag{10}$$

Where

$$m\left(y\right) = \begin{cases} 1 & if \quad y \geq 0 \\ 0 & otherwise \end{cases} \tag{11}$$

3.3 Feature Concatenation

The proposed methodology utilized textural pattern features such as LTP, LOOP, and LGP, along with grid-based features obtained from image skeletonization to analyze facial expressions in a database. These multiple features were concatenated and fed to a BiLSTM classifier for training, with the joint view depiction of all the features highly conserving the original information. The trained video and test video were then provided as input for modeling, with the performance of the BiLSTM classifier optimized and well-tuned utilizing the CMO. The optimization algorithm minimized the total number of variables in the features to attain an optimal solution. The trained video was used to teach the performance of the classifier and identify patterns, with the model accuracy calculated using the test data.

3.4 Optimized BiLSTM classifier for emotion classification

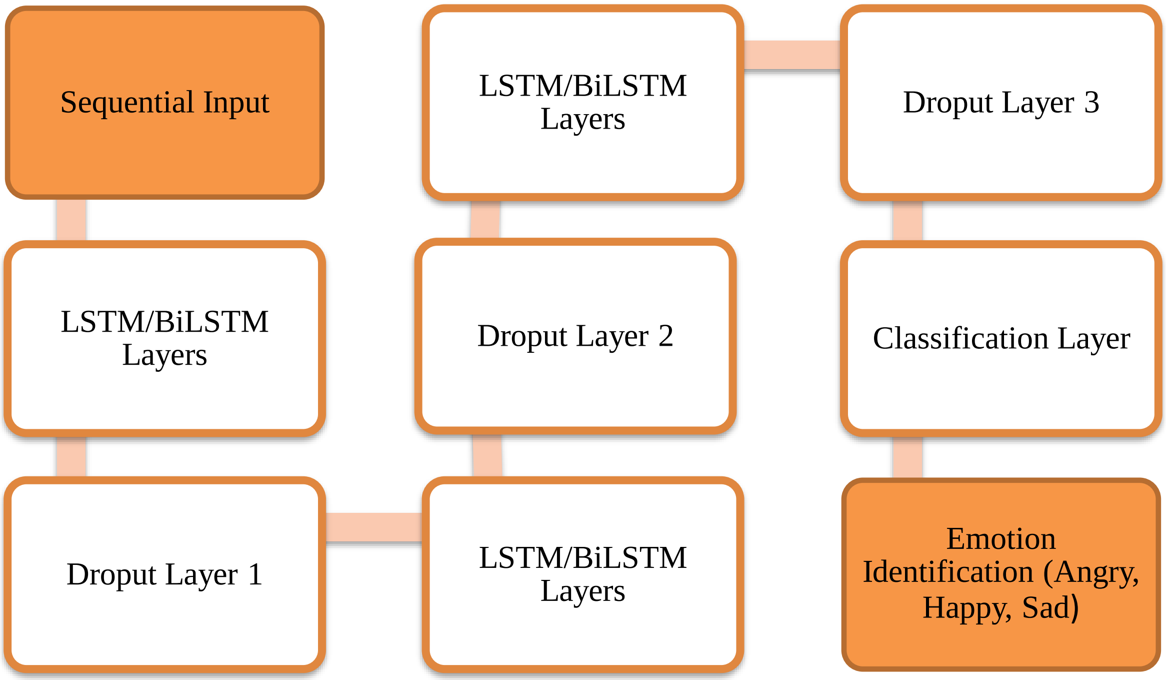


Figure 2. BiLSTM Classifier for the Emotion Recognition Model

The BiLSTM network is composed of specific layers that perform individual tasks, with the activation layer transforming the output from the input weights during the forward layer. The output of the BiLSTM neural network depends on both preceding and following segments, as both the forward and backward layers are utilized in the network with backpropagation adjusting weights and biases. A deep BiLSTM classifier is utilized for emotion recognition, extracting key features to enhance analysis. The classifier is bidirectional, allowing for accurate emotion prediction by evaluating input data in both directions. The bias and weight matrices are ideally adjusted via the CMO method, with input values preserved in the weight vector & biases maintained in every pre-output layer. Hidden layers store calculation results. The equation for determining the gate activation function is formulated as,

$$h_{\mathrm{T}} = \zeta\left(U\left[\mathrm{M}_{\mathrm{T}}, \mathrm{N}_{\mathrm{T}-1}, J_{\mathrm{T}-1}\right] + j_h\right) \tag{12}$$

The input is represented as M_{T}, the output of the last block is represented as $\mathrm{N}_{\mathrm{T}-1}$ and the memory of the terminal block of LSTM is described as $J_{\mathrm{T}-1}$. The weight vector is denoting as U for each input with the bias vector as j_h and the logistic sigmoid function is given as ζ.

4. Results& Discussion

This section provides the suggested Chiropteran Mahi optimization-based BiLSTM classifier with effectual implementation and performance evaluation. The system's accuracy, sensitivity, specificity are the considered performance metrics. By contrasting the suggested approach with current ones, taking into account the measurements with the epoch, or validating preliminary report, the effectiveness of the suggested approach is examined. The developed Chiropteran Mahi optimization-based BiLSTM classifier is employed using the EWalk (Emotion Walk) dataset in the Python tool, implemented in Windows 10 OS with 8GB RAM. The proposed method efficiently identifies the emotions of people while walking, using three performance measures: accuracy, sensitivity, and specificity.

4.1 Preliminary Results

In this section, the Initial outcomes of the suggestedCMO-based BiLSTM classifier are described. The conductance of the suggestedCMO-based BiLSTM classifier's various features is revealed in figure 3. The input image from the EWalk dataset with the emotion of sad, angry, and happy are provided as input. The preprocessed picture is the next stage of the provided input picture & then the image skeletonization occurs for removing the textual characteristics of the picture for the accurate prediction of emotions. The textual features such as the LTP, LGP, and LOOP are considered for emotion identification in the suggested system.

Input Image Sad, Angry, and Happy			
Pre-processed Image Sad, Angry, and Happy			
Image Skeletonization Sad, Angry, and Happy			

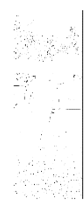
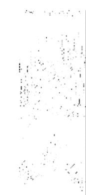
LTP features Sad, Angry, and Happy			
LGP features Sad, Angry, and Happy			
LOOP features Sad, Angry, and Happy			

Figure 3. Prelimiary Results of Textual Features

4.2 Performance Evaluations

The conductance of the proposed CMO-based BiLSTM classifier is estimated using the EWalk dataset for emotion identification while walking. The evaluation of the proposed method is primarily based on the K-fold and Training percentage techniques, using performance metrics such as Accuracy, Sensitivity, and Specificity.

i) Evaluation depend on K-fold value

Figure 4 indicates the effectiveness of the constructed CMO-depend BiLSTM classifier and evaluates the suggested approach by taking into account the different epochs in means of the K-fold.Diagram 4 a) demonstrates the accuracy dependon the K-fold value with the various epochs, ensuring the effectiveness of the suggested approach. Initially, the developed CMO-based BiLSTM classifier accuracy is 89.777 % for epoch 20 with a K-fold value of 2. When the K-fold value is increased with the increasing number of epochs up to 20 to 100, the accuracy of the suggested technique is highly improved. The accuracy is increased to 94.720 % with the K-fold value of 10 for the final epoch 100. Similarly, figure 4 b) shows the sensitivity depend on the K-fold value with the various epochs, showing that the developed Chiropteran Mahi optimization-based BiLSTM classifier sensitivity is 97.079 % for epoch 60 with a K-fold value of 6. When the K-fold value is increased with the increasing number of epochs up to 20 to 100, the sensitivity of the developed method is highly improved. The sensitivity is increased to 98.964 % with the K-fold value of 10 for the final epoch 100, as shown in figure 4 b). Diagram 4 c) demonstrates the specificity based on the K-fold value with the various epochs, ensuring the effectiveness of the suggested approach.Initially, the developed CMO-based BiLSTM classifier specificity is 90.399

% for epoch 80 with a K-fold value of 8. When the K-fold value is increased with the increasing number of epochs up to 20 to 100, the accuracy of the suggested technique is highly improved. The accuracy is increased to 91.423 % with the K-fold value of 10 for the final epoch 100.

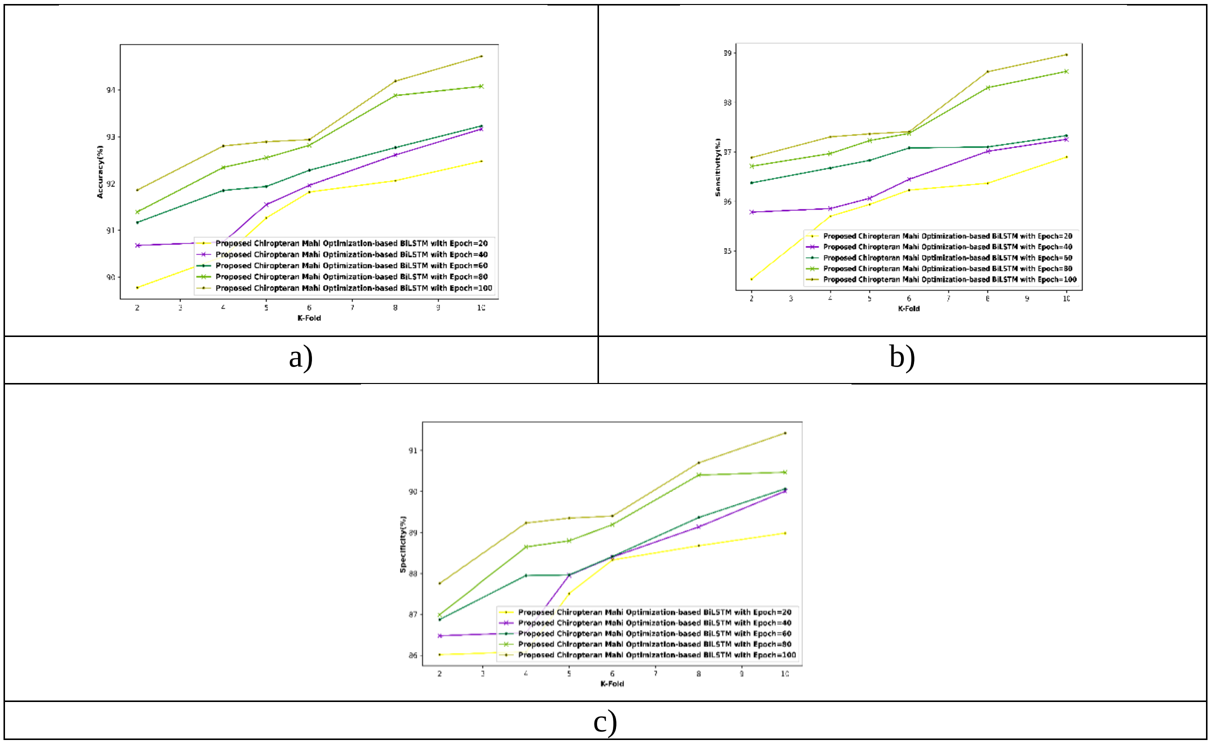


Figure 4. Performance Evaluation depend on Epoch a) Accuracy b) Sensitivity, c) Specificity

ii) Evaluation depend on Training Percentage

The efficacy of the developed Chiropteran Mahi optimization-based BiLSTM classifier is presented in figure 5. By taking into account the different epochs in terms of training percentage, the suggested method is examined.The accuracy, sensitivity, specificity depend on the training percentage value with the various epoch are demonstrated in figure 5 a), b), and c), respectively, for ensuring the efficiency of the proposed method. Initially, the developed Chiropteran Mahi optimization-based BiLSTM classifier accuracy, sensitivity, specificity are 61.152%, 87.270%, and 89.596%, respectively, for epoch 20 with a training percentage value of 2. When the training percentage value is increased with the increasing number of epochs up to 20 to 100, the accuracy, sensitivity, specificity of the developed approach are highly improved. The accuracy is increased to 89.949% with the training percentage value of 10 for the final epoch 100, and the sensitivity and specificity are increased to 90.235% and 90.563%, respectively, with the training percentage value of 10 for the final epoch 100.

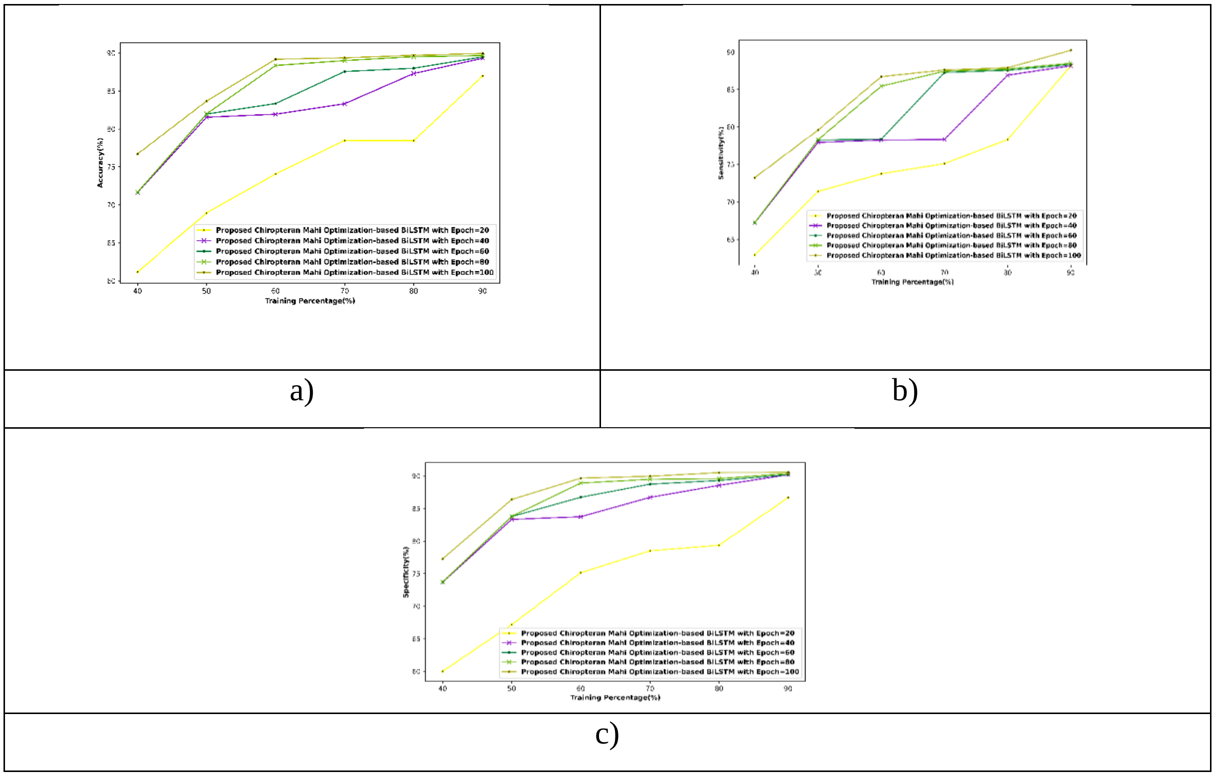


Figure 5. Performance Evaluation depend on Epoch a) Accuracy b) Sensitivity, c) Specificity

4.6 Comparative Methods

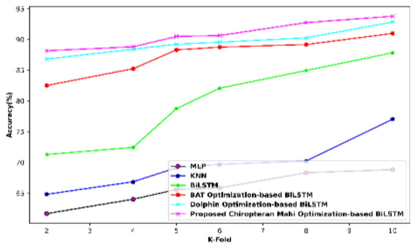
The proposed Chiropteran Mahi optimization-based BiLSTM classifier is related with the existing methods like MLP [18], KNN [19], BiLSTM [20], BAT optimization-based BiLSTM, Dolphin optimization-based BiLSTM.

4.6.1 Modified Discussion

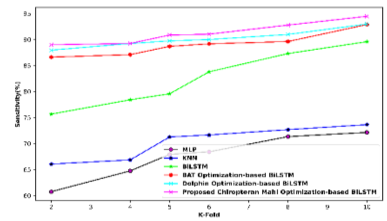
The developed Chiropteran Mahi optimization-based BiLSTM classifier is related to the existing methods for the accurate emotion recognition of people while walking from the EWalk dataset. The K-fold value & the device's training % are used to evaluate how well the suggested scheme performs.

i) Comparison based on K-Fold

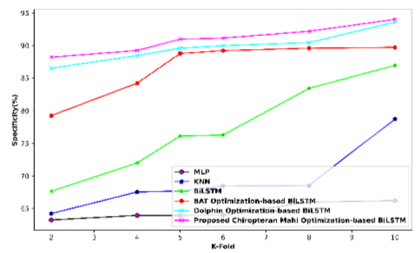
The efficacy of the proposed Chiropteran Mahi optimization-based BiLSTM classifier is demonstrated in Figure 6 with the K-fold value. Various existing methods, including MLP, KNN, BiLSTM, BAT optimization-based BiLSTM, and Dolphin optimization-based BiLSTM, are available for accurately identifying emotions while walking. Figure 6a shows that the developed Chiropteran Mahi optimization-based BiLSTM classifier achieves an accuracy improvement of 12.951% when compared to the BiLSTM classifier with a K-fold value of 5. Diagram 6b demonstrates that the developed classifier achieves a sensitivity improvement of 2.049% when compared to the BAT optimization-based BiLSTM classifier with a K-fold value of 6. Furthermore, Diagram6c reveals that the developed classifier achieves a specificity improvement of 1.899% when compared to the Dolphin optimization-based BiLSTM classifier with a K-fold value of 8.



a)



b)

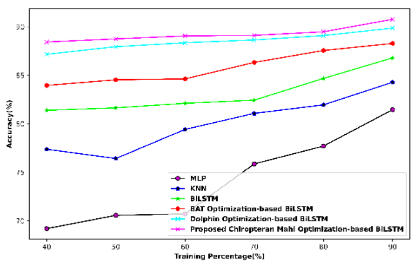


c)

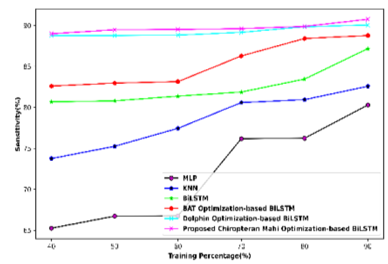
Figure 6. Comparability Evaluation depend on K-fold Value with a) Accuracy b) Sensitivity, c) Specificity

ii) Comparison based on Training Percentage

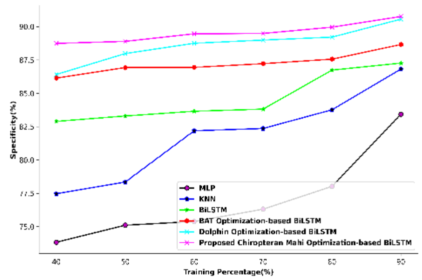
This paragraph presents the performance evalaution of the suggested Chiropteran Mahi optimization-focused BiLSTM classifier for identifying emotions while walking. The existing methods including MLP, KNN, BiLSTM, BAT optimization-based BiLSTM, and Dolphin optimization-based BiLSTM are also discussed. The accuracy, sensitivity, specificity of the suggested classifier are compared with these existing methods depend on the training percentage value, which is demonstrated in Figure 7. The suggested approachobtains an accuracy improvement of 7.793% compared to BiLSTM, a sensitivity improvement of 3.718% compared to BAT optimization-based BiLSTM, and a specificity improvement of 0.831% compared to Dolphin optimization-based BiLSTM.



a)



b)



c)

Figure 7. Comparability Analysis basedon Training Percentage with a) Accuracy b) Sensitivity c) Specificity

4.7 Modified Discussion

By evaluating the K-fold value and the trained percent of the provided EWalk samples with assessment criteria like accuracy, sensitivity, or specificity, the efficacy of the developed system is confirmed. The improvement in accuracy, sensitivity, and specificity are revealed in comparative table 1 for the EWalk dataset.

Table 1. Comparative Discussion for the Proposed and Current Approachdepend on Training Percentage & the K-fold Value

Methods	EWalk Dataset					
	K-fold 10			Training percentage 90		
	Accuracy %	Sensitivity %	Specificity %	Accuracy %	Sensitivity %	Specificity %
MLP	68.838	72.150	66.215	81.456	80.296	83.432
K-Nearest Neighbor	77.038	73.639	78.734	84.278	82.582	86.817
BiLSTM	87.839	89.595	86.961	86.789	87.179	87.268
Bat optimization-based BiLSTM	90.981	92.885	89.725	88.278	88.779	88.660
Dolphin optimization-based BiLSTM	92.832	92.985	93.607	89.858	90.053	90.562
Proposed Chiropteran Mahi	**93.770**	**94.467**	**94.011**	**90.763**	**90.763**	**90.763**

optimization-based BiLSTM						

5. Conclusion

The focus of this research paper is on demonstrating emotions, namely Sad, Angry, and Happy, that are expressed through body movements. The Deep BiLSTM system, which was built via CMO, successfully recognizes human emotions when they are presented. Gaits, movies, or emotional responses that have been tagged as the first input for the suggested emotion recognizing method. Before feeding the data into the BiLSTM classifier, the most relevant features are extracted and combined. The Chiropteran Mahi optimization method is proposed to improve the performance by optimizing the hyperparameters of the classifier to better recognize emotions from walking data. To identify human emotions from walking data, the most important step is skeletonization, which separates features based on the shape of the object and organizes body movements based on the order of human joint locations. The resulting features are grid-based. The production of the suggested technique is well increased in means of the accuracy as 12.951 % based on the training percentage and 7.793 % depend on the K-Fold value when compared to the existing methods. Research progress will improve emotional expression utilizing walking data by combining several models with further optimization methods.

References

[1] Mona Algarni , Faisal Saeed , Tawfik , Al-Hadhrami , Fahad Ghabban , Mohammed Al-Sarem,“Deep Learning-Based Approach for Emotion Recognition Using Electroencephalography (EEG) Signals Using Bi-Directional Long Short-Term Memory (Bi-LSTM),” national centre of Biotechnology information,8,2022.

[2] Zihao Ye, Tao Zuo, Waner Chen, Yingxin Li & Ziyu Lu, “Textual emotion recognition method based on ALBERT-BiLSTM model and SVM-NB classification,” Soft Computing Volume 27, pp. 5063–5075 ,2023.

[3] Zheng Liu, Xin Kang & Fuji Ren, “Improving Speech Emotion Recognition by Fusing Pre-trained and Acoustic Features Using Transformer and BiLSTM,” *Intelligent Information Processing* , pp 348–357,2022.

[4] Om Prakash Yadav , Laxmi Prasad Bastola , Jagdish Sharma , “Speech Emotion Recognition using Convolutional Recurrent Neural Network,” Proceedings of 10th IOE Graduate Conference, Volume 10, 2021.

[5] Mustaqeem, Sajjad, M., & Kwon, S., “Clustering-Based Speech Emotion Recognition by Incorporating Learned Features and Deep BiLSTM,” *IEEE Access*, 8, 79861–79875.

[6] Fachang Cui ,ORCID,Ruqing Wang ,Weiwei Ding,Yao Chen andLiya Huang, “A Novel DE-CNN-BiLSTM Multi-Fusion Model for EEG Emotion Recognition,” *IEEE*,2022.

[7] Chen, C., & Li, Q., “A Multimodal Music Emotion Classification Method Based on Multifeature Combined Network Classifier.,” *Mathematical Problems in Engineering*, 1–11,2020.

[8] T. B. Alakus, I. Turkoglu, “Emotion recognition with deep learning using GAMEEMO data set,” *IET*,2020.

[9] Fabian Benitez-Quiroz, C., Srinivasan, R., and Martinez, A.M.,"Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in CVPR, 2016.

[10] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," arXiv preprint arXiv:1801.07481, 2018.

[11] J. Bruna, W. Zaremba, A. Szlam and Y. Lecun, "Spectral networks and locally connected networks on graphs,"*In proceedings of 2nd International Conference on Learning Representations, ICLR*, 2014.

[12] C. Si, W. Chen, W. Wang, L. Wang and T. Tan, "An attention enhanced graph convolutional LSTM network for skeleton-based action recognition,"*In proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Computer Vision Foundation, IEEE*, pp.1227–1236, 2019.

[13] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng and H. Li, "T-GCN: a temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, pp.1–11, 2019.

[14] R. Ying, R. He, K. Chen, P. Eksombatchai, W.L. Hamilton and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," *In proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD, pp. 974–983, 2018.

[15] Amir Hossein Gandomi, Xin-She Yang, Amir Hossein Alavi, SiamakTalatahari, "Bat algorithm for constrained optimization tasks," *Neural Computing and Applications*, vol.22, no.6, pp.1239-1255, 2013.

[16] Kaveh, A., and Farhoudi, N., "A new optimization method: Dolphin echolocation," Advances in Engineering Software, vol.59, pp.53-70, 2013.

[17] EWalk (Identifying Emotions from Walking using Affective and Deep Features) dataset http://gamma.cs.unc.edu/GAIT/

[18]Palo, H.K., Mohanty, M.N. and Chandra, M., 2015. Use of different features for emotion recognition using MLP network. In Computational Vision and Robotics (pp. 7-15). Springer, New Delhi.

[19] Rieger, S.A., Muraleedharan, R. and Ramachandran, R.P., 2014, September. Speech based emotion recognition using spectral feature extraction and an ensemble of kNN classifiers. In The 9th International Symposium on Chinese Spoken Language Processing, IEEE, pp. 589-593.

[20] Sajjad, M. and Kwon, S., 2020. Clustering-based Speech Emotion Recognition by Incorporating Learned Features and Deep BiLSTM. IEEE Access, 8, pp.79861-79875.