International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 28, No. 1, February, 2023, pp. 169-174
Journal homepage: http://iieta.org/journals/isi

Facial Emotion Recognition Using HOG and Convolution Neural Network

Raghad Ghalib Abd*, Abdul-Wahab Sami Ibrahim, Ameen A. Noor

Computer Science Department, College of Education, University of Almustansirya, Baghdad 10045, Iraq

Corresponding Author Email: Raghadghalib@uomustansiriyah.edu.iq

https://doi.org/10.18280/isi.280118

Received: 26 October 2022
Accepted: 20 December 2022

Keywords:
HOG, CNN, median filter, facial emotion recognition, face detection

ABSTRACT

Due to numerous difficulties, including the variation in face shapes between individuals, the challenge of recognizing dynamic facial attributes, the poor quality of digital images, etc., detecting human emotion depending on facial expression is difficult for the computer vision community. Thus, in this study, we propose an approach for emotion recognition depending on facial expression using histogram of oriented gradients and convolution neural network (HOG-CNN). The HOG-CNN composed of three stages, median filter, HOG, and CNN. The first stage is preprocessing using median filter. The second stage is feature extraction using HOG. The third stage is classification using CNN. The proposed method was tested and evaluated on the UMD face database. The system attained a high performance with a mean average accuracy of 98.07%, average precision of 94.78%, and average recall of 97.15%.

1. INTRODUCTION

Emotions play a significant part in daily life and directly impact human decisions, attention, reasoning, quality of life, and prosperity. Emotions and facial expressions help individuals to communicate with one another. The development of human-computer interaction (HCI) has become more essential due to the effect of computers on human life and the automation of people's daily lives. The interaction between computers and humans is highly interesting [1]. Emotion substantially influences human cognitive processes, including perception, attention, learning, memory, reasoning, and problem-solving. Emotion has a powerful influence on attention, especially modulating the selectivity of attention and motivating behavior and action. This executive and attentional control is intimately related to the process of learning, as intrinsically limited attentional abilities are better concentrated on relevant information. Emotion also aids in the information retrieval and efficient encoding [2, 3].

A common challenge in the study of computer vision is the ability to identify emotion from a face picture. Image sequence-based and still image-based techniques can be used to classify it. The performance of the image sequence-based method is often better than a still image-based method when it comes to improving recognition performance by extracting valuable temporal characteristics from the image sequences. The spatiotemporal characteristic might be based on geometric characteristics as well as appearance [4, 5].

One of the essential tasks in security, computer vision, education, telecommunications, and psychiatry is automatic emotion recognition. Facial expression recognition has been one of the fastest-growing fields due to its applications in biometrics, emotion analysis, and image retrieval. There has been a great deal of research into facial expression recognition by addressing issues with recognizing facial expressions in various gestures, illuminations, and other situations [6].

Deep neural network attained significantly good result in modelling complex patterns. In this work, a deep learning based method for detection emotion of human is introduced. Proposed method utilizes median filter to enhance the image at preprocessing step, uses the HOG for extraction of features, and then apply convolution neural network for classification (CNN). The experimental results demonstrate that the extracted attributes enhance the training accuracy and speed. The proposed method was tested and evaluated on the UMD face database. The system attained a high performance with a mean average accuracy of 98.07%, average precision of 94.78%, and average recall of 97.15%.

The research is structured as follows: Section 2 discuss the related work, methodology is illustrated in section 3, proposed method is presented in section 4, results are illustrated in section 5, and finally conclusion in section 6.

2. RELATED WORK

Emotion recognition and face detection were and still a research topic that requires improvement. Numerous works have been suggested in the literature to perform this task. Particularly those research's that rely on deep learning methods that raise the limits of the traditional handcrafted methods and boost the state-of-the-art. Clawson et al [7], studied hierarchical and single modeling approaches, using regional inputs for CNN learning, and investigated the effects of data augmentation and image preprocessing. By applying the CNN to learn facial regions independently and combine their output with an SVM meta layer. The model experimented on the CK+ dataset and achieved an accuracy of 93.3%.

Kwolek [8] accomplished facial recognition by extracting characteristics from the Gabor filter. Research demonstrated that adding a Gabor filter increases CNN accuracy from 79 percent to 87.5 percent Wu et al. [9] Claimed that facial emotion recognition might be accomplished using Gabor motion energy filters (GME). GME filter and Gabor energy (GE) filter were compared in this procedure, and the results revealed that the suggested method was 7% better.

Dagher et al. [10], proposed a three stage Support Vector Machine (SVM) in which the first stage consists of twenty-one SVM, which are binary combination of seven emotions The first stage is sufficient if one motion is dominating. If two are predominant, then the 2nd stage is utilized, and, if three are predominant, then the 3rd stage is utilized. At processing step the input image converted to gray and resized, viola jones detection is applied, and border adjustment and cropping. The experiment is conducted on CK+ and JAFFE datasets. The model attained an accuracy of 93.29% on CK+ database and 96.71% on JAFFE database.

Said and Barr [11], presented the face-sensitive convolutional neural network (FS-CNN) to identify human emotions. On large-scale pictures, the suggested FS-CNN is utilized to identify faces. Next, facial landmarks are examined to forecast expressions for emotion identification. In the first step, faces in high-resolution photos are found and cropped for further processing. In the next step, a CNN was utilized to analyze scale invariance on pyramid images and forecast facial emotion based on landmark analytics. The UMD faces database was utilized to test and the suggested FS-CNN. High performance was attained with a mean average accuracy of nearly 95%.

3. METHODOLOGY

3.1 UMD faces dataset

The dataset of UMD faces [12], which was obtained from the internet and was gathered utilizing well-known search engines like google, yahoo, Bing etc. There are 367,888 RGB photos in this collection. It was noted for gender identification, facial detection, and keypoint localization. An overview of the images with annotations provided by the UMD face database is shown in Figure 1.

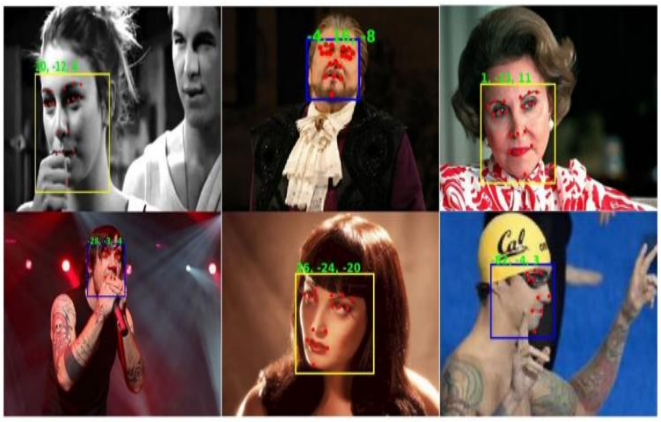

Figure 1. Images with annotations from the UMD face database

3.2 Median filter

The FER system's performance could be affected by the noise that frequently distorts facial images. An image or signal's noise can be removed using the median filter (MF), a non-linear digital filtering method. Such a pre-processing procedure to reduce noise and enhances the outcomes of processing [13].

3.3 Histogram of oriented gradient

HOG is an attribute descriptor that analyzes how gradients are distributed throughout an image [14]. Two properties determine the histogram. The gradient's amplitude and direction are as follows. The magnitude specifies how much to populate, and the direction specifies which bin to populate. Because computing the gradient around an object's corners and edges causes a significant reaction, this is helpful for object recognition. This may define the location and angle of lips and the angle around the eyebrows and face for facial expressions. The properties of each expression may be identified using this descriptor [14]. Figure 2, where the HOG descriptor is applied to a facial image displaying the angry expression, shows the gradient orientations [15, 16].

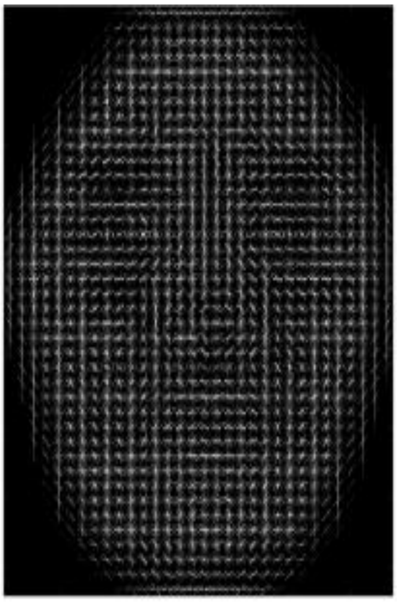

Figure 2. Visualization of the HOG descriptor on facial image [15]

HOG is a feature descriptor that was first suggested for detecting pedestrians. It keeps track of how frequently gradient orientation appears in a detection window. The following is a summary of the primary steps in computing HOG features [17]:

(1) Gradient computation. The spatial gradients in the horizontal and vertical axes are computed in this stage. The gradient magnitudes and angles are then calculated using these two gradients.

(2) Orientation binning. The picture is split into small, interconnected regions known as cells in this stage. Based on the gradient angle, the gradient magnitude of every pixel in a cell is divided into several orientation bins.

(3) Feature description. Blocks of neighboring cells are created in this stage. The L2-norm is used to normalize each block. A descriptor is produced in a detection window by merging the normalized block histograms.

3.4 Convolution neural network

Pooling, convolution, and fully linked layers are some of the structural components of CNN architecture. After a stack of multiple convolution layers and a pooling layer are repeated, one or more fully linked layers are added. Forward propagation is the procedure of converting inputs data into output using these layers (Figure 3) [18].

In an experiment conducted in 1962, Hubel and Wiesel [19] demonstrated that four distinct neuronal cells in the brain only reacted when edges of a specific orientation were present. They discovered that the neurons were organized into a columnar design and that those neurons could create visual perception when they worked together. CNN provides an output layer with a vector of highly distinct attributes connected to the pre-classified category after accepting input data such as images, video, and audio with an optional dimension as the input layer. CNN employed labeled training data, a collection of training instances, as a form of supervised learning. CNN examines the training data and derives a technique for mapping new instances, just like other supervised learning algorithms.

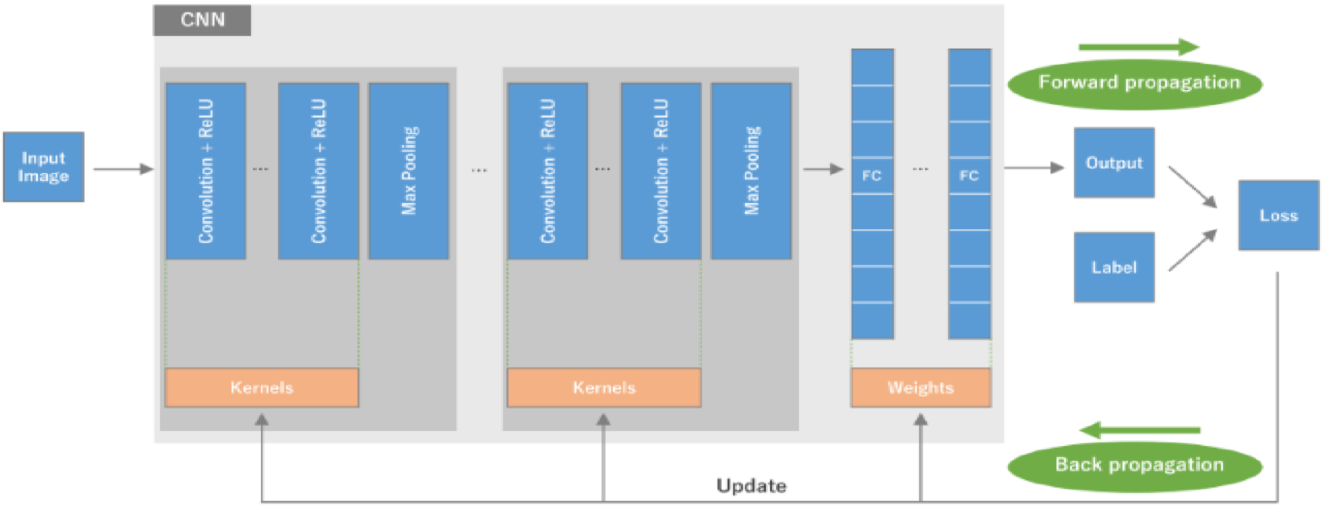

Figure 3. Overview of a convolution neural network [16]

CNN typically needs a lot of training data, and adding additional data can increase the inference accuracy. By using a deeper and larger network, CNN can also increase its inferences' accuracy. However, longer training times will result from bigger training sets or/and more complex models. A few unique sorts of layers are frequently utilized in CNNs [20, 21], such as loss, convolutional, ReLu (Rectified Linear Units), pooling, and fully connected layer [22]. Deep neural network used in many fields such as the study [23-28].

3.5 Evaluation metrics

The proposed model evaluation was performed based on recall, accuracy, and precision, that utilize the true positive (TP), true negative (TN), false positive (fp), and false negative (fn) terms [16], and these metrics are calculated as follows: [29, 30]

$$Accuracy(ACC) = \frac{Tp + Tn}{Total\ popoulation} \tag{1}$$

$$recall = \frac{Tp}{Tp + Fn} \tag{2}$$

$$precision = \frac{Tp}{Tp + Fp} \tag{3}$$

4. PROPOSED METHODS

Figure 4 shows the proposed method for the facial expression recognition system. The proposed method consists of four parts: 1) Preprocessing; 2) Face detection; 3) Feature extraction using HOG; 4) Classification using CNN algorithm. In the first part, the image is enhanced using the median filter as preprocessing step. In the second part, the face is detected using bounding box algorithm. Whereas in the third stage, the attributes are extracted using a histogram of gradient descent. Finally, the image is classified as one of the seven emotions (afraid, sad, angry, smile, happy, disgust, and surprise). The CNN algorithm composes of four layers: input, two hidden, and output layer. A median filter removes histogram equalization and noise from the images to enhance the contrast.

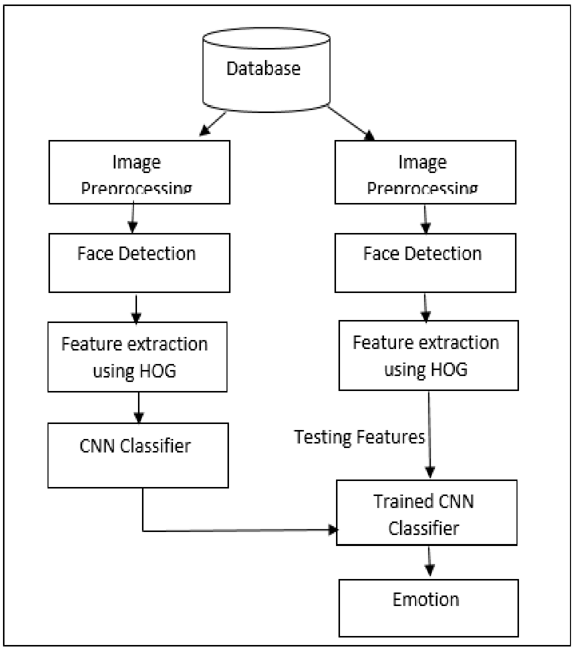

Figure 4. Proposed facial emotion recognition system Block Diagram

The fundamental concept behind the system proposed is based on the fact that every facial image person has various distinctive features. These attributes vary from one face image to another. In this work, the HOG descriptor is used to extract features from the image of the face. These algorithms depend on the idea of extracting features from face images of persons to discover the seven emotions. The algorithms rely on feature extraction from The Person's Face image to identify the emotions (afraid, sad, angry, smile, happy, disgust, and surprise).

5. EXPERIMENTAL RESULTS

The images are extracted from UMD face dataset and FAD (Face Attributes Dataset) to estimate the proposed system's performance accuracy, recall, and precision. The images are enhanced using a median filter, and images of various persons with several emotions are utilized to extract the parameters with the help of HOG; 140 images are utilized in the training stage and 70 images in the testing stage. Here we are classifying seven emotions (afraid, sad, angry, smile, happy, disgusts, and surprise).

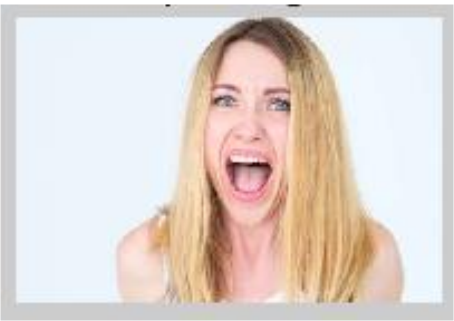

Figure 5. Input image

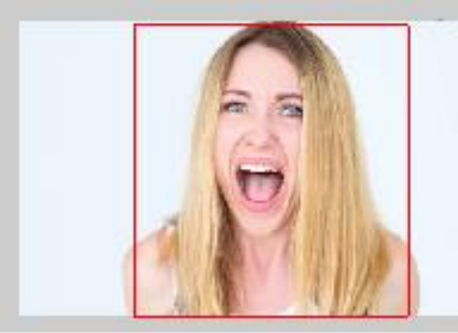

Figure 6. FACE image

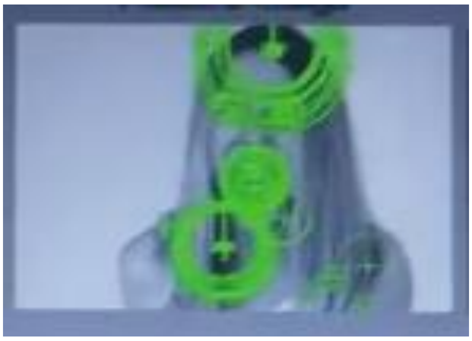

Figure 7. HOG image

Figure 5 shows input image. The image is first enhanced using median filter. First is detect the face part of the image as shown in Figure 6. Then the face image is filtered using histogram of oriented gradients. The HOG image is illustrated in Figure 7. From the result of Table 1, it is observed that accuracy of afraid emotion is 99.07%, sad emotion is 99.09%, angry emotion is 99.03%, happy emotion is 98.09%, smile emotion is 95.09%, disgust emotion is 97.03%, and surprised emotion is 99.06%.

In comparison to earlier work, the proposed method attained higher accuracy. In the research [11], the authors tested their FS-CNN method based on UMD face dataset and attained an average accuracy of 95%. In contrast, the proposed method attained an average accuracy of 98.07%. Accuracy of HOG-CNN method is illustrated in Figure 8, precision in Figure 9, and recall in Figure 10.

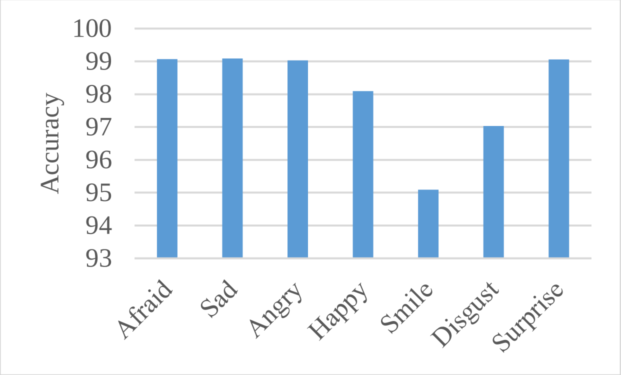

Figure 8. Accuracy of HOG-CNN method

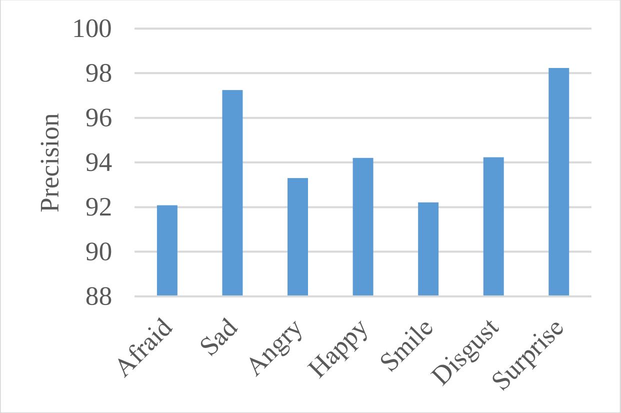

Figure 9. Precision of HOG-CNN method

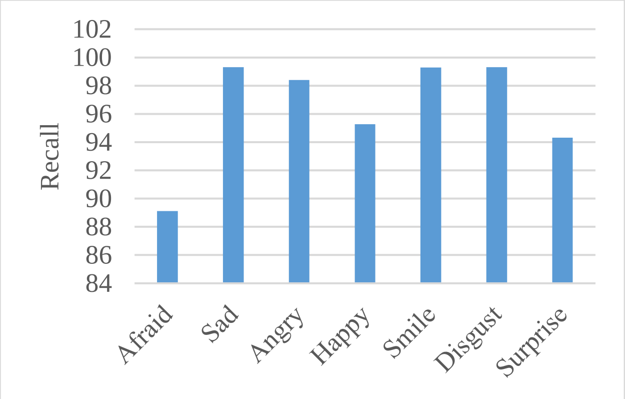

Figure 10. Recall of HOG-CNN method

Confusion matrix results of the proposed method is illustrated in Table 2.

For fair comparison the model is tested on CK+ dataset [31] and achieved a mean average accuracy of 94.48%. A comparison between proposed method and related work is illustrated in Table 3.

Table 1. Results of proposed HOG-CNN method

Emotion	Precision	Recall	Accuracy
Afraid	92.09%	89.12%	99.07%
Sad	97.24%	99.32%	99.09%
Angry	93.3%	98.4%	99.03%
Happy	94.21%	95.28%	98.09%
Smile	92.21%	99.28%	95.09%
Disgust	94.24%	99.32%	97.03%
Surprised	98.24%	94.32%	99.06%

Table 2. Confusion matrix

Reference	Dataset	Method	Accuracy
[5]	CK+	SVM+CNN	93.3%
[10]	CK+	HOG+ Three stage SVM	93.29%
[11]	UMD	FS-CNN	95%
[32]	CK+	ASM+Quadratic classifier	92.42%
Our method	UMD CK+	HOG-CNN	98.07% 94.48%

Table 3. Performance comparison with related work

	Afraid	Angry	Disgust	Happy	Sad	Smile	Surprise
Afraid	94.31	0.00	4.33	0.00	1.34	0.00	0.00
Angry	0.77	99.89	0.00	5.21	3.22	0.00	0.00
Disgust	0.00	4.98	90.32	1.12	3.00	0.00	12.4
Happy	2.14	8.90	0.00	98.19	0.00	8.01	0.77
Sad	1.29	0.00	0.10	0.00	92.42	0.31	0.00
Smile	0.00	0.00	9.12	0.09	0.05	95.67	0.06
Surprise	0.00	0.00	0.02	2.01	0.00	0.00	91.98

6. CONCLUSIONS

An adequate approach of addressing the FER problem was presented. A number of actions were carried out at the image preprocessing stage to assure that the characteristics obtained from the face detection holds a relevant and appropriate contribution to the classification stage. The additional attributes around ears on the side of face and those around chin and neck on bottom of mouth turned out to be useless for the classification stage. Consequently, they had to be cropped. Furthermore, the mouth needed to be kept whole and not cropped.

The muscle movements on the face form the facial expressions. HOG features detect subtle movements due to the fact that the HOG descriptor is sensitive to the object's shape. Experimental result has been conducted on images extracted from UMD face dataset and FAD (Face Attributes Dataset). From the result of Table 1, it is observed that average accuracy of proposed method is 98.07%.

For future work, different feature extraction techniques will utilize to extract features. Transfer learning will be used to increase the training data size and collect real images to construct a robust FER model.

REFERENCES

[1] Zadeh, M.M.T., Imani, M., Majidi, B. (2019). Fast facial emotion recognition using convolutional neural networks and gabor filters. In 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), pp. 577-581. https://doi.org/10.1109/KBEI.2019.8734943

[2] Tyng, C.M., Amin, H.U., Saad, M.N.M., Malik, A.S. (2017). The influences of emotion on learning and memory. Front in Psychology, 8: 1454. https://doi.org/10.3389/fpsyg.2017.01454

[3] Jamal, M., Hassan, T.A. (2022). Speech coding using discrete cosine transform and chaotic map. Ingénierie des Systèmes d'Information, 27(4): 673-677. https://doi.org/10.18280/isi.270419

[4] Jung, H., Lee, S., Yim, J., Park, S., Kim, J. (2015). Joint fine-tuning in deep neural networks for facial expression recognition. In 2015 IEEE International Conference on Computer Vision, pp. 2983-2991. https://doi.org/10.1109/ICCV.2015.341

[5] Subramanian, B., Yesudhas, H.R., Enoch, G.J. (2020). Channel-based encrypted binary arithmetic coding in wireless sensor networks. Ingénierie des Systèmes d'Information, 25(2): 199-206. https://doi.org/10.18280/isi.250207

[6] Chowdary, M.K., Nguyen, T.N.,Hemanth, D.J. (2021). Deep learning-based facial emotion recognition for human-computer interaction applications. Neural Computing and Appllicaiton, 8. https://doi.org/10.1007/s00521-021-06012-8

[7] Clawson, K., Delicato, L.S., Bowerman, C. (2018). Human centric facial expression recognition. In BCS Learning and Development Ltd. Proceedings of British HCI 2018, pp. 1-12. http://dx.doi.org/10.14236/ewic/HCI2018.44

[8] Kwolek, B. (2005). Face detection using convolutional neural networks and gabor filters. In Artificial Neural Networks: Biological Inspirations-ICANN 2005, pp. 551-556. https://doi.org/10.1007/11550822_86

[9] Wu, T.F., Bartlett, M.S., Movellan, J.R. (2010). Facial expression recognition using Gabor motion energy filters. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 42-47. https://doi.org/10.1109/CVPRW.2010.5543267

[10] Dagher, I., Dahdah, E., Al Shakik, M. (2019). Facial expression recognition using three-stage support vector machines. Visual Computing for Industry, Biomedicine, and Art, 2(1): 1-9. http://dx.doi.org/10.1186/s42492-019-0034-5

[11] Said, Y., Barr, M. (2021). Human emotion recognition based on facial expressions via deep learning on high-resolution images. Multimed Tools and Application, 80(16): 25241-25253. https://doi.org/10.1007/s11042-021-10918-9

[12] UMD faces dataset (n.d.). [Online]. Available: http://umdfaces.io/.

[13] Boateng, K.O., Weyori, A.B., Laar, D.S. (2012). Improving the effectiveness of the median filter. International Journal of Electronics and Communication Engineering, 5(1): 85-97.

[14] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern

Recognition, 1: 886-893. https://doi.org/10.1109/CVPR.2005.177

[15] Jan, A. (2017). Deep learning based facial expression recognition and its applications. Brunel University London.

[16] Hussein, A.K. (2019). Histogram of gradient and local binary pattern with extreme learning machine based ear recognition. Journal of Southwest Jiaotong University, 54(6): 1-6. http://dx.doi.org/10.35741/issn.0258-2724.54.6.31

[17] Zhou, W., Gao, S.Y., Zhang, L., Lou, X. (2020). Histogram of oriented gradients feature extraction from raw bayer pattern images. IEEE Transactions on Circuits and Systems II: Express Briefs, 67(5): 946-950. https://doi.org/10.1109/TCSII.2020.2980557

[18] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K. (2018). Convolutional neural networks : an overview and application in radiology. Insights Imaging, 9(4): 611-629. https://doi.org/10.1007/s13244-018-0639-9

[19] Hubel, D.H., Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. The Journal of Physiology, 160(1): 106-154. https://doi.org/10.1113/jphysiol.1962.sp006837

[20] Liu, T.Y., Fang, S.S., Zhao, Y.H., Wang, P., Zhang, J. (2015). Implementation of training convolutional neural networks. Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.1506.01195

[21] Pitaloka, D.A., Wulandari, A., Basaruddin, T., Liliana, D.Y. (2017). Enhancing CNN with preprocessing stage in automatic emotion recognition. Procedia Computer Science, 116: 523-529. https://doi.org/10.1016/j.procs.2017.10.038

[22] Gu, S.Q., Pednekar, M., Slater, R. (2019). Improve image classification using data augmentation and neural networks. SMU Data Science Review, 2(2): 1-43.

[23] Hussein, A.K. (2021). Fast learning neural network based on texture for Arabic calligraphy identification. Indonesian Journal of Electrical Engineering and Computer Science, 21(3): 1794-1799. https://doi.org/10.11591/ijeecs.v21.i3.pp1794-1799

[24] Hussein, A.K. (2019). Subject review: Machine learning and deep learning based arabic handwriting recognition. International Journal of Engineering Research and Advanced Technology, 5(10): 9-14. https://doi.org/10.31695/IJERAT.2019.3578

[25] Ayad, H., Ghindawi, I.W., Kadhm, M.S. (2020). Lung segmentation using proposed deep learning architecture. International Journal of Online and Biomedical Engineering, 16(15): 141-147. http://dx.doi.org/10.3991/ijoe.v16i15.17115

[26] Atya, B.A., Ali, O.T. (2019). Predict the relationship between autism and the use of smart devices by children in the coming years using neural networks. In 2019 First International Conference of Computer and Applied Sciences (CAS), pp. 79-83. http://dx.doi.org/10.1109/CAS47993.2019.9075775V

[27] Hoomod, H.K., Amory, Z.S. (2020). Temperature Prediction Using Recurrent Neural Network for Internet of Things Room Controlling Application. In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 973-978. https://doi.org/10.1109/ICCES48766.2020.9137885

[28] Hussein, K.A., Al-Ani, Z.T.A. (2022). Iraqi license plate recognition based on neural network technique. In Journal of Physics: Conference Series, 2322(1): 12025. http://dx.doi.org/10.1088/1742-6596/2322/1/012025

[29] Saeed, N.A., Al-Ta’i, Z.T.M. (2019). Feature selection using hybrid dragonfly algorithm in a heart disease predication system. International Journal of Engineering and Advanced Technology, 8(6): 2862-2867. http://dx.doi.org/10.35940/ijeat.F8786.088619

[30] Saeed, N.A., Al-Ta’i, Z.T.M. (2020). Heart disease prediction system using optimization techniques. In New Trends in Information and Communications Technology Applications, pp. 167-177. https://doi.org/10.1007/978-3-030-55340-1_12

[31] Kanade, T., Cohn, J.F., Tian, Y.L. (2010). Comprehensive database for facial expression analysis. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG’00), pp. 484-490. https://doi.org/10.1109/AFGR.2000.840611

[32] Ayache, F., Alti, A. (2020). Performance evaluation of machine learning for recognizing human facial emotions. Revue d'Intelligence Artificielle, 34(3): 267-275. https://doi.org/10.18280/ria.340304

NOMENCLATURE

HOG	Histogram of oriented gradient
PCA	Principal components analysis
CNN	Convolution neural network
GME	Gabor motion energy filters