Detection and Classification Expression of Visual Information Based on Artificial Intelligent Review Work

Nashwan J. Hussein[*], Amar A. Mahawish[**], Ibrahim Sami Attar[***]

[*] Electrical Power Engineering Department, AL Hussain University College, Karbala, Department of Programming, College of Information Technology, University of Babylon, Iraq
Email: nashwan.jasem@huciraq.edu.iq
https://orcid.org/0000-0001-8447-2846

[**] Department of Computer Engineering, College of Engineering, Al-Iraqia University, Baghdad, Iraq
Email amar.mahawish@aliraqia.edu.iq
https://orcid.org/0000-0002-2188-9940

[***] University Sains Malaysia, George Town, Malaysia
Email ibrahim_sami@student.usm.my
https://orcid.org/0000-0001-7766-5391

Abstract

Human Emotion Expression is difficult to define and categorize, making it a difficult research topic. Teaching machines to decipher facial expressions is difficult. Dynamic facial expressions include modest muscular movements that shape their appearance. Advanced algorithms and methods are needed to capture and analyze these minute differences. Deep learning's popularity has expanded emotion recognition applications. The ability to recognize facial expressions has several practical uses. In the medical field, it can help doctors better understand their patients' mental health issues and formulate effective treatment plans. Emotion recognition may be used in the field of customer service to get a feel for how satisfied customers are and adjust strategies accordingly. User experiences may be improved through human-computer interaction when emotion detection allows for more natural and individualized interactions with technology. Gaming and entertainment could employ emotional reaction recognition to tailor and engage players based on their moods. Emotion detection can also improve situational awareness and identify high-stress situations in law enforcement and security. Recognition of pupils' facial expressions can assist teachers in improving classroom instruction and comprehending their emotions. Deep learning models provide an advantage over prior emotion-recognition systems, but they need additional research for high-robust expression detection.

Keywords- Emotion Expression, Face Detection, Human Emotion Analysis, CNN Algorithm, Static and Dynamic expression.

I. INTRODUCTION

The ability to recognize emotion is built into computers, and one day, they may even be programmed to experience emotions. Emotional theory has a long and illustrious history; the classic Aristotelian theory of emotions by Aristotle examines the evolution of his thinking on emotions by defining and explaining a wide range of emotions, contrasting and comparing them, and characterizing the emotions themselves. Remarkable insights emerged from his ideas, such as the following: Emotions, such as anger, pity, fear, and their opposites, are the reasons people transform, have different perspectives, and experience both positive and negative feelings [1]. The usage of biometric data is becoming increasingly commonplace, with examples ranging from fingerprint scanning technology for logging into secure databases to face recognition via passport photos for gaining entry to a nation. The primary goal is to provide a more foolproof method of identification than, say, a password. By removing the need for human intervention, this technology streamlines the process and increases security, reducing the likelihood of forgery or fraud. Biometric identification can be accomplished using a variety of human characteristics, including the face, iris, voice, and fingerprints [2, 3, 4]. A further subfield of Human-computer interaction (HCI) is devoted to processing feelings; this area is known as affective computing. Researchers in the field of computer vision use these datasets and attempt to decipher their machine-readable meaning. Many studies are focusing on this problem, hoping to find a way to automatically identify affective feelings [5, 6]. One's choices in many domains can be influenced by awareness of one's emotional condition.

The main objectives of this review paper are to analyze and evaluate Deep Learning DL, Convolution Neural Network CNN, and other models on emotion recognition using facial expressions, implement the Histogram of Oriented Gradients feature extraction technique and analyze real-time emotion detection frameworks for audio and video multimodal. The DEAP dataset is used to classify emotion detection. The analysis and results are obtained by comparing various models, as shown in the next section.

II. REVIEW OF EMOTION DETECTION AND ALGORITHM ANALYSIS

The proposed methodology for recognizing facial expressions and detecting moods incorporates a comprehensive explanation of dataset integration and architecture design of Deep Learning (DL) models. Furthermore, this study aims to determine how effective a deep learning model is in recognizing facial expressions and detecting moods through the evaluation of the algorithm. This means that this literature review on human emotions will focus on facial expression analysis. It will explain how various computer programs can be utilized to recognize facial expressions. The current state of research and practical applications for identifying these feelings will be given via a review of the literature. Where DL has emerged to help put accurate Machine Learning (ML) into practice, particularly through the application of neural networks for both learning and prediction [6]. Various deep learning algorithms that can be used for face expression recognition are covered. In the twenty-first century, computers are ubiquitous and play a crucial role in society. The ability to recognize emotion is being built into computers, and one day, they may even be programmed to experience emotions. Emotional theory has a long and illustrious history, dating back to the Stoics, Plato, and Aristotle of Ancient Greece [7]. The classic Aristotelian theory of emotions by Aristotle examines the evolution of his thinking on emotions by defining and explaining a wide range of emotions, contrasting and comparing them, and characterizing the emotions themselves. Remarkable insights emerged from his ideas, such as the following: Emotions, such as anger, pity, fear, and their opposites, are the reasons people transform, have different perspectives and experience both positive and negative feelings [8]. The usage of biometric data is becoming increasingly commonplace, with examples ranging from fingerprint scanning technology for logging into secure databases to face recognition via passport photos for gaining entry to a nation. The primary goal is to provide a more foolproof method of identification than, say, a password. By removing the need for human intervention, this technology streamlines the process and increases security, reducing the likelihood of forgery or fraud. Biometric identification can be accomplished using a variety of human characteristics, including the face, iris, voice, and fingerprints [9]. The facial muscles contract to produce an expression that other people may see. Ekman [10]. Six core facial expressions were identified, representing the full range of human emotions (happiness, surprise, fear, sadness, anger, and disgust). The basic displays of emotion can be expanded upon to convey a wide range of nuanced sentiments. In 2021, Researchers Sujanaa et al. employed a dataset that consisted of still frames of videos showing images of people's mouths conveying various emotions. Images of people's faces have their mouths removed using a Haar-based cascade classifier, and frames are taken from the video at a rate of 20 per second. Each histogram in the feature set represents a different image of a person's mouth, and each histogram is based on using tools like Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP). Two methods, Speed Up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT), are utilized to separate individual data points [11]. Zamani and Wulansari classified and proposed two models that combine the best features of the In order to achieve this objective, utilizing the One-Dimensional Convolutional Neural Network (1D-CNN) as well as the Recurrent Neural Network (RNN). The RNN design includes Gated Recurrent Units (GRUs) and Long Short-Term Memories (LSTMs) to address the vanishing gradient issue common to time series data. High-Value-High-Arousal (HVHA), High-Value-Low-Arousal (HVLA), Low-Value-High-Arousal (LVHA), and Low-Value-Low-Arousal (LVLA) are the four emotional zones that our model distinguishes LVLA. The popular DEAP is a dataset for emotional analysis that uses physiological and audiovisual signals in this experiment. According to the experiments, the training accuracy of the suggested approaches is 96.3% for the 1D-CNN-GRU model and 97.8% for the 1D-CNN-LSTM model. This emotion classification task is, therefore, well within the capabilities of both models [12]. The Researchers Srinivas and Mishra define a multimodal system for emotion recognition that integrates characteristics from disparate modalities, such as audio and video. Energy, zero crossing rate, and Mel-Frequency Cepstral Coefficients (MFCC) are all strategies considered while extracting audio features. The findings from MFCC are very encouraging. First, using a spatial temporal Gaussian Kernel, the films are split into frames and saved in a linear scale space. Applying a Gaussian weighted function to the second momentum matrix of linear scale space further extracts characteristics from the photos. The audio and video features are fused using the Marginal Fisher Analysis (MFA) fusion method, and the combined features are then sent into the Facial Expression Recognition CNN (FERCNN) model for analysis [13]. Table (1) explains the summary Details of the Researcher's work.

TABLE I. explain the summary Details about the Researcher's work.

Methods Using	Result with Precision	Name-and Year
1D CNN	62.48%	Ristea et al In-2019[5].
LSTM+CNN	79.1%	Ryumina-and Karpov in 2020[6].
D1-CNN	90.23%	Sujanaa-and-Palanivel [11].
CNN-LSTM	78.53%	Hans and Rao [20].
SVM and 1D-CNN	97.44%	Sujanaa-etal-In2020 [21].
FERCNN model	97.88%	Srinivas-and Mishra-In 2022[13].
CNN	93.59%	Vijay and Yasutomo (2022) [22].
1DCNN-GRU 1DCNN-LSTM	96.3%& 97%	Zamani and Wulansari (2021) [12].

III. Attributes of Color Image

The use of color digital images is extensive across multiple fields, such as computer vision, image processing, and multimedia applications. Various models have been created to handle the representation and manipulation of color information in digital images. One commonly employed model is the RGB model, which combines different intensity levels of these three primary colors to represent colors. The RGB model is widely supported by imaging software and devices due to its intuitive nature. Another well-known color model is CMYK, which is frequently utilized in printing and graphic design practices. The CMYK color model is utilized to represent colors by utilizing the concept of subtractive color mixing. This involves combining different inks to achieve the desired colors.

Moreover, there are alternative color models such as HSV, Lab, and YUV/YCbCr that offer different perceptual properties and aid specific image processing tasks [14]. These models are significant in analyzing, manipulating, and comprehending color images. They enable the creation of algorithms and techniques for various applications related to color analysis. Using a Deep Learning Model for Emotion Recognition necessitates the following preprocessing steps:

A. Process of Grayscale Conversion

While RGB (Red, Green, Blue) images have three, converting the RGB image to grayscale can lower the processing burden and simplify the data. The formula for the transformation is as follows [15].

$$\text{Grayscale Value} = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

Computers and televisions both employ this color display technology. A digital color picture is composed of three RGB 2-D matrices, one for each basic color. The visual on the screen is constructed by multiplying the values of these three matrices together; typically, 8 bits are used to represent each of the three components of the three matrices. As can be seen in Figure (1), there are a total of 24 bits in a color pixel (3 x 8).

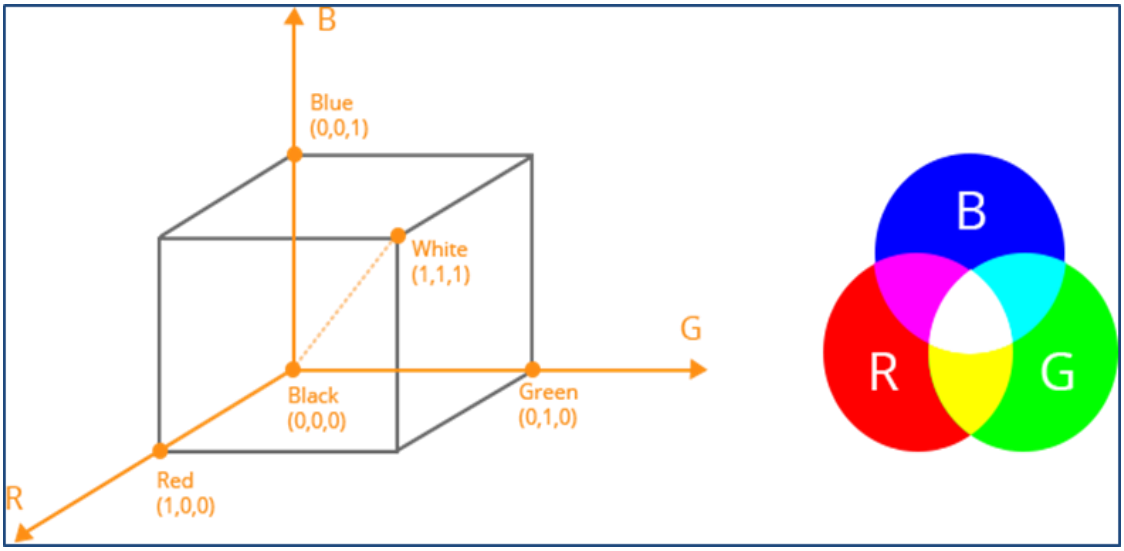

Figure 1. Explain the process conversion of RGB to Gray

B. Enhancements of Digital Images

Various digital image enhancement techniques may significantly improve photographs' visual quality and interpretability. This section will focus on histogram equalization, one of these techniques. Histogram equalization is one of the most common techniques for boosting the contrast of digital photos and improving their overall look.

C. Histogram Equalization process

Histogram equalization dispersing intensity values improves visual contrast. Histogram equalization is performed on an input picture having pixel values from 0 to L-1 (where L is the number of intensity levels, generally 256 for an 8-bit image) [16].

1. Create a histogram of the supplied image to see how often each brightness level appears. The histogram value at intensity level I will be denoted by H (i).

2. Calculate the cumulative distribution function (CDF) of the histogram. The CDF, denoted by CDF (i), represents the cumulative probability of each intensity level up to level i. It is computed using the following equation [17].

$$CDF(i) = \Sigma H(j) \text{ for } j = 0 \text{ to } i \quad (2)$$

Map intensity values from [0, L-1) to [0, L-1] by normalizing the cumulative distribution function (CDF). The new intensity values are comprehensive because of the normalizing procedure. The following formula is used to calculate the intensity at any level I [32]:

$$NewValue(i) = round((CDF(i) - CDF(min)) / (M * N - 1) * (L - 1)) \quad (3)$$

Where M*N is the number of pixels in the picture, L is the number of intensity levels, New Value(i) is the new intensity value, CDF(i) is the cumulative probability of intensity level I is the minimal cumulative probability among all intensity levels, and.

3. Apply the following equation to each pixel in the supplied picture, replacing its value with New Value (i). By using the histogram equalization method, the final image will have better aesthetic appeal and be more appropriate for further image analysis and processing because of the increased contrast and more even distribution of intensity values [18].

Algorithm 1: explain the processing of the histogram equalization equation

Input value: The pixel values of a grayscale facial image span from 0 to 255.

Output value: enhanced Gray

BEGIN

Step 1: Understanding the dimensions of the grayscale image (E*F) requires pixel values between 0 and 255. Set every element of a matrix G of size 256 to 0

Step 2: To create an image histogram, the corresponding elements must be updated in the matrix. – The image matrix is updated by scanning each pixel to create the histogram.

G [gray value (pixel)] = G [gray value (pixel)] + 1

Step 3: Using Eq. (2.2), CDF calculations can be performed.

CH [0] = H [0]

Each pixel (1 to 255) has CH[i] equal to CH [pixel-1] plus H[pixel].

Step 4: Referring to Eq (2), compute the updated pixel values through the process of general histogram equalization.

Step 5: Update image-wise; new values replace the original grayscale image.

NewImage [E][F] = T[OldImage[E][F]]

End

D. Standard Scaler

The standard scaler is a core preprocessing method in machine learning, with the goal of normalizing a dataset's features such that they all have the same mean value of zero and the same standard deviation of one. For many algorithms, especially those that are sensitive to the magnitude of input characteristics, this change is crucial for boosting convergence and speed. By subtracting the mean (μ) of the feature's values from each value and dividing by the standard deviation (σ) of the feature's values, the standard scaler calculates the Z-score for each feature [19]:

$$z = \frac{x-\mu}{\sigma} \quad (4)$$

Each data point is denoted by x, the feature mean is mu, the feature standard deviation is sigma, and the converted Z-score is z. In order to prevent bigger magnitude characteristics from dominating the learning process, the data is transformed so that it is centered around 0. This transformation also accounts for differences in scale. Therefore, the standard scale is an important preprocessing step that contributes to the development of more accurate machine learning models across a variety of datasets.

IV. Algorithm Using to Detect Expression

One area of computer-human interaction is facing detection, a branch of object detection. The technique of object detection involves finding instances of items belonging to a specific class (such as people, cars, buildings, or faces) in an image or video. In this new era, object detection has a wide variety of uses, including pedestrian and face detection [21, 22, 23].

A. Real-Time Visual Expression Algorithm Detection

The Viola and Jones technique is extensively used because it efficiently detects faces in real-time using a cascade of weak classifiers. Due to its high detection accuracy and fast processing speed, the Viola & Jones algorithm is well-suited for use in real-time settings [24]. Simple rectangular filters that record local intensity variations in the image form the basis of its operation, and these filters are known as Haar-like features. The system can recognize faces of diverse sizes and orientations thanks to the computation of these features at many scales and locations across the image [25].

The Viola and Jones face detection method consists of several sub-steps, each of which aids in the recognition and localization of individual facial features. In brief, the algorithm consists of the following steps [26].

1. Haar-like features are basic rectangular filters that capture local picture intensity fluctuations. These traits allow face recognition. The method chooses Haar-like characteristics that can distinguish face and non-facial areas.

2. Calculating the integral picture accelerates Haar-like feature calculation. The integral picture sums pixel intensities in a rectangle region from the top-left corner. This equation efficiently computes

$$\text{Integral Image} = \mathrm{I}_{\mathrm{Integ}(x_2, y_2)} + \mathrm{I}_{\mathrm{Integ}(x_1-1, y_1-1)} - \mathrm{I}_{\mathrm{Integ}(x_1-1, y_2)} - \mathrm{I\_Integ}(x_2, y_1 - 1) \quad (5)$$

Where Integral Image is the integral picture value at coordinate (x_2, y_2) and Image (x_1, y_1) is the pixel intensity in the original image.

3. The approach uses the Adaboost machine learning technique to combine numerous weak classifiers into a strong one. The system picks a subset of training pictures and iteratively modifies the weights of weak classifiers depending on their performance in distinguishing face and non-facial areas.

4. The Viola & Jones technique uses a cascade of classifiers to reject non-facial areas and focus computation on face regions efficiently. Each cascade level contains multiple weak classifiers. An area is deleted if it does not fulfill a stage's criteria, decreasing computation for succeeding stages.

5. The technique uses a sliding window to scan the picture at multiple sizes and places. At each window point, the integral image computes Haar-like characteristics, and the cascade of classifiers sequentially determines if the region includes a face.

6. Eliminates redundant face detections and improves accuracy. Selecting the region with the greatest detection confidence score eliminates overlapped detections.

B. Classification and Feature Extract

The initial step in the first model involves feature extraction utilizing the HOG (Hand-crafted Features) algorithm, followed by the application of a 1D CNN. The algorithm employed in this thesis effectively captures and represents the local gradient patterns present in the input image [27, 28]. These patterns are subsequently converted into a feature vector for further analysis and processing. The features obtained from the Histogram of Oriented Gradients (HOG) algorithm are subsequently inputted into a one-dimensional Convolutional Neural Network (CNN) model. The architectural design of the 1D Convolutional Neural Network (CNN) is carefully constructed to analyze and interpret the facial expressions in video sequences effectively. This particular architecture enables the model to accurately perceive and understand the dynamic changes in facial expressions that occur over some time. The system utilizes the retrieved characteristics as input and proceeds with the classification job, intelligently assigning the input sequences to their

corresponding facial expression categories. The 2D Convolutional Neural Network (CNN) will be the second model, and the process of feature extraction and classification will be integrated. The proposed model utilizes input images and applies a sequence of frames by using Time Distributed to convolutional layers to extract spatial patterns and hierarchical representations of facial expressions. The 2D convolutional neural network (CNN) model effectively leverages its inherent capacity to acquire discriminative features from unprocessed image data in order to extract pertinent information. Subsequently, the system proceeds to conduct classification utilizing the acquired features, effectively assigning appropriate categories to the facial expressions depicted in the input images.

C. Oriented Gradients

The HOG technique is used in the proposed system in order to extract features from the video frames. The HOG technique records the distribution of gradients within the localized facial areas, which are key cues for understanding facial expressions. These gradients serve as essential cues since they vary from region to region on the face. In order for the HOG method to function, the facial region is first segmented into smaller cells, and then the gradient magnitude and orientation within each cell are calculated. The gradient orientations are then arranged in a histogram, with the histogram bins being determined by the magnitude of the gradients [29, 30].

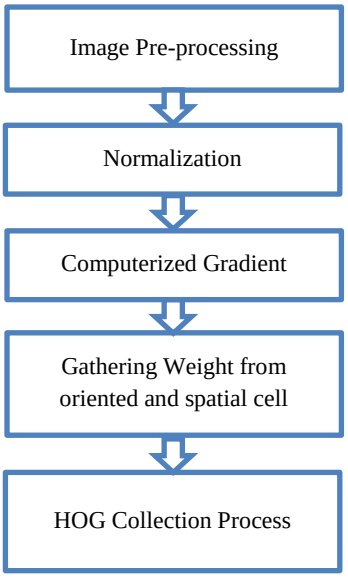

Figure 2. Explain the HOG Algorithm and Gradient

D. Classification Process Using First and Second Convolution Neural Network

Network (1D CNN) and a Two-Dimensional Convolutional Neural Network (2D-CNN). The first stage of the process focuses on pre-processing processes designed to improve the quality and interpretability of the supplied data. Following the feature extraction step, the use of the 1D CNN model is implemented, capitalizing on its ability to proficiently categorize emotions in video sequences [31, 32]. Furthermore, the utilization of a 2D-CNN model enables the concurrent extraction of features and classification, hence facilitating a thorough assessment and efficient categorization of emotions [33, 34].

V. CONCLUSION

In this study, DL, CNNs, and related algorithms were evaluated for emotion recognition of the face. Despite the impressive progress made in the analysis of detection, the performance of the CNN model is heavily influenced by the careful selection of hyper-parameters such as kernel size, stride, and filter types. In addition, the significance of both the quality and amount of data cannot be overstated in the attainment of elevated levels of accuracy. While DL produces high quality in the processing of emotion recognition, it introduces additional complexity. This review highlights the important role of feature extraction methods by using the Histogram of Oriented Gradients (HOG) approach, which is of utmost importance in the feature extraction process from video frames inside the suggested system. The HOG approach is proficient at capturing substantial texture and shape information, hence enhancing the system's capability to interpret facial emotions properly. Accuracy in recognizing and classifying facial expressions has also been extensively researched.

REFERENCES

[1] E. Goceri, "Deep learning based classification of facial dermatological disorders," Comput Biol Med, vol. 128, 2021, doi: 10.1016/j.compbiomed.2020.104118.

[2] K. Chlasta, K. Wołk, and I. Krejtz, "Automated speech-based screening of depression using deep convolutional neural networks," Procedia Computer Science, vol. 164, pp. 618–628, 2019, doi: 10.1016/j.procs.2019.12.228.

[3] W. H. Abdulsalam, R. S. Alhamdani, and M. N. Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," Int J Mach Learn Comput, vol. 9, no. 1, pp. 14–19, 2019, doi: 10.18178/ijmlc.2019.9.1.759.

[4] D. A. S. Devi, C. Satyanarayana, and D. S. Rekha, "Facial Emotion Recognition Using Hybrid Approach for DCT and DBACNN," in Lecture Notes in Electrical Engineering, Springer Science and Business Media Deutschland GmbH, 2022, pp. 411–423. doi: 10.1007/978-981-16-9885-9_34.

[5] N. C. Ristea, L. C. Dutu, and A. Radoi, "Emotion recognition system from speech and visual information based on convolutional neural networks," in 2019 10th International Conference on Speech Technology and Human-Computer Dialogue, SpeD 2019, Institute of Electrical and Electronics Engineers Inc., Oct. 2019. doi: 10.1109/SPED.2019.8906538.

[6] E. Ryumina and A. Karpov, "Facial expression recognition using distance importance scores between facial landmarks," in CEUR Workshop Proceedings, CEUR-WS, 2020. doi: 10.51130/graphicon-2020-2-3-32.

[7] Z. Wang et al., "Automated Rest EEG-Based Diagnosis of Depression and Schizophrenia Using a Deep Convolutional Neural Network," IEEE Access, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3197645.

[8] B. Victor, K. Bowyer, and S. Sarkar, "An evaluation of face and ear biometrics," Proceedings - International Conference on Pattern Recognition, vol. 16, no. 1, 2002, doi: 10.1109/icpr.2002.1044746.

[9] Y. Ma, Z. Huang, X. Wang, and K. Huang, "An Overview of Multimodal Biometrics Using the Face and Ear," Mathematical Problems in Engineering, vol. 2020. 2020. doi: 10.1155/2020/6802905.

[10] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2008. doi: 10.1007/978-3-540-89991-4_6.

[11] J. Sujanaa, S. Palanivel, and M. Balasubramanian, "Emotion recognition using support vector machine and one-dimensional convolutional neural network," Multimed Tools Appl, vol. 80, no. 18, pp. 27171–27185, Jul. 2021, doi: 10.1007/s11042-021-11041-5.

[12] F. Zamani and R. Wulansari, "Emotion Classification using 1D-CNN and RNN based On DEAP Dataset," Academy and Industry Research Collaboration Center (AIRCC), Dec. 2021, pp. 363–378. doi: 10.5121/csit.2021.112328.

[13] P. V. V. S. Srinivas and P. Mishra, "Human Emotion Recognition by Integrating Facial and Speech Features: An Implementation of Multimodal Framework using CNN," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, 2022, doi: 10.14569/ijacsa.2022.0130172.

[14] K. Kumar, R. K. Mishra, and D. Nandan, "Efficient Hardware of RGB to Gray Conversion Realized on FPGA and ASIC," in Procedia Computer Science, 2020. doi: 10.1016/j.procs.2020.04.215.

[15] H. Mutahira, B. Ahmad, M. S. Muhammad, and D. R. Shin, "Focus Measurement in Color Space for Shape from Focus Systems," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3098753.

[16] S. Sood, H. Singh, and M. Malarvel, "Image quality enhancement for Wheat rust diseased images using Histogram equalization technique," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021, doi: 10.1109/iccmc51019.2021.9418023.

[17] H. Kh. Omar and N. E. Tawfiq, "Face Recognition Based on Histogram Equalization and LBP Algorithm," Academic Journal of Nawroz University, vol. 8, no. 3, p. 33, Aug. 2019, doi: 10.25007/ajnu.v8n3a394.

[18] S. H. Gangolli, A. Johnson Luke Fonseca, and R. Sonkusare, "Image Enhancement using Various Histogram Equalization Techniques," 2019 Global Conference for Advancement in Technology (GCAT), Oct. 2019, doi: 10.1109/gcat47503.2019.8978413.

[19] D. D. Olatinwo, A. Abu-Mahfouz, G. Hancke, and H. Myburgh, "IoT-Enabled WBAN and Machine Learning for Speech Emotion Recognition in Patients," Sensors, vol. 23, no. 6, p. 2948, Mar. 2023, doi: 10.3390/s23062948.

[20] A. S. A. Hans and S. Rao, "A CNN-LSTM based deep neural networks for facial emotion detection in videos," International Journal of Advances in Signal And Image Sciences, vol. 7, no. 1, pp. 11–20, Mar. 2021, doi: 10.29284/ijasis.7.1.2021.11-20.

[21] Christy, A., Vaithyasubramanian, S., Jesudoss, A., & Praveena, M. A. (2020). "Multimodal speech emotion recognition and classification using convolutional neural network techniques". International Journal of Speech Technology, 23(2), 381-388.

[22] V. John and Y. Kawanishi, "Audio and Video-based Emotion Recognition using Multimodal Transformers," in Proceedings - International Conference on Pattern Recognition, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 2582–2588. doi: 10.1109/ICPR56361.2022.9956730.

[23] Z. Dahirou, M. Zheng, and M. Yuxin, "Face Detection with Viola Jones Algorithm," 2020 7th International Conference on Information Science and Control Engineering (ICISCE), Dec. 2020, doi: 10.1109/icisce50968.2020.00130.

[24] P. Xayachack and J. Zhang, "Robust Face detection based on Viola-jones Algorithms," IOP Conference Series: Materials Science and Engineering, vol. 768, no. 6, p. 062079, Mar. 2020, d899x/768/6/062079.oi: 10.1088/1757.

[25] J. Huang, Y. Shang, and H. Chen, "Improved Viola-Jones face detection algorithm based on HoloLens," EURASIP J Image Video Process, vol. 2019, no. 1, 2019, doi: 10.1186/s13640-019-0435-6.

[26] "Comparative of Viola-Jones and YOLO v3 for Face Detection in Real time," Iraqi Journal of Computer, Communication, Control and System Engineering, 2022, doi: 10.33103/uot.ijccce.22.2.6.

[27] A. Jadhav, S. Lone, S. Matey, T. Madamwar, and S. Jakhete, "Survey on Face Detection Algorithms," Int J Innov Sci Res Technol, vol. 6, no. 2, 2021.

[28] V. Upadhyay and D. Kotak, "A Review on Different Facial Feature Extraction Methods for Face Emotions Recognition System," 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Jan. 2020, doi: 10.1109/icisc47916.2020.9171172.

[29] M. Moe Htay, "Feature extraction and classification methods of facial expression: a survey," Computer Science and Information Technologies, vol. 2, no. 1, pp. 26–32, Mar. 2021, doi: 10.11591/csit.v2i1.p26-32.

[30] P. N. Maraskolhe and A. S. Bhalchandra, "Analysis of Facial Expression Recognition using Histogram of Oriented Gradient (HOG)," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Jun. 2019, doi: 10.1109/iceca.2019.8821814.

[31] R. C. Ng, K. M. Lim, C. P. Lee, and S. F. A. Razak, "Surveillance system with motion and face detection using histograms of oriented gradients," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 2, 2019, doi: 10.11591/ijeecs.v14.i2.pp869-876.

[32] M. Shatnawi, N. Almenhali, M. Alhammadi, and K. Alhanaee, "Deep Learning Approach for Masked Face Identification," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, 2022, doi: 10.14569/ijacsa.2022.0130637.

[33] F. J. Díaz-Pernas, M. Martínez-Zarzuela, D. González-Ortega, and M. Antón-Rodríguez, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," Healthcare (Switzerland), vol. 9, no. 2, Feb. 2021, doi: 10.3390/healthcare9020153.

[34] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, A. Karim, and A. Islam, "TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study," IEEE Access, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3179577.