JOURNAL LA MULTIAPP

VOL. 04, ISSUE 01 (001-010), 2023
DOI: 10.37899/journallamultiapp.v4i1.773

# Movements Recognition in the Human Body Based on Deep Learning Strategies

**Muthana S. Mahdi**[1]

[1]*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*

**Corresponding Author: Muthana S. Mahdi*
*Email:* muthanasalih@uomustansiriyah.edu.iq

Article Info

*Article history:*
Received 21 February 2023
Received in revised form 23 June 2023
Accepted 12 July 2023

*Keywords:*
Emotion Recognition,
Body Movement Recognition,
Convolutional Neural Network,
Non-Verbal Connection,
Human Interaction with The Computer

**Abstract**

*These days, the study of human body movements for the purpose of emotion identification is an absolutely necessary component of social communication. Several different contexts call for the implementation of non-verbal communication strategies such as gestures, eye movements, facial expressions, and body language. Among them, emotion detection based on body movements. It can also identify the emotions of a person even if they are too far away from the camera. Other studies have shown that body language can express emotional states more effectively than words can. In this research study, an emotional state is determined by the human motion of the entire body. The architecture of a deep convolution neural network is used, and multiple parameter settings are considered. Both the University of York's emotion dataset, which includes 15 different kinds of emotions, and dataset of GEMEP corpus, which includes five emotions, can be used to assess the proposed system. The results of the experiments demonstrated that the proposed system has a higher degree of recognition accuracy.*

## Introduction

Emotional intelligence, which could be fostered via bettering human-computer interactions and human communication, is one area where emotion recognition could prove useful (Shirbhate & Talele, 2016). The way one moves about says more than words ever could. New findings from the field of experimental psychology highlight the fundamental role that feelings play in both decision making and logical analysis. Humans use a wide range of emotional expressions in their day-to-day interactions with one another. Both verbal and nonverbal cues play important roles in human interaction. The term "non-verbal communication" refers to the exchange of information or cues without the use of words.

Body language (kinesics) and outward appearance are two examples (Gavrilescu, 2015) of such visual clues. Body language and posture can reveal a person's emotional state. There is information sent by posture that cannot be conveyed with words or expressions alone. Human posture, for instance, can be used to infer a person's mental state from a great distance. To deduce human emotion from non-verbal communication, it is sufficient to record body language (Elfaramawy et al., 2017). When happy, angry, or surprised, the body typically turns toward the source of those emotions; when scared, the body contracts; when joyful, the arms open wide and speed up; when afraid or sad, the body turns away; (Asaju & Vadapalli, 2022) to name just a few examples. Recent research has demonstrated that we are adept at reading the emotional nuances in others' nonverbal cues and drawing conclusions about their mental states.

Gestures are a collection of specific physical movements. All you need to do is use your brain, hands, and arms to complete the task. Taken as a whole, these cues reveal both the emotional climate and the nature of the interactions taking place. Identifying emotions from body language has numerous uses, and this work is supported by psychological research. Alarm system behaviour anomaly detection Automatic emotion recognition through bodily cues has several potential uses, including in the areas of self-aware computing, human-computer interaction, healthcare, and the support of people with autism (Niewiadomski et al., 2015). The purpose of this work is to deduce human emotions from the ways in which people's bodies move in response to various stimuli.

The need for this study was inspired by the fact that a blurry face vision can occur in a surveillance setting if the camera is too far from the subject. Capturing head, hand, leg, and torso movements for emotion recognition is a promising approach to addressing this type of problem.

This study's contribution is a feedforward deep convolution neural network (FDCNN) that uses deep convolutional features to infer a person's emotional state based on their motions.

What follows is the text of the aforementioned document. The works that are relevant to this discussion are briefly in Section Two. The proposed work is explained in Section three. In Section four, we present the findings of our experiments. At last, in conclusions, some final thoughts are offered, and plans for the future are outlined.

## Related Works

The characteristics of a gesture's dynamic are proposed for an emotion recognition scheme, and supervised learning techniques used to evaluate them (Arunnehru & Kalaiselvi Geetha, 2017). Within the domain of body motion, a method for defining high-level parameters using the discrete elements of a convolutional auto encoder has been created (Holden et al., 2016). Using STIP characteristics, a method is suggested for recognizing the affective state of a human (Gunes et al., 2015). The results on analysis of sentiment will be discussed, along with the practical uses of deep learning (Zhang et al., 2018; Brock, 2018). Using the backpropagation technique, a deep learning algorithm that works with enormous data sets has been constructed (LeCun et al., 2015). A neuronal architecture capable of self-organization that was designed in order to discern emotional states based on full-body motion patterns (Elfaramawy et al., 2017). For the purpose of recognizing emotions from video, a model that combines CNNs and RNNs has been presented (Khorrami et al., 2016).

Deep learning models built with DCNN have been designed specifically for the multimodal emoFBVP database. A model for the recognition of non-verbal emotions that uses a hierarchical representation of the features was tested, and the results showed a considerable improvement in accuracy (Ranganathan et al., 2017). The new method of an artificially intelligent scheme using promising neural network topologies is offered for the purpose of emotion recognition (Tamhane et al., 2022). The design of the method used for emotion identification is a CNN-RNN network, which allows it to achieve superior results to those obtained using other methods (Ebrahimi Kahou et al., 2015) A set of emotional body gestures has been designed to distinguish across cultures as well as between genders in order to create a foundation for the automatic recognition of emotional body gestures (Noroozi et al., 2018).

## The Proposed Scheme

This stage will be divided into two parts as follows:

### Network of Convolutional Neural

Convolutional layer: The four stages of convolution that make up the convolutional layer as follow: First, you need an image and a detector for features. second, you'll want to multiply

each of the image's pixels by its associated feature pixel. Third, Discover the total amount. Fourth, the sum must be divided by the feature's total pixel count. Which can be determined by the following formula:

$$C(x_{u,v}) = \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{i=\frac{m}{2}}^{\frac{m}{2}} f_k(i,j) x_{u-i,v-j} \qquad (1)$$

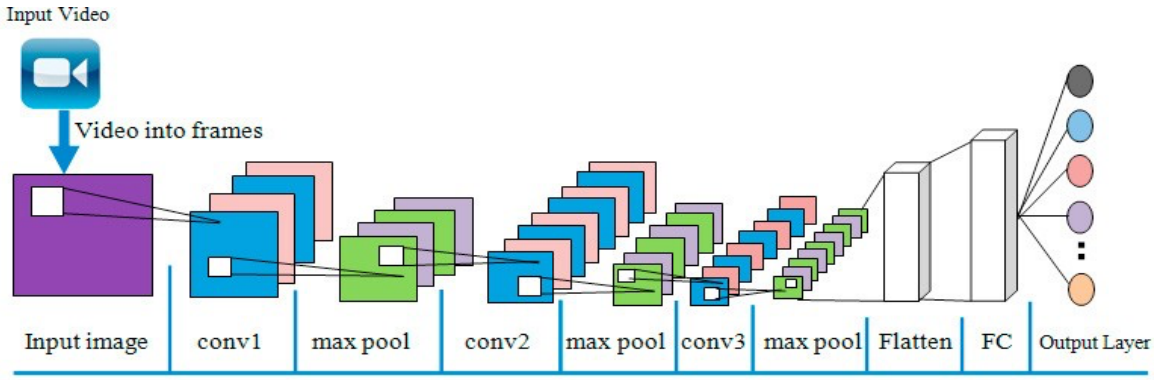



*Figure 1.* Model of Convolutional Neural Network.

Where, x is input image, n × m is kernel size, and fk is a filter.

Pooling layers: Maps can be made smaller via pooling or subsampling layers. The pooling function is implemented by the following four procedures; (1) The size of window is determined; (2) The step movement is chosen; (3) The filtered images are shown when the window is shifted; (4) Each window's maximum value is extracted. Equation 2 can be used to determine this value:

$$M(x_i) = Max\left\{x_{i+k,i+l} \middle| |k| \leq \frac{m}{2}, |l| \leq \frac{n}{2}\ k, k \epsilon N\right\} \qquad (2)$$

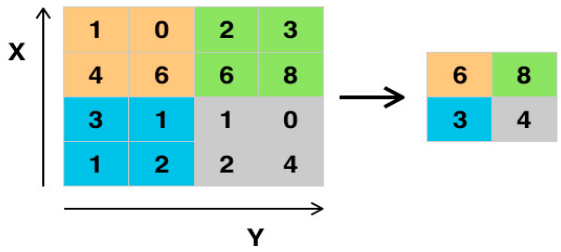



*Figure 2.* Max-Pooling Example

A Unit of rectified Linear: The activation function when the outcome is zero, and the entry is less than zero. It is computed as follows:

$$R(x) = \max(0, x) \qquad (3)$$

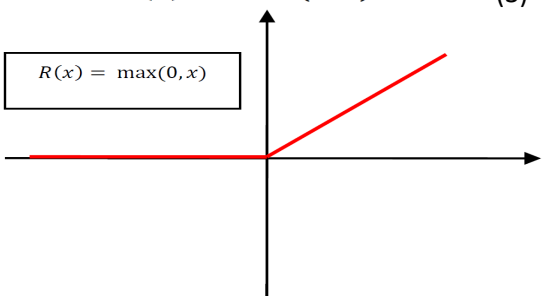



*Figure 3.* Max-Pooling Example

Fully connected layer: all of the neurons in this layer communicate with all of the neurons in the preceding layer; this is the same as in neural networks. You may figure it out by using the formula:

$$F(x) = \sigma(W * x) \quad (4)$$

Softmax layer: this is the layer where backpropagation can take place. Through error backpropagation, networks improve their efficiency. Consider an input vector of size N. where S(x): R →[0, 1]N. It is computed by:

$$S(x)_j = \frac{x^{xi}}{\sum_{i=0}^{N} e^{xi}} \quad (5)$$

so, N >= j >= 1

Output layer: the number of classes determines the depth of this layer. It is a representation of the image's input class.

$$C(x) = \{ i | \exists i \, \forall j \neq i : x_j \leq x_i \} \quad (6)$$

**Feed forward Deep Convolutional Neural Network (FDCNN)**

Frames are extracted from the input videos and stored in two distinct directories, one for the training set and the other for the validation set. First-layer input is now the raw photos themselves. The FDCNN model is depicted in Figure 4. The input picture is 150 pixels wide by 150 pixels high by three bytes in size (where three is the number of color channels). The filters, or weights, in this network have a size of 3x3 across all layers. Sliding, also known as convolving, refers to the process of multiplying the original pixel value by a set of weight values. These additions are added to generate a single value known as the receptive field. A unique integer is generated by each receptive field. Acquire the Feature Map with Scale at Last (150x150x3). There are 32 stacked feature maps and 32 filters used in the first layer. The spatial resolution of the representation is shrunk by the subsampling layer (75x75x32). There are 64 filters used in the second layer, and those filters are overlaid with 64 feature maps. As a result, the layer of max-pooling is decreasing the dimension of feature to (37 x 37 x 64). In this case, the max-pooling layer's output is a dimensionality reduction of the features to (18x18x128). Each of the maximum pooling layers has a 2x2 grid in its center. As a final step, 512 hidden units are placed in completely connected layers, and 15 neurons are allocated per class to display the anticipated emotions.

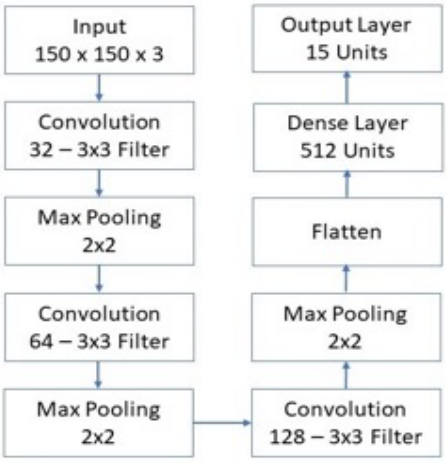



*Figure 4. Model Architecture*

This stage will be divided into four sections as follows:

**The Dataset of Emotion**

This dataset features 29 actors displaying a range of emotions and behaviors (jumping, sitting, strolling, etc.) across five categories: happiness, anger, sadness, distrust, and fear. The

resolution of the recorded video is 1920 pixels wide by 1080 pixels in height, and the frame rate is 25 frames per second (Santhoshkumar et al., 2017).

**The Dataset of GEMEP**

The Geneva Multimodal Emotion Portrayals (GEMEP) are a compilation of many types of media used to convey feelings. Ten actors expressing five core emotions—anger, joy, fear, sadness, and pride—were employed in this production. Every one of the 720p 576 videos has 25 frames per second (Santhoshkumar & Geetha, 2019; Marrero Fernandez et al., 2019).

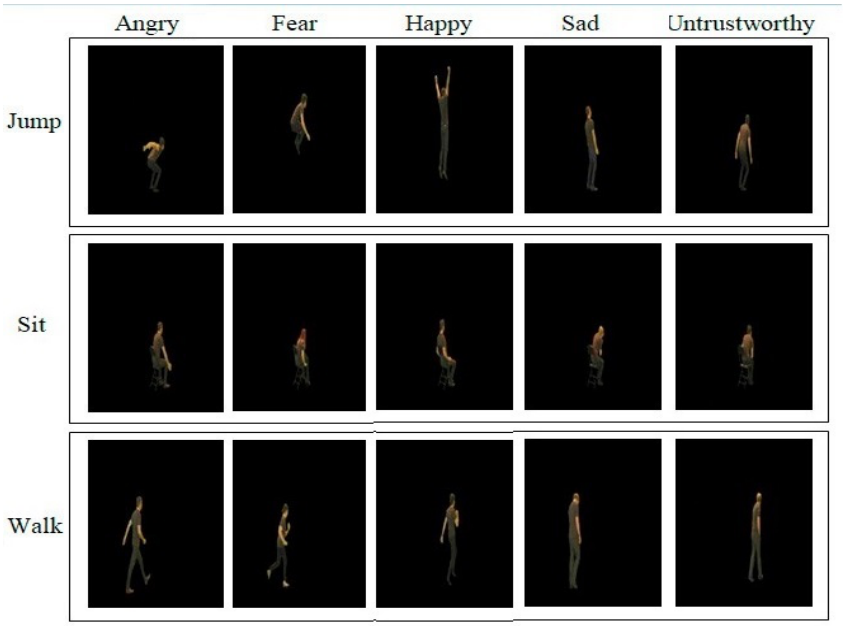
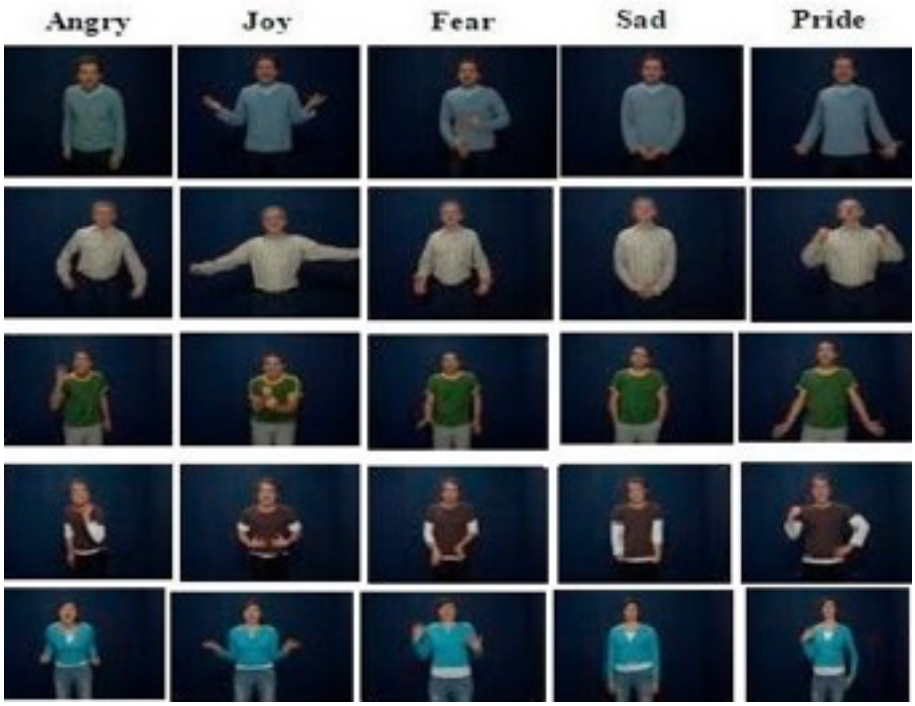



*Figure 5. Example frame of three activities and Five emotions*

The experiments were run on a Windows 10 computer running the Jupyter notebook environments and Anaconda Python, with an Intel core i7 processor. A 64-batch-size and 10 training iterations are used to prepare the dataset.



**Criteria for Evaluating Performance**

The confusion matrix is a two-dimensional table in which the rows represent predictions and the columns provide actual classifications (Sati, et al., 2021).

For native patterns, TP (true positives) is the total count of those that were correctly identified as native. False negatives (FN) represent the number of times native patterns were wrongly labeled as exotic. Number of non-native patterns that were wrongly identified as native (FP). Number of foreign patterns that were accurately identified as being alien; also known as "true negatives" (TN) (Acharjya et al., 2017).

| | | True Set | |
|---|---|---|---|
| | | Native | Foreign |
| Predicted set | Native | TP | FP |
| | Foreign | FN | TN |

*Figure 6. Confusion matrix*

Accuracy, Recall, F-Score, Specificity, and Precision are all useful metrics for gauging the quality of a proposal's performance. The proportion of times a model gets a forecast right is the accuracy, and it may be computed with Eq. 7.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{FP}+\text{TN}} = \frac{\text{TP}+\text{TN}}{\text{P}+\text{N}}$$

The accuracy with which an unusual emotion is recalled is revealed via recall.

$$Recall = \frac{tp}{tp+fn} \quad (8)$$

The F-score is the median of the two measures of accuracy used to evaluate symphonies.

$$F - Score = 2\frac{Precision\ X\ Recall}{Precision+Recall} \quad (9)$$

The degree to which a technique is able to effectively identify negative emotion is measured by its specificity.

$$Specificity = \frac{tn}{tn+fp} \quad (10)$$

Finally, the fraction of correct categorization is displayed in Precision and can be determined mathematically by (LeCun et al., 2015).

$$Precision = \frac{tp}{tp+fp} \quad (11)$$

## Results and Discussion

The Percentage of Confusion Matrix in the Emotion Dataset is shown in Table (1) and Table (2) respectively. Where the relationship is described between the angry jump, the happy jump, and the untrustworthy jump indicating that these images are somewhat similar. This scheme performed well in predicting walking and sitting (sad and happy). Which indicates that learning this class of emotions is optimal compared to other emotions.

Table 1. Percentage of Confusion Matrix in the Data set of Emotion (part1)

| | AS | AJ | AW | HS | FS | FJ | FW | HJ |
|---|---|---|---|---|---|---|---|---|
| AS | 96 | 0 | 0 | 2 | 1 | 0 | 0 | 0 |
| AJ | 0 | 95 | 0 | 0 | 0 | 1 | 0 | 2 |
| AW | 0 | 0 | 96 | 0 | 0 | 0 | 1 | 0 |
| HS | 1 | 0 | 0 | 97 | 1 | 0 | 0 | 0 |
| FS | 1 | 0 | 0 | 1 | 95 | 0 | 0 | 0 |
| FJ | 0 | 2 | 0 | 0 | 0 | 94 | 0 | 1 |
| FW | 0 | 0 | 3 | 0 | 0 | 0 | 95 | 0 |
| HJ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 96 |
| HW | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| SJ | 0 | 4 | 0 | 0 | 0 | 1 | 0 | 1 |
| SS | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| SW | 0 | 0 | 4 | 0 | 0 | 0 | 2 | 0 |
| UJ | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 3 |
| US | 3 | 0 | 0 | 1 | 2 | 0 | 0 | 0 |
| UW | 0 | 0 | 2 | 0 | 0 | 0 | 5 | 0 |

Note that AJ is Angry Jump, AS is Angry Sit, AW is Angry Walk, FJ is Fear Jump, FS is Fear Sit, FW is Fear Walk, HJ is Happy Jump, HS is Happy Sit, HW is Happy Walk, SJ is Sad Jump, SS is Sad Sit, SW is Sad Walk, UJ is Untrust worthy Jump, us is Untrust worthy Sit, UW is Untrust worthy Walk.

Table 2. Percentage of Confusion Matrix in the Emotion Dataset (part2)

| | UJ | SJ | HW | UW | SS | US | SW |
|---|---|---|---|---|---|---|---|
| AJ | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| AS | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| AW | 0 | 0 | 1 | 2 | 0 | 0 | 0 |
| FJ | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| FS | 0 | 0 | 0 | 0 | 2 | 1 | 0 |
| FW | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| HJ | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| HS | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| UJ | 91 | 4 | 0 | 0 | 0 | 0 | 0 |
| SJ | 0 | 94 | 0 | 0 | 0 | 0 | 0 |
| HW | 0 | 0 | 94 | 2 | 0 | 0 | 3 |
| UW | 0 | 0 | 1 | 91 | 0 | 0 | 1 |
| SS | 0 | 0 | 0 | 0 | 95 | 1 | 0 |
| US | 0 | 0 | 0 | 0 | 2 | 92 | 0 |
| SW | 0 | 0 | 0 | 2 | 0 | 0 | 92 |

Table 3. Performance measures for proposed work on dataset of emotion

| | F-Measure | TP Rate | FP Rate | Recall | Precession |
|---|---|---|---|---|---|
| AW | 0.95 | 0.92 | 0.02 | 0.92 | 0.97 |
| FJ | 0.93 | 0.90 | 0.03 | 0.90 | 0.97 |
| HJ | 0.95 | 0.94 | 0.03 | 0.94 | 0.97 |
| HS | 0.93 | 0.93 | 0.03 | 0.90 | 0.97 |
| FS | 0.96 | 0.92 | 0.02 | 0.94 | 0.97 |
| FW | 0.92 | 0.89 | 0.03 | 0.88 | 0.97 |

| | | | | | |
|---|---|---|---|---|---|
| UJ | 0.89 | 0.85 | 0.02 | 0.82 | 0.98 |
| US | 0.92 | 0.93 | 0.02 | 0.87 | 0.98 |
| HW | 0.95 | 0.95 | 0.03 | 0.93 | 0.97 |
| SJ | 0.93 | 0.85 | 0.02 | 0.88 | 0.98 |
| UW | 0.94 | 0.94 | 0.03 | 0.91 | 0.97 |
| AJ | 0.89 | 0.83 | 0.02 | 0.83 | 0.97 |
| AS | 0.94 | 0.89 | 0.01 | 0.90 | 0.99 |
| SS | 0.93 | 0.89 | 0.03 | 0.89 | 0.97 |
| SW | 0.95 | 0.93 | 0.02 | 0.93 | 0.98 |

Table 4. Percentage of Confusion Matrix in the Data set of GEMEP

| | **Fear** | **Joy** | **Pride** | **Sad** | **Angry** |
|---|---|---|---|---|---|
| Fear | 95 | 1 | 2 | 1 | 1 |
| Joy | 1 | 96 | 1 | 1 | 1 |
| Pride | 1 | 0 | 97 | 1 | 1 |
| Sad | 1 | 2 | 0 | 96 | 1 |
| Angry | 1 | 2 | 1 | 1 | 95 |

The link between joy and anger is described in table 3, which indicates that the two feelings are somewhat closer to one another than they are to each other. The performance of this network is better than other emotion predictors at predicting fear, sadness, and pride.

Table 5. Performance measures for proposed work on dataset of GEMEP.

| | **F-Measure** | **TP Rate** | **FP Rate** | **Recall** | **Precession** |
|---|---|---|---|---|---|
| Fear | 0.94 | 0.92 | 0.03 | 0.91 | 0.97 |
| Pride | 0.94 | 0.91 | 0.03 | 0.91 | 0.97 |
| Sad | 0.97 | 0.97 | 0.03 | 0.96 | 0.97 |
| Joy | 0.93 | 0.89 | 0.02 | 0.89 | 0.98 |
| Angry | 0.89 | 0.83 | 0.02 | 0.82 | 0.97 |

Table 6. Compared between the several related works with the proposed work.

| **Papers** | **Accuracy [%]** |
|---|---|
| Gavrilescu [2] | 86.4% |
| Radoslaw [5] | 73.0% |
| Nourhan [3] | 90.7% |
| The proposed work | 95.4% |

## Conclusions

In this research, we present the FDCNN model for emotion prediction from video sequences of human body movements. Deep convolutional feature representations are used in this model to learn about saliency at different levels. Both the GEMEP dataset and the Emotion dataset from the University of York are used to test the effectiveness of the strategy that has been proposed. In comparison to the baseline models, the performance of this scheme is significantly improved. The work that will be done in the future seeks to produce research apps that can identify the feelings of children who have autism spectrum condition (ASD).

## Acknowledgment

The authors thank the Department of Computer Science, College of Science, Mustansiriyah University, for supporting this work.

## References


Acharjya, D. P., Geetha, M. K., & Sanyal, S. (Eds.). (2017). Internet of Things: novel advances and envisioned applications.

Arunnehru, J., & Kalaiselvi Geetha, M. (2017). Automatic human emotion recognition in surveillance video. *Intelligent techniques in signal processing for multimedia security*, 321-342.

Asaju, C., & Vadapalli, H. (2022, January). A temporal approach to facial emotion expression recognition. In Artificial Intelligence Research: Second Southern African Conference, SACAIR 2021, Durban, South Africa, December 6–10, 2021, Proceedings (pp. 274-286). Cham: Springer International Publishing.

Barros, P., Jirak, D., Weber, C., & Wermter, S. (2015). Multimodal emotional state recognition using sequence-dependent deep hierarchical features. *Neural Networks, 72*, 140-151.

Brock, H. (2018, February). Deep learning—Accelerating Next Generation Performance Analysis Systems?. In Proceedings (Vol. 2, No. 6, p. 303). MDPI.

Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015, November). Recurrent neural networks for emotion recognition in video. *In Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 467-474).

Elfaramawy, N., Barros, P., Parisi, G. I., & Wermter, S. (2017, October). Emotion recognition from body expressions with a neural network architecture. *In Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 143-149).

Elfaramawy, N., Barros, P., Parisi, G. I., & Wermter, S. (2017, October). Emotion recognition from body expressions with a neural network architecture. *In Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 143-149).

Gavrilescu, M. (2015, November). Recognizing emotions from videos by studying facial expressions, body postures and hand gestures. In 2015 23rd Telecommunications Forum Telfor (TELFOR) (pp. 720-723). IEEE.

Gunes, H., Shan, C., Chen, S., & Tian, Y. (2015). *Bodily expression for automatic affect recognition*. Emotion recognition: A pattern analysis approach, 343-377.

Holden, D., Saito, J., & Komura, T. (2016). A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG),* 35(4), 1-11.

Khorrami, P., Le Paine, T., Brady, K., Dagli, C., & Huang, T. S. (2016, September). How deep neural networks can improve emotion recognition on video data. *In 2016 IEEE international conference on image processing (ICIP)* (pp. 619-623). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

Marrero Fernandez, P. D., Guerrero Pena, F. A., Ren, T., & Cunha, A. (2019). Feratt: Facial expression recognition with attention net. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Niewiadomski, R., Mancini, M., Varni, G., Volpe, G., & Camurri, A. (2015). Automated laughter detection from full-body movements. *IEEE Transactions on Human-Machine Systems, 46*(1), 113-123.

Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing, 12*(2), 505-523.

Ranganathan, H., Venkateswara, H., Chakraborty, S., & Panchanathan, S. (2017, September). Deep active learning for image classification. *In 2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3934-3938). IEEE.

Santhoshkumar, R., & Geetha, M. K. (2019). Deep learning approach for emotion recognition from human body movements with feedforward deep convolution neural networks. *Procedia Computer Science*, 152, 158-165.

Santhoshkumar, R., Geetha, M. K., & Arunnehru, J. (2017). SVM–KNN based Emotion Recognition of Human in Video using HOG Feature and KLT Tracking Algorithm. *International Journal of Pure and Applied Mathematics, 117*(15), 621-634.

Sati, V., Sánchez, S. M., Shoeibi, N., Arora, A., & Corchado, J. M. (2021). Face detection and recognition, face emotion recognition through NVIDIA Jetson Nano. *In Ambient Intelligence–Software and Applications: 11th International Symposium on Ambient Intelligence* (pp. 177-185). Springer International Publishing.

Shirbhate, N., & Talele, K. (2016, December). Human body language understanding for action detection using geometric features. *In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 603-607). IEEE.

Tamhane, S., Shrirao, A., Shah, M., & Patil, D. (2022). Emotion Recognition Using Deep Convolutional Neural Networks. Available at SSRN 4096405.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1253.