International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 6
ISSN: 2088-5334

Mental Health State Classification Using Facial Emotion Recognition and Detection

Adel Aref Ali Al-zanam [a], Omer Hussein Abdou Elsayed Hussein Alhomery [a,*], Choo Peng Tan [a]

[a] Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Bukit Beruang, Melaka, Malaysia
Corresponding author: [*]omarhusseinalhomery@gmail.com

Abstract—**Analyzing and understanding emotion can help in various aspects, such as realizing one's attitude, behavior, etc. By understanding one's emotions, one's mental health state can be calculated, which can help in the medical field by classifying whether one is mentally stable or not. Facial Recognition is one of the many fields of computer vision that utilizes convolutional networks or Conv Nets to perform, train, and learn. Conv Nets and other machine learning algorithms have evolved to adapt better to larger datasets. One of the advancements in Conv Nets and machines is the introduction of various Conv architectures like VGGNet. Thus, this study will present a mental health state classification approach based on facial emotion recognition. The methodology comprises several interconnected components, including preprocessing, feature extraction using Principal Component Analysis (PCA) and VGGNet, and classification using Support Vector Machines (SVM) and Multilayer Perceptron (MLP). The FER2013 dataset tests multiple models' performances, and the best model is employed in the mental health state classification. The best model, which combines Visual Geometry Group Network (VGGNet) feature extraction with SVM classification, achieved an accuracy of 66%, demonstrating the effectiveness of the proposed methodology. By leveraging facial emotion recognition and machine learning techniques, the study aims to develop an effective method.**

Keywords—**Principal component analysis; multilayer perceptron; facial emotion recognition; CNN; visual geometry group.**

Manuscript received 8 Jan. 2023; revised 14 Jun. 2023; accepted 21 Sep. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

As human beings undergo various moments and events throughout their daily life, they encounter countless ordeals and occasions that may affect their emotions. Emotions are forms of expression that humans naturally show when faced with an ordeal or other occasions. In addition, a human's behavior, thoughts, and feelings can be affected by emotion. Emotions can be either positive or negative; positive emotions are associated with pleasant ones, such as joyfulness, and negative ones are associated with negative ones, such as anger [1]. Nonetheless, while undergoing an emotional phase, the specific emotion associated with that phase can affect one's mental health [2]. Mental health is concerned with one's collective psychological welfare, which is related to one's ability to maneuver their feelings along with their consciousness in dealing with ordeals [2]. As those gratified with good emotional states can positively handle themselves, dealing with their surroundings and antagonizing their negative emotions, while those with bitter emotional states can witness difficulties in dealing with the aforementioned aspect. Thus, deducing that mental health state is a vital component of one's overall well-being, and timely detection and appropriate management of mental health conditions are essential for enhancing one's outcomes. However, accurately diagnosing and monitoring mental health states remains challenging, necessitating the development of innovative, objective, and non-invasive assessment tools. Humans tend to utilize their facial expressions to communicate different types of messages in several contexts. Facial expression definitions include a wide spectrum of concepts, from simple, maybe basic socioemotional terms like "surprise," to complex communication like "carelessly." Facial expressions are utilized in different circumstances, from responses to outside events to linguistic formulations in sign language. Effective interpersonal communication requires accurate processing and interpretation of emotions conveyed through facial expressions; thus, it is widely recognized that facial expressions are commonly connected to emotions [3]. An emotion can be extracted by focusing and analyzing the facial features, such as eyebrows, lips, nose, mouth, and facial muscles, since they tend to behave based on one's feelings, regardless of whether one attempts to fake their feelings [4]. When recognizing facial expressions, seven universal facial expressions have received widespread scientific approval. Happiness, sadness, fear, disgust, anger, contempt, and surprise are among them [5]. These phrases contain useful information that may be used in various situations. For example, in commerce, a customers' facial expressions, for example, are utilized to determine whether they are truly interested in purchasing or to comprehend their shopping purpose [6]. This suggests that a person's behavioral patterns are intimately linked to their facial expression [7]. Therefore, the mental health state of a person can be closely determined by using facial expressions [8].

Thus, this paper proposes an approach for determining mental health states by aggregating facial emotion scores over seven days. This innovative technique allows for a comprehensive and dynamic representation of an individual's emotional landscape, facilitating a more accurate classification of mental health status.

II. Materials and Method

Due to the importance of facial emotion recognition in mental health state classification, multiple methods for emotion recognition were introduced and developed. Regardless, computer vision is always associated with facial expression recognition as it helps analyze images and classify them. The matter is further explained as follows based on various studies regarding facial expression recognition and computer vision.

A. Mental Health State Classification

Image classification case studies (FER) often cover facial expression recognition. Computer vision is one of the many subfields that AI spawned. This subfield helps computers and other systems interpret digital images, videos, and other forms of visual input. Both humans and computers use the object recognition process to recognize and identify objects. However, the criteria and methods utilized by humans and computers in recognizing objects are quite different [9]. Humans usually process visual information in semantic space. Extracting semantic information—line segments, borders, shape, etc.—accomplishes this. Computers can interpret data space by visuals like colors and textures, but they cannot do this like humans. Thus, humans and machines process visual information differently.

Image classification involves classifying and labeling pixels or vectors in an image according to a set of criteria. Spectral or textural attributes can be used for categorizing. Like any classification problem, image classification uses supervised and unsupervised methods. Supervised classification works by selecting a sample of pixels from an image that belongs to different classes and instructing the image processing program to use these training sites to classify the other pixels where the classes can be defined as labels. Whereas unsupervised classification utilizes software analysis of the picture to group pixels with similar characteristics without needing a user's intervention. Computer algorithms can be used to classify related pixels. Regardless, both picture categorization methods have pros and cons, and based on the training data, a categorization method is utilized [10].

Computer vision relies on image classification because object categorization is complex since image categorization involves classifying images, and a picture can have infinite classes. Nonetheless, when classifying images in huge numbers, an enormous amount of time is usually dedicated to the process in case it is manually implemented. Thus, computer vision automation may save time when reviewing and categorizing many images, emphasizing the importance of computer vision.

B. Image Preprocessing

When dealing with data, first, the data would be required to undergo the preprocessing process, which is more like preparing the data, so it is ready for use. Various research studies have been carried out on facial emotion detection. The studies are normally expressed in two parts: the method of feature extraction or feature selection, along with the classification method, preceded by various image preprocessing techniques. Image preprocessing improves image visibility and removes unwanted noise. Contrast enhancement techniques, such as wiener, regularized, Lucy-Richardson, blind convolution, and median filters, are widely used to enhance the visual quality of low-contrast images [11]. For denoising, homomorphic filters are used to remove unwanted pixel representations in images [12].

Resizing images is also an essential preprocessing step in image classification tasks. Traditional methods include cropping and scaling, while content-aware methods such as warping, multi-operator, and seam carving can also be used [13]. The warping function, which transfers coordinates in a source picture to positions in a destination image, may be used to specify the warping technique. Since the warping function is nonlinear, different areas of the image will have varying magnifications. The warping scaling method, which does not remove other portions of the picture, emphasizes the ROI. The Multi-operative approach combines many techniques to benefit from each one's strengths and minimize weaknesses. The content-aware image scaling operator known as seam carving comprises both reduction and expansion. A seam is a perfect 8-connected path of low-energy pixels that spans the picture from top to bottom or left to right.

C. Feature Extraction

Feature extraction or selection reduces the number of features in a dataset by developing new ones or choosing the most important ones. HOG is a commonly used feature extraction technique in picture feature extraction, which uses a histogram to describe the features of an image based on the direction change data along edges [14] [15]. Other feature extraction methods include facial landmarks [16], deep learning approaches using VGGNet [6] or RedNet [17], and dimensionality reduction using PCA or LDA or a combination of both. PCA approximates the original data using lower-dimensional feature vectors, while LDA projects data into a new linear feature space. In the tests, PCA was employed to decrease picture dimensions before LDA was utilized for feature extraction [18].

D. Image Classification

In recent years, image classification has been a growing research domain that has recorded many advances, especially in deep learning. For facial expression classification, the number of classes is mostly limited to seven, which are the common facial expressions as established by [5]. In the last decade, based on the survey done by [19], the most common image classification method used is the SVM [20], with 51% of the total reviews using SVM as the classification method. The most common image classification method has recently been using a convolution neural network. Based on [21], CNN is determined to perform better in classification accuracy than other classification techniques like Random Forest and SVM. This evaluation is done on hyperspectral images [17], [22], [23], [24]. The architecture arrangement, number of layers, and optimization parameters mostly differentiate the network. Other common machine learning methods for image classification are LSTM used in [6], [16].

E. Facial Emotion and Mental Health State

In order to utilize the facial emotion of humans in determining their current mental health state, establishing some kind of relationship between facial emotion and mental health would be required. Therefore, mental health conditions can be used to establish a correlation with facial emotions. Stress and negative emotions have a significant detrimental impact on mental health in many situations. Chang [25] conducted an experiment revolving around emotion induction, which collected five physiological signals from individuals, including an electrocardiogram, galvanic skin reactions (GSR), blood volume pulse, and pulse to ascertain the physiological condition of the patient (sadness, fear, and pleasure). This physiological state can be used to determine the current mental state of the person, given the current situation [25]. Other emotional expressions may be employed as a deciding element of a person's mental state if physiological states display emotions like sorrow, which can be used to assess a person's mental state. Facial expressions may generally be used to identify emotions.

The study's method or approach to facial emotion recognition and its application to mental health state classification is described in detail in the following section. The section is organized into the following components: system architecture, dataset, preprocessing, feature extraction, classification, and mental health state classification. The system architecture provides an overview of the entire process, while subsequent sections delve into each component's specifics. Following this structured approach, the study aims to develop an effective and accurate method for classifying mental health states based on facial emotion recognition.

F. System Architecture

The proposed system architecture consists of several interconnected components designed for facial emotion recognition and mental health state classification. It starts with the FER2013 dataset, which is divided into training and testing sets. Training images undergo preprocessing and feature extraction using both PCA and VGGNet methods. These features are then used in classification algorithms, SVM, and MLP, resulting in six different models. Testing images undergo preprocessing and feature extraction (PCA and VGGNet) before inputting the trained models. A diagram illustrating the system architecture is provided in Figure 1 to help visualize the flow and components of the system. The best model is selected based on performance and is then used in the mental health state classification.

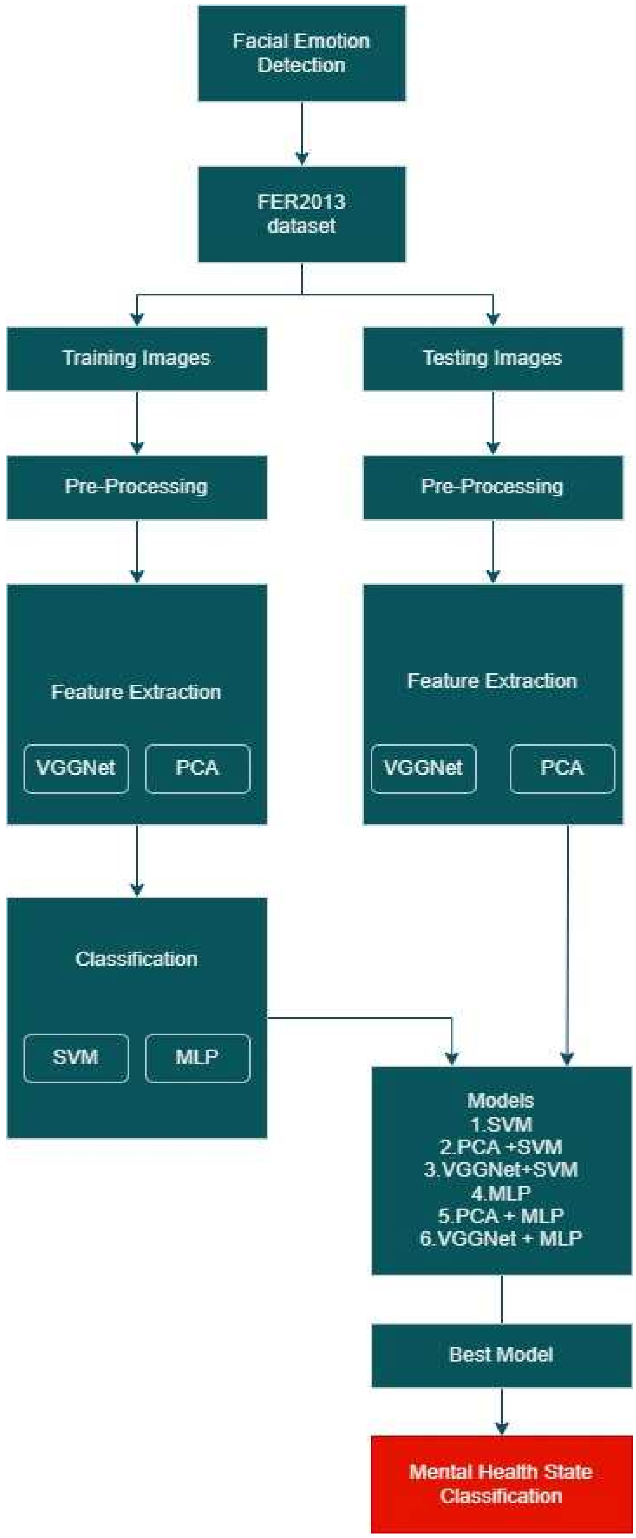


Fig. 1 The flow and components of the proposed system.

G. Dataset

The dataset to be used for this project is the FER2013 facial emotion dataset, which has been previously used in studies such as the study by Sultana [23].

TABLE I
DATASET DISTRIBUTION

Class	Class Number	No of Training Images	No of Validation Images	No of Testing Images	Total Class Images
Angry	0	3995	491	467	4953
Disgust	1	436	55	56	547
Fear	2	4097	528	496	5121
Happy	3	7215	879	895	8989
Sad	4	4830	594	653	6077
Surprise	5	3171	416	415	4002
Neutral	6	4965	626	607	6198
	TOTAL	28709	3589	3589	35887

The FER2013 dataset comprises approximately 35,000 facial RGB images for different facial expressions. All the images in the dataset are of restricted size (48 X 48 pixels). This dataset has seven labels: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral, which are represented by 0 to 6, respectively. The disgusting sample images are the least of all the classes, having only about 500 images. The other classes have over 4,000 images to represent them. The images in this dataset are collected from Google with varying poses, ages, and occlusions. The dataset will be divided into training, validation, and testing sets: 80% for training, 10% for validation, and 10% for testing. The database is open source and available on Kaggle. A summary of the dataset's distribution can be seen in Table I.

H. Preprocessing

The preprocessing step is applied to both the training and testing images. The only preprocessing step performed on the FER2013 dataset is normalization since the images are already in the required size and RGB format. Normalizing the images ensures that each pixel falls within a similar data distribution, which can improve training time and convergence speed. In this project, the image normalization function scales every pixel value between -0.5 and 0.5. Given that pixel values range from 0 to 255, the first step is dividing each pixel value by 255, then subtracting 0.5. Consequently, a pixel with a value of 0 becomes -0.5, while a pixel with a value of 255 becomes 0.5. For example, a pixel with a value of 184 will be assigned a value of 0.22. After normalization, the dataset is saved, separating training, validation, and testing data. The labels of the data are saved as well.

I. Feature Extraction

This study utilized two feature extraction methods: Principal Component Analysis (PCA) and a deep learning technique using the VGGNet model. Each method is discussed below.

1) Principal Component Analysis (PCA): PCA is used to reduce the dimensionality of the dataset by extracting the most informative features, specifically for image data in this study. The variance of image features based on the components is calculated to determine the optimal number of principal components for PCA reduction. The image dimensions are 48 x 48, resulting in 2304 components. A variance graph is plotted to identify the optimal number of principal components that account for 99% of the image variance and 881 components are found to be sufficient. Key Python packages used in the implementation include NumPy, sklearn, pickle, and matplotlib. The results of the variance graph reveal that the variance stops increasing after 1000 components, indicating diminishing returns in capturing additional variance beyond this point. 99% of the variance and 881 components are used for this project.

2) VVGNet: The VGGNet is a CNN architecture that is well known for its capability in image classification [26][27]. Due to its capability VGGNET is employed for the emotion recognition task and it consists of a deep convolutional neural network architecture that consists of four blocks, each containing a convolutional neural network layer, rectified linear unit (ReLU) activation function, batch normalization layer, max pooling layer, and dropout layer. The first block has a 3x3 filter size with 32 filters, while the second block has two 3x3 filters with 64 filters each. The third block has two 3x3 filters with 128 filters each. Finally, the fourth block flattens the output of the convolutional neural network layer and passes it to a dense layer with 128 neurons, followed by two more dense layers with 64 neurons each. The model is trained using a SoftMax classifier and L2 regularization and is specifically designed for the task of emotion recognition from images.

J. Classification

In the classification step, two different machine learning classification algorithms are applied in the study: Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). Each machine learning classification algorithm is discussed below.

1) SVM: SVM is one of the many machine learning classification tools; however, it is well known for being powerful in handling both classification and regression tasks. SVM creates a hyperplane to filter different classes and iteratively find the most fitting hyperplane that minimizes error. Three experiments were conducted in the study using SVM with linear, polynomial, and Gaussian kernels, and with penalty parameters ranging from 0.01 to 30. Multiple tests were performed to extract the most fitting penalty score, as shown in Table II.

TABLE II
DIFFERENT PENALTY PARAMETER TEST AND THEIR RESPECTIVE ACCURACY

Penalty Parameter (C)	Accuracy
0.01	0.6551
0.03	**0.6613**
0.1	0.6539
0.3	0.6490
1	0.6350
3	0.6521
10	0.6242
30	0.6300

The results showed that the Gaussian kernel performed the best, with a penalty score of C = 3, achieving an accuracy of 48%. The linear and polynomial models achieved a maximum accuracy of 39%. PCA was applied to reduce the number of features, which significantly reduced the training time. The best model with PCA used a Gaussian kernel with C = 10 and achieved 45% accuracy, slightly lower than without PCA. Finally, the VGGNet was used as a feature extraction method before training with SVM. The best model achieved 66% accuracy with a Gaussian kernel and a penalty parameter of

0.03. Based on the results, the VGGNet feature extraction method proved to be the most effective for SVM classification.

2) MLP: The study evaluated the performance of different feature reduction techniques using MLP. The MLP had an input layer, three hidden layers withholding a ReLU activation function, and one output layer withholding a SoftMax activation function. The study conducted three experiments using MLP, which are, first, training the MLP on the dataset without any feature reduction; second, applying PCA to reduce the dimensionality of the dataset and then training the MLP on the reduced dataset; and third, using the VGGNet architecture in conjunction with MLP. However, the best performance was achieved using SVM with VGGNet feature extraction, achieving 66% accuracy with a Gaussian kernel and a penalty parameter of 0.03. The MLP models achieved maximum accuracy of 43% and 59% without and with PCA, respectively, as shown in Table III.

TABLE III
MULTI-LAYERED PERCEPTRON MODEL

Epochs	30		50		75	
	Accuracy	TT (sec)	Accuracy	TT (sec)	Accuracy	TT (sec)
MLP Only	0.4341	66	0.4166	110	0.4288	166
PCA + MLP	0.4113	27	0.3954	46	0.4079	76
VGGNET + MLP	0.5912	7200	0.6300	14400	0.6242	18000

Overall, the results shown in Table III suggest that using VGGNet feature extraction in conjunction with SVM can lead to improved performance in image classification tasks.

K. Mental Health State Classification

The method described here utilizes detected emotions for seven days to determine the mental health state of a person. Scores are assigned to each of the seven emotions: disgust, fear, anger, surprise, happy, sad, and neutral. At the end of the seven-day evaluation, an aggregate score is calculated. If this aggregate score exceeds the threshold value of 50, the user is considered mentally unstable.

TABLE IV
FACIAL EMOTION AND MENTAL HEALTH STATE CORRELATION ADOPTED FROM SIMCOCK [28]

Mental Health	SP12 Depression
Anger	-0.16
Fear	-0.06
Happy	0.15
Neutral	0.40
Sad	-0.04
Disgust	-0.06
Surprise	0.08

The correlation between mental health and facial emotion from Simcock [28] is utilized to derive the mental health score assigned to various emotions. The table has no specified value for disgust and surprise because these emotions were not considered in this study. Therefore, the correlation value of fear is assigned to surprise and half the correlation value of happy to disgust. Table IV shows the correlation between facial emotion and SP12, a chronic type of depression adopted by Simcock [28].

Interpolation is used to derive the mental health score for each emotion, with scores ranging from 0 to 100. Considering that correlation values lie between -1 and +1, the interpolation formula shown in Equation 1 below is applied, where (x1, x2) = (-1, 1) represent the correlation values and (y1, y2) = (100, 0) represent the mental health state [28], [29].

$$y = y1 + (x - x1)\ \frac{(y2-y1)}{(x2-x1)} \tag{1}$$

Interpolation is used on another equation that can be seen below, where x represents the correlation value, and y represents the mental score (1) [26], [30]. The result is used as the mental health score for each emotion described in Table V.

$$\mathrm{MS} = 50 - 50\mathrm{CR} \tag{2}$$

where MS = Mental Health Score
CR = Correlation value

TABLE V
MENTAL SCORE DISTRIBUTION

Emotion	Mental Score
Disgust	53
Angry	58
Fear	53
Surprise	46
Happy	43
Sad	53
Neutral	30

The higher the score, the worse the mental health. The neutral emotion has the lowest score, while the angry emotion has the highest score. If a person exhibits anger for all seven days, they will have the highest possible aggregate score for their mental health state, indicating mental instability. In contrast, a neutral emotion for all seven days will produce the lowest aggregate score, suggesting mental stability.

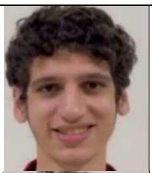
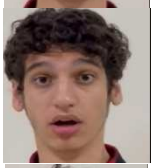
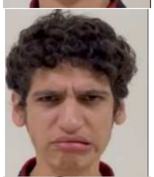
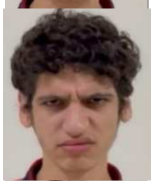
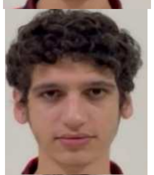
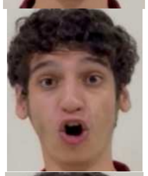
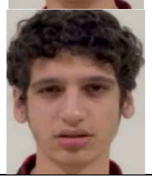
A directory containing seven images is sent to the model to evaluate the mental health score. For each image, the system first calculates the image's heights and width respectively using the OpenCV function, where it then extracts the ROI for an image, resizes the image to 48 x 48 dimensions, normalizes the image's pixels, and then converts the pixels to an array. It then passes the converted array for model prediction and returns the different emotions with a probability value, then the trained model finally picks the emotion with the highest probability as the image emotion prediction.

The mental health state of the user is then determined based on the mental score for each emotion over the seven-day period. If the aggregate mental score passes the threshold value of 50, the user is considered mentally unstable. In summary, the system developed in this study offers an approach for classifying mental health based on facial emotion detection, potentially improving mental health diagnosis and treatment.

III. RESULTS AND DISCUSSION

By collecting samples from two different users for seven consecutive days to perform emotion detection and thus, performing a mental health classification, the sample were tested throughout the different models for evaluation. The images collected from the first and second user are shown in Table VI and Table VII under the image's column. The images are collected through mobile devices. The emotion from these images is detected, and the result is shown in Table VI and Table VII under the emotion classification column. In Fig. 2 beneath Table VI and Fig. 3 beneath Table VII, each of the seven images classified emotions based on maximum probability of the emotion classification. The result is used to derive the mental health score for each emotion. The resulting mental health is calculated as the average of all seven emotions score. The average result helps determines if the user is stable or unstable based on the threshold score.

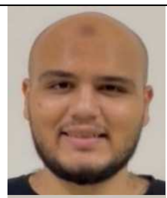
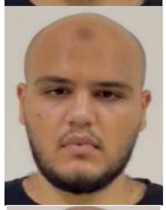
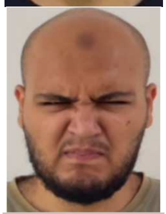
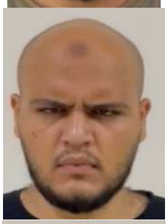
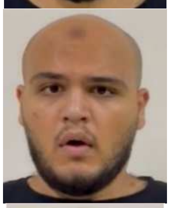
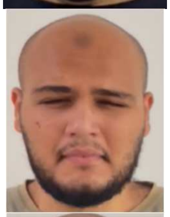
TABLE VI
SAMPLE IMAGES FOR EMOTION DETECTION FROM THE FIRST USER

Images	Emotion Classification	Mental Health Score
	angry: 0.10% disgust: 0.01% scared: 0.04% happy: 98.05% sad: 0.09% surprised: 0.05% neutral: 1.67%	Classification = Happy Mental Score = 43
	angry: 5.31% disgust: 0.22% scared: 7.58% happy: 2.07% sad: 4.73% surprised: 68.94% neutral: 11.15%	Classification = Surprised Mental Score = 46
	angry: 91.81% disgust: 0.58% scared: 1.24% happy: 0.21% sad: 3.59% surprised: 0.10% neutral: 2.47%	Classification = Angry Mental Score = 58
	angry: 87.15% disgust: 0.58% scared: 1.15% happy: 0.47% sad: 3.88% surprised: 0.14% neutral: 6.63%	Classification = Angry Mental Score = 58
	angry: 14.66% disgust: 0.21% scared: 1.03% happy: 13.61% sad: 5.31% surprised: 0.21% neutral: 64.97%	Classification = Neutral Mental Score = 30
	angry: 29.87% disgust: 0.49% scared: 5.05% happy: 1.67% sad: 0.69% surprised: 60.60% neutral: 1.63%	Classification = Surprised Mental Score = 46
	angry: 11.17% disgust: 0.30% scared: 1.76% happy: 4.98% sad: 8.19% surprised: 0.35% neutral: 73.26%	Classification = Neutral Mental Score = 30

```
Image  1 : happy , Mental Score:  43
Image  2 : surprised , Mental Score:  46
Image  3 : angry , Mental Score:  58
Image  4 : angry , Mental Score:  58
Image  5 : neutral , Mental Score:  30
Image  6 : surprised , Mental Score:  46
Image  7 : neutral , Mental Score:  30
Mental Health State = Stable
average score: 44.42857142857143
```

Fig. 2 Detected emotion for each emotion and final mental health score for the first user

TABLE VII
SAMPLE IMAGES FOR EMOTION DETECTION FROM THE SECOND USER

Images	Emotion Classification	Mental Health Score
	angry: 8.00% disgust: 0.36% scared: 5.60% happy: 22.29% sad: 13.86% surprised: 0.24% neutral: 49.65%	Classification = Neutral Mental Score = 30
	angry: 59.27% disgust: 0.89% scared: 10.09% happy: 0.21% sad: 17.82% surprised: 0.62% neutral: 11.09%	Classification = Angry Mental Score = 58
	angry: 96.97% disgust: 0.35% scared: 0.92% happy: 0.08% sad: 0.99% surprised: 0.06% neutral: 0.62%	Classification = Angry Mental Score = 58
	angry: 85.25% disgust: 0.37% scared: 3.70% happy: 0.11% sad: 5.95% surprised: 0.19% neutral: 4.43%	Classification = Angry Mental Score = 58
	angry: 59.57% disgust: 0.42% scared: 5.29% happy: 1.44% sad: 8.55% surprised: 0.40% neutral: 24.32%	Classification = Angry Mental Score = 58
	angry: 19.28% disgust: 0.57% scared: 9.11% happy: 4.00% sad: 46.84% surprised: 0.27% neutral: 19.92%	Classification = Sad Mental Score = 53
	angry: 0.03% disgust: 0.00% scared: 0.02% happy: 99.81% sad: 0.01% surprised: 0.02% neutral: 0.10%	Classification = Happy Mental Score = 43

```
Image  1 : neutral , Mental Score:  30
Image  2 : angry , Mental Score:  58
Image  3 : angry , Mental Score:  58
Image  4 : angry , Mental Score:  58
Image  5 : angry , Mental Score:  58
Image  6 : sad , Mental Score:  53
Image  7 : happy , Mental Score:  43
Mental Health State = Unstable
average score: 51.142857142857146
```

Fig. 3 Detected emotion for each emotion and final mental health score for the second user

Based on the mental health score of the first user, the aggregate score is 44.4, and it is determined that this user is mentally stable because it is less than the threshold value. However, the second user has a very high mental health score of 51.14, which shows the second user is unstable, at which the aggregate score is above the threshold value. By witnessing the results, the suggested approach proves that it has the quintessential capabilities to calculate and classify one's mental health state. In addition to its simplicity for reaching that classification, it can be used by any individual, given that they provide seven images.

IV. Conclusion

Different classification algorithms were implemented in this project to determine the best one suited for this problem. The best-performing model used VGGNet as the feature extraction method, and the classification was done using SVM. Thus, in evaluating this model, one of the limitations faced has been the inability to take samples at times unknown to the user. This system is to be used in the background of mobile devices, and images are taken randomly during the day while the user is active on the phone. In addition, the training sample of the dataset is not evenly distributed, and the disgusted emotion sample has extraordinarily little compared to the other sample. This made classifying images disgustingly difficult for the model. This project can be further improved by implanting other feature extraction methods to extract more facial image features. In addition, the system can be integrated with mobile devices to show and evaluate its applicability in real-world scenarios. Furthermore, it is worth noting that the number of representative images chosen does not influence the accuracy of the whole model.

Acknowledgment

We affirm that the participants provided informed consent for the publication of the images in Tables VI and VII.

References

[1] S. Tivatansakul, M. Ohkura, S. Puangpontip, and T. Achalakul, "Emotional healthcare system: Emotion detection by facial expressions using Japanese database," 2014 6th Computer Science and Electronic Engineering Conference (CEEC), Sep. 2014, doi:10.1109/ceec.2014.6958552.

[2] M. Smith, R. S., "Improving Emotional Health," May 15, 2014 [Online]. Available: http://www.helpguide.org/mental/mental_emotional_health.html. [Assessed Jan. 5, 2023].

[3] E. A. Elliott and A. M. Jacobs, "Facial Expressions, Emotions, and Sign Languages," Frontiers in Psychology, vol. 4, 2013, doi:10.3389/fpsyg.2013.00115.

[4] J. M. Garcia-Garcia, V. M. R. Penichet, and M. D. Lozano, "Emotion detection," Proceedings of the XVIII International Conference on Human Computer Interaction, Sep. 2017, doi:10.1145/3123818.3123852.

[5] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," Proceedings of the National Academy of Sciences, vol. 111, no. 15, Mar. 2014, doi:10.1073/pnas.1322355111.

[6] X. Chen, S. Sun, H. Li, Z. Ma, and K. Zhang, "Better than humans: a method for inferring consumer shopping intentions by reading facial expressions," 2021 14th International Symposium on Computational Intelligence and Design (ISCID), Dec. 2021, doi:10.1109/iscid52796.2021.00041.

[7] E.-M. Seidel, U. Habel, M. Kirschner, R. C. Gur, and B. Derntl, "The impact of facial emotional expressions on behavioral tendencies in women and men.," Journal of Experimental Psychology: Human Perception and Performance, vol. 36, no. 2, pp. 500–507, 2010, doi:10.1037/a0018169.

[8] E. Back and T. R. Jordan, "Revealing Variations in Perception of Mental States from Dynamic Facial Expressions: A Cautionary Note," PLoS ONE, vol. 9, no. 1, p. e84395, Jan. 2014, doi:10.1371/journal.pone.0084395.

[9] B. Zhang, "Computer vision vs. human vision," 9th IEEE International Conference on Cognitive Informatics (ICCI'10), Jul. 2010, doi:10.1109/coginf.2010.5599750.

[10] R. Sathya and A. Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification," International Journal of Advanced Research in Artificial Intelligence, vol. 2, no. 2, 2013, doi: 10.14569/ijarai.2013.020206.

[11] P. Pandey, K. K. Dewangan, and D. K. Dewangan, "Enhancing the quality of satellite images by preprocessing and contrast enhancement," 2017 International Conference on Communication and Signal Processing (ICCSP), Apr. 2017, doi:10.1109/iccsp.2017.8286525.

[12] M. Huang, M. Zhuang, J. Zhou, and X. Wu, "Preprocessing Method Flow of Under-screen Fingerprint Image," 2020 IEEE 14th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Oct. 2020, doi:10.1109/asid50160.2020.9271693.

[13] Priyanka C. Dighe, S. K. (2014). Survey on Image Resizing Techniques. *International Journal of Science and Research (IJSR)*, 1444–1448. ISSN: 2319-7064

[14] Surasak, T. T. (2018). Histogram of oriented gradients for human detection in video. Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business and Social Science, 172–176. doi:10.1109/ICBIR.2018.8391187

[15] L. Zhang, W. Zhou, J. Li, J. Li, and X. Lou, "Histogram of Oriented Gradients Feature Extraction Without Normalization," 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Dec. 2020, doi: 10.1109/apccas50809.2020.9301715.

[16] M. Kolahdouzi, A. Sepas-Moghaddam, and A. Etemad, "Face Trees for Expression Recognition," 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Dec. 2021, doi: 10.1109/fg52635.2021.9666986.

[17] H. Zhao, Q. Wang, Z. Jia, Y. Chen, and J. Zhang, "Bayesian based Facial Expression Recognition Transformer Model in Uncertainty," 2021 International Conference on Digital Society and Intelligent Systems (DSInS), Dec. 2021, doi: 10.1109/dsins54396.2021.9670628.

[18] J. Li, B. Zhao, H. Zhang, and J. Jiao, "Face Recognition System Using SVM Classifier and Feature Extraction by PCA and LDA Combination," 2009 International Conference on Computational Intelligence and Software Engineering, Dec. 2009, doi: 10.1109/cise.2009.5364125.

[19] S. D. Fabiyi et al., "Comparative Study of PCA and LDA for Rice Seeds Quality Inspection," 2019 IEEE AFRICON, Sep. 2019, doi: 10.1109/africon46755.2019.9134059.

[20] F. Ayeche and A. Alti, "Face Recognition and Facial Expression Recognition Algorithm Based on Novel Features Extraction Descriptors HDG and HDGG," 2021 International Conference on Information Systems and Advanced Technologies (ICISAT), Dec. 2021, doi: 10.1109/icisat54145.2021.9678425.

[21] M. Ahmad et al., "Hyperspectral Image Classification—Traditional to Deep Models: A Survey for Future Prospects," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 968–999, 2022, doi: 10.1109/jstars.2021.3133021.

[22] R. Gill and J. Singh, "A Deep Learning Approach for Real Time Facial Emotion Recognition," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), Dec. 2021, doi: 10.1109/smart52563.2021.9676202.

[23] A. Sultana, S. K. Dey, and Md. A. Rahman, "A Deep CNN based Kaggle Contest Winning Model to Recognize Real-Time Facial Expression," 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Nov. 2021, doi: 10.1109/iceeict53905.2021.9667824.

[24] S. Tegani and T. Abdelmoutia, "Using COVID-19 Masks Dataset to Implement Deep Convolutional Neural Networks For Facial Emotion Recognition," 2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Dec. 2021, doi: 10.1109/isaect53699.2021.9668345.

[25] C.-Y. Chang, C.-W. Chang, J.-Y. Zheng, and P.-C. Chung, "Physiological emotion analysis using support vector regression,"

Neurocomputing, vol. 122, pp. 79–87, Dec. 2013, doi: 10.1016/j.neucom.2013.02.041.

[26] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," Journal of Informatics and Web Engineering, vol. 1, no. 2, pp. 45–54, Sep. 2022, doi: 10.33093/jiwe.2022.1.2.4.

[27] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," The Visual Computer, vol. 36, no. 2, pp. 405–412, Jan. 2019, doi: 10.1007/s00371-019-01630-9.

[28] G. Simcock et al., "Associations between Facial Emotion Recognition and Mental Health in Early Adolescence," International Journal of Environmental Research and Public Health, vol. 17, no. 1, p. 330, Jan. 2020, doi: 10.3390/ijerph17010330.

[29] Ashish Kumar Srivastav, D. V., "Interpolation (Definition, Formula) | Calculation with Examples," March 26, 2023, Available: https://www.wallstreetmojo.com/interpolation/. [Assessed April 23, 2023].

[30] T. Hastie, A. Montanari, S. Rosset, and R. J. Tibshirani, "Surprises in high-dimensional ridgeless least squares interpolation," The Annals of Statistics, vol. 50, no. 2, Apr. 2022, doi: 10.1214/21-aos2133.