

Research on Emotion Recognition Based on Deep Learning for Mental Health

Xianglan Peng

School of Humanities, Henan Mechanical and Electrical Vocation College, Zhengzhou, Henan 451191, China

E-mail: pe79193@163.com

Keywords: deep learning, artificial intelligence, facial expression, emotion recognition

Received: January 27, 2021

This paper briefly introduced the support vector machine (SVM) based and convolutional neural network (CNN) based healthy emotion recognition method, then improved the traditional CNN by introducing Long Short Term Memory (LSTM), and finally carried out simulation experiments on three emotion recognition models, the SVM, traditional CNN, and improved CNN models, in the self-built face database. The results showed that the CNN model converged faster in training and had a smaller error when it was stable after introducing LSTM; compared with the SVM and traditional CNN models, the improved CNN had a higher recognition accuracy for facial expressions; the time consumed by the improved CNN model was the shortest in both training and testing stages.

Povzetek: Analiziranih je bilo več metod strojnega učenja, tudi globoke mreže, za iskanje čustev v povezavi z mentalnim zdravjem.

1 Introduction

With the progress of science and technology and the improvement of computer performance, artificial intelligence appeared and has been widely used in mechanical operation fields, such as translation, image recognition, and classification, which are not difficult but highly repetitive [1]. The ultimate goal of artificial intelligence is to achieve good human-computer interaction, thus replacing humans to carry out dangerous or repetitive work. However, in the current development of artificial intelligence, although it has been able to realize the recognition and classification of objects, such as images and audio, in human-computer interaction, the perception of human emotions by artificial intelligence is still at a low level [2]. Artificial intelligence needs better emotion recognition ability to achieve better human-computer interaction services [3]. Also, human beings express their emotions in various forms, including actions, language, physiological signals, and facial expressions. These emotions usually reflect their psychological state, especially physiological signals and facial expressions. People's physiological state will directly affect the psychological state, and the psychological state will react to the physiological state. Changes in physiological signals will reflect changes in physiological state, thus indirectly reflecting the psychological state [4]. Facial expression can directly reflect people's emotions, and the changes of emotions also reflect the state of mental health. However, the monitoring of physiological signals needs quite professional equipment, and the collection process is complex, which may delay the judgment of people's mental health. Changes in facial expression are relatively easy to collect as long as a good camera is configured to collect mental health-related images. When mental health is judged by the emotion reflected by the facial expression, the manual observation needs rich clinical experience and

has low efficiency. Artificial intelligence has a fast computing speed, and it can extract relevant feature rules from face images more effectively and then judge whether people's emotions are in a healthy state. Atkinson et al. [5] proposed a feature-based emotion recognition model based on an electroencephalogram, which combined the mutual information-based feature selection method with kernel classifier to improve the accuracy of emotion classification tasks. The experimental results verified the effectiveness of the proposed method. Kaya et al. [6] proposed to replace the Deep Neural Network (DNN) and support vector machine (SVM) with the extreme learning machine (ELM) in audio and visual emotion recognition. The results showed that the method could achieve better accuracy in emotion classification in audios and videos. Shojaeilangari et al. [7] proposed a new pose invariant dynamic descriptor to encode the relative motion information of facial landmarks. The results showed that the method could deal with speed changes and continuous head pose changes to realize fast emotion recognition. This paper briefly introduced the emotion recognition method based on SVM and convolutional neural network (CNN), then improved the traditional CNN by introducing Long Short Term Memory (LSTM), and finally carried out simulation experiments on three emotion recognition models, the SVM, traditional CNN, and improved CNN models, in the ORT human face database and self-built face database.

2 Recognition of mental health emotion based on deep learning

Unless specially controlled, the expression of ordinary people is usually rich, and the emotion of the other person can be confirmed by observing the change of expression [8]. Artificial intelligence is difficult to understand the emotions represented by different expressions in images

taken by cameras and judge the mental health level represented by the emotions; therefore, artificial intelligence needs relevant algorithms to improve the experience of human-computer interaction and the accuracy of artificial intelligence in judging the user's mental health.

2.1 Traditional recognition method based on SVM

At present, artificial intelligence needs to recognize emotions through machine learning, and SVM is one of the traditional machine learning methods [9]. The basic principle of SVM for health emotion recognition is to find a hyperplane for space division in the vector space of expression features. The expression on one side of the hyperplane is classified as a healthy emotion, and the other side is classified as one kind of unhealthy emotion. In short, SVM is a classification algorithm, which classifies the expression images collected by cameras to identify whether the emotion is in a healthy psychological state. Since the expressions collected by cameras are generally image data, it is necessary to extract features of expression images to obtain expression features when using SVM for recognition [10]. There are various methods for extracting image features. In this paper, facial expression features are extracted by the LDP (local directional pattern) algorithm. The principle of the LDP algorithm is directional edge statistics. It is assumed that x is a pixel in an image. The gray value of a 3×3 field that centers on pixel x is convoluted with Kirsch template [11] M to obtain the corresponding edge response, $|m_i|$. Then, the edge responses are sorted according to their gradients. The first k edge responses are marked as code 1, and the rest is marked as code 0. The calculation formula of LDP code [12] is as follows:

$$\begin{cases} m_k = kth(M) \\ M = |m_0, m_1, \dots, m_7| \\ LDP_k(r, c) = \sum_{i=0}^7 b_i(m_i - m_k) \times 2^i, (1) \\ b_i(m_i - m_k) = \begin{cases} 1 & m_i - m_k \geq 0 \\ 0 & m_i - m_k < 0 \end{cases} \end{cases}$$

where m_k is the k -th edge response, M is Kirsch template, $LDP_R(r, c)$ is the LDP code of central point c , and r is the domain radius, which is set as 3 in this paper. The extraction steps are as follows: ① eight Kirsch templates and equation (1) are combined to convert each pixel in the original face image into LDP code; ② the LDP code image of the human face is constructed according to the LDP code; ③ the LDP code image is divided into $a \times b$ blocks, the histogram of all the blocks is extracted; ④ the histogram of the blocks is connected end to end to get the final feature vector.

After obtaining the expression feature vector, it can be used as a training sample to train SVM to obtain the decision function of SVM. The calculation formula is:

$$\begin{cases} f(x) = sgn(\sum_{i=1}^l a_i y_i K(x_i, x_j) + b) \\ \sum_{i=1}^l a_i y_i = 0 \quad 0 \leq a_i \leq C \end{cases}, (2)$$

where a is the set of a_i , a_i is the Lagrangian coefficient [13], l is the sample size, $K(x_i, x_j)$ is the kernel function, C is the penalty parameter, y_i is the result of classification, and x_i is the sample data.

2.2 Healthy emotion recognition method based on LSTM-CNN

In addition to SVM, neural network, a kind of deep learning algorithm, has also widely used in artificial intelligence. Neural network realizes machine learning by imitating neural cells of the human brain, which is relatively better in learning effect. CNN is one of the neural networks [14]. Compared with other kinds of neural networks, CNN is more suitable for image recognition. The basic structure of CNN is the input layer, convolution layer, pooling layer, and output layer. The convolution layer and pooling layer are the hidden layers of the neural network, and the number of them depends on the operation requirements. The more the number is, the better the learning effect is, but the lower the efficiency is.

The basic process of emotion recognition by CNN is as follows. An image is inputted into the input layer after preprocessing and then convoluted through convolution kernels in the convolution layer. The image features are extracted. The convoluted image is processed by pooling in the pooling layer (equivalent to compressing the image, including mean-pooling and max-pooling). After repetitive convolution and pooling operations, the results are output in the form of full connection in the output layer according to the transfer formula. The results will be compared with the expected results; if they are not consistent, the weights and bias terms in the hidden layer will be adjusted reversely, and the weights and bias terms in CNN will be adjusted through repetitive training to make the output as close to the expected output as possible.

One of the advantages of CNN in image recognition is that it does not need feature extraction. The convolution operation in the convolution layer plays the role of feature extraction. Moreover, the hidden layer of CNN has an activation function operation, which can transform the linear input data into nonlinear to fit the hidden rules between features better; therefore, it can classify more accurately than the hyperplane in SVM [15].

Although CNN can effectively identify the emotion in expression images, in practical applications, when artificial intelligence recognizes the images taken by cameras, not all of the images are taken under good lighting and from proper angles, and most of the images have incomplete facial expression features. Moreover, the facial expression features have multi-dimensional and multi-scale changes, making it difficult to improve the recognition rate. This study introduces LSTM [16] into CNN to improve its recognition rate of expression and emotion.

The main structure of LSTM includes the input gate, forget gate, and output gate. Parameters to be calculated are input into the input gate, mainly including the current input of cell (x_t), the state of the last hidden layer (h_{t-1}), and the last state of cell (C_{t-1}). A matrix is constructed by these parameters and corresponding weights to determine the number of new information in the cell. The relevant formula is:

$$\begin{cases} i_t = g(\omega_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(\omega_C \cdot [h_{t-1}, x_t] + b_C), \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \end{cases}, \quad (3)$$

where i_t is the proportion of the new information that can be memorized, \tilde{C}_t is the cell state of the new information added, C_t is the current cell state after the addition of the new information, ω_i and ω_C are the corresponding weights, and b_i and b_C are the corresponding offsets.

The forget gate determines the number of original information to be abandoned, and its formula is:

$$f_t = g(\omega_f \cdot [h_{t-1}, x_t] + b_f), \quad (4)$$

where f_t is the proportion of information that is not forgotten in C_{t-1} , ω_f is the corresponding weight, and b_f is the corresponding offsets.

The output gate is a structure that obtains the output result based on the parameters of the first two structures. The output result can be the final result or the hidden variable when the content is updated next time. The formula is as follows:

$$\begin{cases} o_t = g(\omega_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases}, \quad (5)$$

where o_t is the weight that determines the final output information quantity and h_t is the final output or the next hidden state.

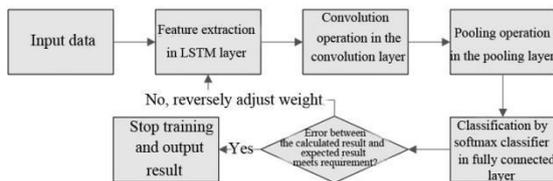


Figure 1: The expression and emotion recognition process based on LSTM-CNN.

After introducing LSTM into CNN, it can effectively associate the changes of expression before and after to obtain the regular features of expression changes. Moreover, the continuously varying features that can reflect emotions in the human face can be more prominent, thus reducing the influence of irrelevant background features and reflect the emotions contained in continuously changing expressions. The training flow of

the expression and emotion recognition model based on the LSTM and CNN is shown in Figure 1.

① The data were input, and the relevant parameters were initialized, including convolution kernel, weights in structure layers, offset, etc.

② In the LSTM layer, features were extracted from the image according to equations (3), (4), and (5). The feature map that was needed by the subsequent convolution was constructed according to the extracted h_t .

③ The feature map processed by the LSTM layer was input into the convolution layer for convolution operation by the convolution kernel. The convolution formula is:

$$x_j^l = f(\sum_{i \in M} x_i^{l-1} \cdot W_{ij}^l + b_j^l), \quad (6)$$

where x_j^l is the output feature map after the activation of the j -th convolution kernel in the l -th convolution layer, x_i^{l-1} is the feature output of the i -th convolution kernel in the last convolution layer after pooling, W_{ij}^l is the weight parameter between the i -th convolution kernel and the j -th convolution kernel, b_j^l is the offset of j convolution kernels of l layers, M is the number of convolution kernels in the l -th convolution layer, and $f(\bullet)$ is the activation function.

④ The convoluted feature map was input into the pooling layer for pooling. The pooling operation included mean-pooling and max-pooling. In this study, the max-pooling operation was adopted. The target box slid on the feature map for some distance, and the largest pixel in the target box was taken as the compression result of the target box.

⑤ The convolution and pooling operations mentioned above were performed many times, depending on the number of convolution layers and pooling layers. After convolution and pooling, the result was output to the fully connected layer. Then the expression images were classified by using a softmax classifier in the fully connected layer.

⑥ The recognition results of CNN were compared with the expected results (the recognition results refer to the results obtained by calculating the input image layer by layer with CNN, and the expected results refer to the corresponding result label of the training sample), and the weights and offset parameters in the calculation formula were adjusted reversely according to the error until the error was within the predetermined range or converged to stability. The calculation formula of error is as follows:

$$E = -\sum_{k=1}^n t_k \log(y_k), \quad (7)$$

where E stands for the error between the calculated output vector and the actual output vector, n is the number of output layer nodes, y_k is the probability of belonging to such kind of label output to the output layer after the forward calculation of the fully connected layer, and t_k is the label of the actual correct solution that is set.

⑦ When the error converged to stability or is within a predetermined range, the training of the recognition model ended, and then the model was tested using the testing set.

3 Simulation experiment

3.1 Experimental environment

The CNN model was simulated and analyzed using MATLAB software [17]. The experiment was carried out on a laboratory server. The server configurations were the Windows7 system, I7 processor, and 16 G memory.

3.2 Experimental data

In this study, a self-built facial expression database was used. Facial expression images came from 100 students randomly selected from Henan Mechanical and Electrical Vocation College after explaining the use of face images to them and obtaining their approval. Since the purpose of this study was to realize the recognition of mental health emotion of human expression fast by artificial intelligence, when collecting the facial expression image data of the volunteers, the corresponding mental health test was carried out, and the corresponding mental health labels were added for facial expression images. To ensure the time correspondence between the expression image and the degree of mental health in the database (i.e., the mental state reflected by the expression image was indeed the psychological state when the expression was collected), the expression data were collected by making the psychological evaluation of the volunteers to judge the mental health status and capturing the volunteers' expressions synchronously during the psychological evaluation [18]. Finally, 30 facial expression images were collected from each volunteer. The results of the psychological evaluation were statistically analyzed, and it was found that 68 volunteers had healthy psychology (2040 expression images), 26 volunteers had sub-healthy psychology (780 expression images), and six volunteers had poor mental psychology (180 expression images). As the number of people with healthy psychology was the largest, followed by people with sub-healthy psychology and people with poor mental psychology, the number of images collected for three mental health states was unbalanced, which would affect the final training result; therefore, the expression images were extended through means such as rotation, extension, and mirroring. The number of images for sub-healthy psychology and poor mental psychology was extended to 2040. The external performance of three kinds of mental health states was described briefly. The volunteers with healthy psychology were relaxed when receiving the psychological counseling test. They smiled unconsciously in the process of communication and showed a bright smile when the communication was smooth. The volunteers with sub-healthy psychology were not relaxed in the process of mental health assessment, but most of them were not tense in facial expression. In the communication process, the expression was relatively flat, and the communication is relatively smooth. Most of the smiles appeared when they

talked about the topic of interest. The volunteers with poor mental health were usually tight in facial expression. Although they achieved communication, they gave a sense of tension and anxiety, and some had sweating. Moreover, the atmosphere presented by the dialogue in the process of communication was relatively repressive. When testing the three recognition models, 20 expression images of each volunteer in the database were used as the training set, and the remaining ten images were used as the testing set.

In the simulation test, 60% of the images were taken from every mental health status as the training set, and the remaining 40% was taken as the test set. There were 1224 images in the training set and 816 images in the test set.

3.3 Experimental setup

In this study, the expression recognition model was improved by introducing LSTM to CNN. The structural parameters of CNN are as follows. There were three convolutional layers. Every convolutional layer had 64 convolution kernels in a size of 5×5 . Relu function was used as the activation function. There were three pooling layers. In every pooling layer, the size of the pooling box was 2×2 , and the moving step length of the pooling box was 2. The size of the image in the input layer was 150×100 . In the LSTM layer, there were 64 hidden neurons, weights were initialized using gloriot_normal, and the offset was set as 0.

Moreover, to verify the effectiveness and excellence of the improved expression recognition model, it was compared with the SVM model and the traditional CNN model. The comparative experiment was carried out in the same face database. The parameters of the traditional CNN model were consistent with the CNN in the LSTM-CNN model. SVM adopted the sigmoid kernel function, and the penalty parameter was set as 1.

3.4 Experimental results

The SVM model fits the decision function according to the extracted feature vector in training, thus to obtain the hyperplane in the feature vector space for the classification of different expressions; therefore, different from the CNN model, the SVM model needed to be trained repeatedly. Figure 2 shows the convergence curves of the traditional CNN model and the improved CNN model in training. It was seen from Figure 2 that the training error of the two CNN models gradually decreased in the process of training iteration and finally stabilized at a low level. The comparison of the curves showed that the improved CNN model converged to stability faster than the traditional CNN model: the traditional CNN model converged to stability after about 250 times of iterations, and the improved CNN converged to stability about 150 times of iterations; the error of the improved CNN model after convergence to stability was significantly smaller than that of the traditional CNN model.

After training the SVM, traditional CNN, and improved CNN models with the training set in the self-built database, the trained models were tested using the

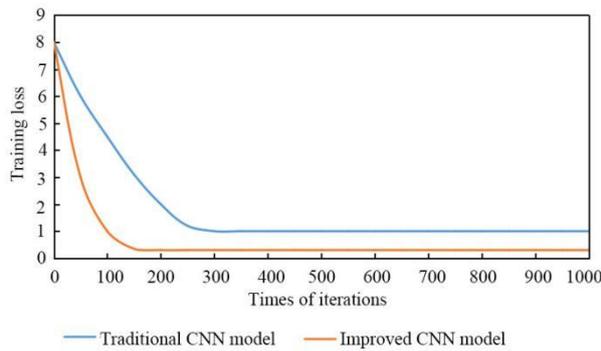


Figure 2: The convergence curves of the traditional and improved CNN models in training.

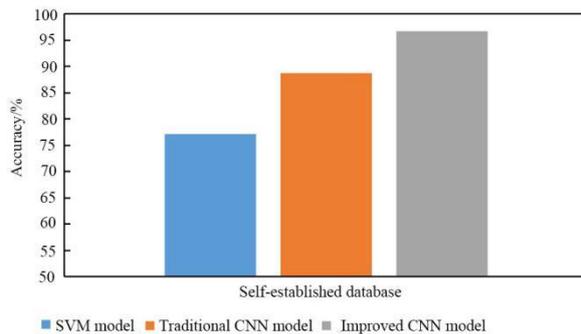


Figure 3: The recognition accuracy of three healthy emotion recognition models in two kinds of face databases.

corresponding testing set, and the results are shown in Figure 3. In the self-built database, the recognition accuracy of the SVM model was 77.1%, that of the traditional CNN model was 88.6%, and that of the improved CNN model was 96.6%. In the same database, the SVM model had the lowest accuracy, that of the traditional CNN model was the second, and that of the improved CNN model was the highest.

For the artificial intelligence that was used for judging emotions, in addition to the high accuracy for emotional judgment, the speed of judgment is also very important. Table 1 shows the time spent in training and testing three healthy emotion recognition models. In the training stage, the training time of the SVM model was 20.2 min, that of the traditional CNN model was 20.4 min, and that of the improved CNN model was 15.3 min; in the testing stage, the SVM model took 835 ms, the traditional CNN model took 621 ms, and the improved CNN model took 378 ms. In the training stage, the SVM model could not perform a parallel operation on the data in the training set but gradually fit it; thus, it needed a long training time. Although the other two CNN models could perform a parallel calculation on the data in the training set, they needed repeated training and gradual adjustment when adjusting the internal weight; thus, they also needed a long time. However, the improved CNN model eliminated the background features that would produce interference from the image as much as possible and highlighted the expression features; therefore, it converged faster and spent less time. In the testing stage, the three models have been trained, and the results could be calculated step by

	The SVM model	The traditional CNN model	The improved CNN model
Training time	20.2 min	20.4 min	15.3 min
Testing time	835 ms	621 ms	378 ms

Table 1: Time consumption of three recognition models in training and testing.

step as long as the data were input; thus, the time consumed was much shorter than that in the training stage.

4 Discussion

For the human body, health includes not only physical health but also mental health. However, different from the physiological health state, it is difficult to see the mental health state intuitively. If the physical health status can be directly obtained through various detection instruments, such as blood state, body temperature, etc., then the mental health status needs to be gradually judged by professionals in the process of communication. It not only requires a high professional quality of the tester but also consumes a lot of time. Artificial intelligence has the advantages of fast learning and high work efficiency. Based on the progress of machine vision technology, artificial intelligence has been gradually applied to the judgment of people’s emotions. Artificial intelligence combined with machine vision technology can judge mental health through the emotion reflected by the change characteristics of human facial expressions. This paper briefly described two intelligent algorithms, the image LDP features-based SVM algorithm and the LSTM-introduced CNN algorithm. Then, 100 student volunteers were taken to establish the database of facial emotional and mental health. The performance of the SVM, traditional CNN, and improved CNN recognition models was compared. The final experimental results showed that the improved CNN model could identify the mental health state behind the expression more accurately than the other two models, and the training and testing time was shorter. On the one hand, compared with the SVM model, the CNN model did not need to extract image features deliberately as its convolution operation has obtained features; on the other hand, the improved CNN model could associate the images before and after to further extract the core features from the changing expression, which reduced the interference of background features and improved the recognition efficiency and accuracy.

The psychological evaluation on volunteers was carried out by professionals to ensure the accurate correspondence between expression and mental state, and facial expression changes were captured in time in the process of psychological assessment. As a thank to the volunteers, after the psychological assessment, professionals provided guidance and suggestions on the mental health of volunteers according to the evaluation results. After summing up the results of the psychological evaluation, it was concluded that most of the volunteers

had healthy psychology, some volunteers had sub-healthy psychology, and fewer volunteers had unhealthy psychology. The advice to the volunteers with healthy psychology was to keep the current good mood. The reason for the sub-healthy state was mostly related to the heavy academic pressure and the chaotic daily schedule. The advice to the volunteers with a sub-healthy state was to adjust work and rest, be relaxed in the face of study, and attempt to formulate a study schedule. Besides the heavy academic pressure, the reasons for the unhealthy mental state of the volunteers also included introversion, inferiority, and little communication with others. The final suggestion for the volunteers with an unhealthy mental state was to set a good daily routine, walking outside, and starting communication with acquaintances first.

5 Conclusion

This paper briefly introduced the SVM-based and CNN-based healthy emotion recognition methods, then improved the traditional CNN by introducing LSTM, and finally carried out simulation experiments on the SVM, traditional CNN, and improved CNN models through the self-built human face database. The results are as follows: (1) compared with the traditional CNN model, the improved CNN model converged faster and had a smaller error after stabilization; (2) the recognition accuracy of the improved CNN model was the highest, followed by the traditional CNN model and SVM model; (3) the improved CNN model took the least time in the training stage and the shortest time in the testing stage.

References

- [1] Jenke R, Peer A, Buss M (2017). Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing*, 5, pp. 327-339. <https://doi.org/10.1109/TAFFC.2014.2339834>.
- [2] Anagnostopoulos C N, Iliou T, Giannoukos I (2015). Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, 43, pp. 155-177. <https://doi.org/10.1007/s10462-012-9368-5>.
- [3] Suja P, Tripathi S (2015). Analysis of emotion recognition from facial expressions using spatial and transform domain methods. *International Journal of Advanced Intelligence Paradigms*, 7, pp. 57. <https://doi.org/10.1504/IJAIP.2015.070349>.
- [4] Atkinson J, Campos D (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications*, 47, pp. 35-41.
- [5] Chen ZB (2018). Facial Expression Recognition Based on Local Features and Monogenic Binary Coding. *Informatica*, 43, pp. 117-121. <https://doi.org/10.31449/inf.v43i1.2716>.
- [6] Kaya H, Salah A A (2016). Combining modality-specific extreme learning machines for emotion recognition in the wild. *Journal on Multimodal User Interfaces*, 10, pp. 139-149.
- [7] Shojaeilangari S, Yau W Y, Teoh E K (2016). Pose-invariant descriptor for facial emotion recognition. *Machine Vision & Applications*, 27, pp. 1063-1070.
- [8] Liu S, Tong J, Meng J, Yang J, Zhao X, He F, Qi H, Ming D (2018). Study on an effective cross-stimulus emotion recognition model using EEGs based on feature selection and support vector machine. *International Journal of Machine Learning and Cybernetics*, 9, pp. 721-726. <https://doi.org/10.1007/s13042-016-0601-4>.
- [9] Ghimire D, Lee J (2016). Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, 13, pp. 7714-7734. <https://doi.org/10.3390/s130607714>.
- [10] Viet SD, Bao CLT (2018). Effective Deep Multi-source Multi-task Learning Frameworks for Smile Detection, Emotion Recognition and Gender Classification. *Informatica*, 42, pp. 345-356. <https://doi.org/10.31449/inf.v42i3.2301>.
- [11] Chakraborty S, Singh S K, Chakraborty P (2017). Local directional gradient pattern: a local descriptor for face recognition. *Multimedia Tools & Applications*, 76, pp. 1201-1216.
- [12] Chakraborty S, Singh S K, Chakraborty P (2018). Correction to: Local directional gradient pattern: a local descriptor for face recognition. *Multimedia Tools & Applications*, pp. 1-1.
- [13] Pan H, Xie L, Lv Z, Li J, Wang Z (2020). Hierarchical support vector machine for facial micro-expression recognition. *Multimedia Tools & Applications*, 79, pp. 1-15. <https://doi.org/10.1007/s11042-020-09475-4>.
- [14] Lopes A T, Aguiar E D, Souza A F D, Oliveira-Santos T (2017). Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order. *Pattern Recognition*, 61, pp. 610-628. <https://doi.org/10.1016/j.patcog.2016.07.026>.
- [15] Gjoreski M, Gjoreski H, Kulakov A (2014). Machine Learning Approach for Emotion Recognition in Speech. *Informatica*, 38, pp. 377-384.
- [16] Ghimire S, Deo R C, Raj N, Mi J (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253, pp. 113541.1-113541.20. <https://doi.org/10.1016/j.apenergy.2019.113541>.
- [17] Wang F, Wu S, Zhang W, Xu Z, Zhang Y, Wu C, Coleman S (2020). Emotion recognition with convolutional neural network and EEG-based EFDMs. *Neuropsychologia*, 146, pp. 107506. <https://doi.org/10.1016/j.neuropsychologia.2020.107506>.
- [18] Azam I, Khan SA (2018). Feature Extraction Trends for Intelligent Facial Expression Recognition: A Survey. *Informatica*, 42, pp. 507-514. <https://doi.org/10.31449/inf.v42i4.2037>.