Article

Demystifying Mental Health by Decoding Facial Action Unit Sequences

Deepika Sharma [1], Jaiteg Singh [1,*], Sukhjit Singh Sehra [2,*] and Sumeet Kaur Sehra [2]

1   Chitkara University Institute of Engineering & Technology, Chitkara University, Rajpura 140401, Punjab, India
2   Physics and Computer Science, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada
*   Correspondence: jaiteg.singh@chitkara.edu.in (J.S.); ssehra@wlu.ca (S.S.S.)

Abstract: Mental health is indispensable for effective daily functioning and stress management. Facial expressions may provide vital clues about the mental state of a person as they are universally consistent across cultures. This study intends to detect the emotional variances through facial micro-expressions using facial action units (AUs) to identify probable mental health issues. In addition, convolutional neural networks (CNN) were used to detect and classify the micro-expressions. Further, combinations of AUs were identified for the segmentation of micro-expressions classes using K-means square. Two benchmarked datasets CASME II and SAMM were employed for the training and evaluation of the model. The model achieved an accuracy of 95.62% on CASME II and 93.21% on the SAMM dataset, respectively. Subsequently, a case analysis was done to identify depressive patients using the proposed framework and it attained an accuracy of 92.99%. This experiment revealed the fact that emotions like disgust, sadness, anger, and surprise are the prominent emotions experienced by depressive patients during communication. The findings suggest that leveraging facial action units for micro-expression detection offers a promising approach to mental health diagnostics.

Keywords: micro-expressions; CNN; K-means; emotion recognition; action units

1. Introduction

Mental health could be understood as a condition of optimal wellness where individuals are able to adeptly handle their daily stressors, meet their emotional needs, and make valuable contributions to their professional and societal spheres (World Health Organization (WHO)). The mental well-being of an individual may be influenced by a variety of elements, encompassing environmental factors, socioeconomic circumstances, personal attributes, and genetic predisposition. Symptoms such as stress, anxiety, and melancholy stand out as key signs of declining mental health [1]. The growing awareness about mental health underscores the urgent need for efficient and reliable methods of identification of mental problems and timely intervention. Conventional self-reporting techniques and clinical interviews for estimating mental health could be expensive, time-consuming, subjective, and biased [2]. There is potential for early identification and better diagnosis of mental issues. Subsequently, effective therapies to support mental well-being by investigating alternative strategies could be explored. Understanding facial expressions and the use of machine learning for emotion identification are potential alternatives for estimating mental health. Communication is indispensable for every aspect of humanness. In addition, spoken words, facial expressions, body language, eye movements, and other nonverbal cues do convey a lot of information about an individual's behavior, emotional state, and temperament [3]. The relationship between facial expressions and mental health has been very complex. Individual differences, cultural differences, social contexts, and experiences can influence the interpretation of emotions. The human capacity for emotional concealment makes this analysis even more complex. An individual experiencing depression, for example, may adopt a facade of happiness to mask internal turmoil. It is imperative to utilize objective methods

to decipher the often-enigmatic language of facial expressions to understand mental health. A wide spectrum of thoughts, emotions, and intentions are communicated through facial expressions. Despite cultural and linguistic differences among individuals, most emotions are conveyed through similar facial expressions. These facial expressions are triggered by coordinated contractions of various facial muscles, either in dynamic movement or static postures. Psychological researchers have identified six basic emotions consistently associated with distinct facial expressions: happiness, surprise, fear, anger, sadness, and disgust [4]. Broadly, facial expressions can be categorized into two categories, viz. macro and micro-expressions. Macro-expressions are prominently visible because they involve entire facial muscles and last for two to three seconds. Micro-expressions involve the same facial muscles as macro-expressions but with subtler movements. Micro-expressions last for merely 1/50th to 1/200th of a second and are difficult to identify. This transience makes micro-expressions challenging to detect in real-time conversations with the naked eye, requiring skilled and attentive observers. The significance of micro-expressions lies in their potential to reveal underlying emotions masked by deliberate expressions. However, a major obstacle to developing accurate automated detection tools is the possibility of micro-expressions being confined to a single, localized area of the face [5].

Facial action units (AUs) play a crucial role in providing a reliable and unbiased method of observing facial features. AUs offer a systematic and quantified approach to evaluating facial expressions by dissecting facial motions. Facial AUs refer to minuscule movements of specific facial muscles that contribute to a more detailed comprehension of emotions. These subtle distinctions separate authentic emotions from contrived ones [6]. Facial AUs offer a distinct perspective for evaluating mental well-being, as different combinations of AUs are linked to different mental health disorders. For instance, anxiety is associated with heightened activity in AU12 and AU17, whereas depression is characterized by reduced activity in AU6 and AU12 [7,8]. Examining the interaction between AUs might yield significant observations regarding an individual's psychological condition, facilitating prompt identification, tracking of treatment progress, and tailored healthcare. Table 1 shows the significantly used action units for emotion recognition with their related muscle movement area on the face [9]:

Table 1. Facial action units with their description.

Action Units	Facial Muscle Description	Action Units	Facial Muscle Description
AU1	Inner brow raiser	AU14	Dimpler
AU2	Outer brow raiser	AU15	Lip Corner depressor
AU4	Brow lowerer	AU16	Lower lip depressor
AU5	Upper lid raiser	AU17	Chin raiser
AU6	Cheek raiser	AU20	Lip stretcher
AU7	Lid tightener	AU24	Lip pressor
AU9	Nose wrinkler	AU23	Lip tightener
AU10	Upper lip raiser	AU25	Lip part
AU11	Nasolabial deepener	AU26	Jaw drop
AU12	Lip corner puller	AU43	Eyes closed
AU13	Cheek puffer		

By deciphering facial expressions, one can transcend subjective interpretations and get profound insight into the connection between the internal state and external manifestations of an individual [10]. The identification of micro-expressions (MEs) may contribute across a wide range of domains, including mental health, lie detection, law enforcement, political psychology, medical care, and human–computer interaction [11]. In contrast to conventional methods that rely on the extraction of hand-crafted features and the implementation of

custom methods to analyze these features, this paper introduces a deep-learning approach to identify micro-expressions (MEs) based on facial action units (AUs).

Enabling mental health professionals to detect potential mental disorders at an earlier stage may help in designing effective custom treatment plans. Machine learning has revolutionized the mental health domain by employing facial AUs to comprehend and forecast emotions and mental states. Such integration of AUs and the computational capabilities of machine learning can revolutionize clinical diagnosis of mental health issues and timely intervention. The intricate facial expressions contain vital information regarding the emotions and cognitive conditions of an individual. Machine-learning algorithms, which have been trained on extensive datasets containing labeled facial expressions, can identify intricate patterns in the activation of AUs.

The rest of this paper is organized as follows. Section 2 reviews the existing research on action unit recognition and emotion classification using machine-learning algorithms. Section 3 elaborates on the materials and methodologies employed during the experimentation. Section 4 presents the findings and discusses their significance. Finally, Section 5 concludes the findings and summarizes the key takeaways of this research.

2. Related Work

Ensuring accurate diagnosis of mental health issues remains a critical concern. As stated earlier, conventional approaches, while helpful, can suffer from drawbacks, such as subjectivity in self-reporting, cost constraints, social biases, time-consuming procedures, or scheduling conflicts in clinical interviews. Recent studies have explored the possibility of using facial expressions, particularly brief micro-expressions, as bio-markers to identify the emotional states to overcome these constraints. This related work explores the recent research in automated mental health detection using facial expressions, facial action units, and machine learning. Additionally, the various methodologies employed for the identification of action units were also reviewed.

An automatic facial AU recognition and stress detection model was proposed using the support vector regression algorithm (SVR). The model was trained on BOSPHORUS and UNBC datasets and tested on experimental dataset SRD'15, having stressed and normal states. According to the study, the AUs that are under investigation when stress conditions arise are AU17, AU25, AU1, AU7, and AU26, which are the most modified and can distinguish between different emotional states [8].

To identify the Depression Anxiety Stress Scale (DASS) levels, a support vector machine (SVM) machine-learning-based model was proposed to analyze facial expressions using action units. It was observed that the depressed emotional state has two sadness and happiness emotions with AU6, AU12, AU15, and AU26. Anxiety has surprise and disgust emotions with AU2, AU9, AU25, and AU45. Stress was measured with sadness and disgust emotions with AU1, AU6, AU12, and AU15. The model achieved 87.2%, 77.9%, and 90.2% accuracy on depression, anxiety, and stress, respectively [7].

Facial expressions were used to identify depression and anxiety through a CNN-based approach VGG network. The FER+ dataset was used for the experiment. HAAR cascade was used for feature extraction from the facial images. The VGG model was used for normal and abnormal states as positive and negative emotion classification. Further, the negative emotions predict whether a person is depressed or has anxiety. The overall framework has achieved 95% accuracy [12].

SVM and random forest (RF) machine-learning techniques were employed to develop an algorithm to validate the identified facial expressions with seven emotions. A web application was used as stimuli to show the 70 images randomly to the participants. At a time, a single participant was shown the 35 images' subset for ten minutes with the option buttons having happy, sad, and other emotions. The facial expressions of a total of 31 participants were captured as they observed the displayed images. The participants were then instructed to select the options that best described their emotions. Furthermore, the FeelPix labeled dataset was developed from these participants' recorded facial expressions

based on left meningeal, right meningeal, nasal center, and subnasal center positions. The experiment shows promising results in the form of accuracy, precision, F1-score, and G mean for different emotions. Overall, 74% accuracy has been achieved while testing the algorithm [13].

A long-short term memory (LSTM) algorithm-based depression detection framework was proposed, using facial expressions and action units. The video dataset of depressed and non-depressed samples was used for the experiment. An accuracy of 91.67% was achieved for depression detection. The key findings of this framework are that depressed subjects have shown movements in AU26, AU20, and AU07. Non-depressed facial movements were captured by AU06, AU25, AU14, and AU12 [14].

Micro-expression recognition based on action units is divided into two parts i.e., feature extraction and emotion classification. The goal of ME recognition, such as video ME sequence "s", which consists of "n" frames with fixed classes set "k", is to determine which class each frame belongs to. The main emphasis is on classes that have a high probability of $p(c_j \mid s)$ among the classes with a probability of "k". The overall structure of this process is shown in the Figure 1.
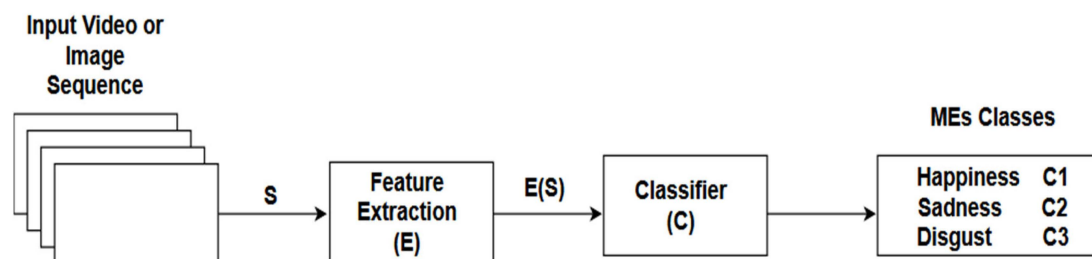
Figure 1. Basic architecture of ME classification.

Based on problem formulation to find the probability, techniques are divided into four categories.

2.1. Conventional Techniques

In conventional techniques, hand-crafted features are employed, and the variance between the features is evaluated to identify expression segments from non-expression segments. Temporal and spatial dimensions are included in the model. The appearance of facial details, which represents the variance in pixel intensity like texture, or the geometric facial information, which includes the forms and placements of facial landmarks, are the two bases on which the features are extracted. These traditional approaches include local binary pattern on three orthogonal planes (LBP-TOP) [15], histogram of oriented optical flow (HOOF) [16], main directional mean optical flow (MDMO) [17], and bi-weighted oriented optical flow (Bi-WOOF) [18]. Feature extraction through LBP-TOP is a combination of LBP histograms of three orthogonal planes named XY, YT, and XT. Histogram concatenation results in duplicate information when neighboring pixels are utilized multiple times in the LBP-TOP computation. Through the extraction of some AUs' substantial optical flow direction, MDMO is employed to characterize localized facial dynamics.

Despite these methods, the mean oriented Riesz features (MORF) technique uses the Riesz pyramid to describe the MEs' evolution in two frames, named MOR image sets, which is then utilized to create a histogram [19]. A 3D histogram of oriented gradient (HOG) is used to recognize the MEs by applying it to twelve facial areas. A fusion of motion boundary histogram (FMBH) is created by combining motion boundary histograms constructed using norms with angles calculated from the optical flow's horizontal and vertical components [20].

2.2. Deep-Learning Techniques

Deep-learning approaches are commonly used to solve tasks related to computer vision, such as detecting objects, as well as tracking and video labeling, including image classification. Numerous researchers have evaluated and used the deep-learning framework for ME analysis because of the deep-learning models' outstanding results on these types of assignments. Most innovative deep-learning systems are built using convolutional neural network (CNN) variants or the combination of CNN and recurrent neural network (RNN). A CNN framework was proposed for ME recognition by employing pre-trained weights for feature extraction trained on ImageNet CK+ and SPOS datasets [21,22]. A lateral accretive hybrid network (LEAR-Net) was proposed; this was a CNN-based system that used an accretion layer to enhance the network's learning capacity, hence improving the salient expression aspects [23]. In another proposed framework, long-short term memory (LSTM) and CNN were employed for temporal and spatial feature extraction. Although accuracy advancements have been considerable for most computer vision applications, advancements in ME evaluation have been relatively moderate [24].

2.3. Hybrid Techniques

Hybrid techniques utilize the conventional and deep-learning approaches to produce better outcomes. Different optical flow methods are used for feature extraction using CNN. Off-ApexNet was proposed for ME recognition by using only onset and apex frames through optical flow for feature extraction. Subsequently, a shallow triple stream 3D CNN was proposed to show the improvements in Off-ApexNet by adding the vertical and horizontal optical flow [24]. An enriched long-term recurrent convolutional network (EL-RCN) was proposed for the computation of optical flow and strain for training the spatial features. In this framework, two approaches were used, namely spatial dimension enrichment and temporal dimension enrichment using the VGG-16 model. For classification, the LSTM algorithm was employed. A spatiotemporal recurrent convolutional network was proposed with two variations. First, based on appearance connectivity in one-dimensional vectors and second, with geometrical connectivity of optical flow [25]. The implementation of conventional approaches to streamline the extraction of spatiotemporal features and the process of classification using deep-learning architecture is an advantage of the techniques in this area.

2.4. Region of Interest-Based Techniques

The previously mentioned techniques focus on the improvement of feature extraction, but region-based strategies target the locality segment in data preprocessing. Instead of analyzing the entire face of MEs, region-of-interest approaches are employed to retrieve reliable and pertinent information [26]. In [27], six distinct regions were originally identified including right and left cheeks, eyes and eyebrows, the nose, and the mouth. In another investigation, a manual selection of active patches was captured through eighteen regions of interest (ROIs). To identify the necessary morphological patches (NMPs), a weight was assigned to every active patch. Based on the characteristics that LBP-TOP has extracted, the weights were calculated using the entropy-weight approach [28]. In subsequent research, rather than selecting the eighteen action patches manually, the six regions were split into smaller blocks to obtain 106 active patches. The random forest (RF) technique was applied to features extracted from action patches using LBP-TOP and optical flow for NMPs [29].

3. Materials and Methods

3.1. Datasets

This study used two datasets—Spontaneous Actions and Micro-Movements (SAMM) and Chinese Academy of Sciences Micro-expression II (CASME II), which are the two most prominent datasets used in micro-expression based studies.

The SAMM dataset is the most recent high-resolution collection of 159 spontaneously produced micro-movements with the greatest demographic heterogeneity. The dataset

was designed to be as varied as practicable to capture a wide range of emotional reactions. A total of 32 participants, with an average age of 33.24 years, were recruited for the experiment, representing an equal number of male and female participants. Based on the seven fundamental emotions, the induction process was captured at 200 frames per second (fps). The resolution of the facial region is around 400×400 pixels, while the overall resolution is 2040×1088 pixels. The SAMM dataset has seven classes including anger, happiness, surprise, contempt, fear, disgust, and others [30].

The CASME II dataset is an improved version of CASME. Every sample in this dataset has spontaneous and dynamic micro-expressions. The CASME II dataset is a collection of 249 samples at 200 frames per second. The total number of participants was 35 (male + female). The resolution of the facial region is around 280×340 pixels, while the overall resolution is 640×480 pixels. The CASME II dataset is divided into five categories, such as happiness, surprise, repression, disgust, and others [31].

Table 2 provides information about the purpose, methodology employed, outcomes, accuracy, and the exploration area for both datasets. The CASME II and SAMM datasets were utilized to identify micro-expression and emotion recognition based on facial muscle movements, action units, facial skin deformations, elasticity, and landmark feature maps.

Table 2. Analysis of the CASME II and SAMM datasets.

| Ref.	Purpose	Dataset	Accuracy (%)	Methodology Used	Outcomes	Exploration Areas
[31]	Micro-expression recognition	CASME II	63.41	LBP-TOP, SVM, LOSO–cross validation technique	ME dataset has been created with 5 class emotions labeled, including action units for 3000 facial muscle movements.	ME variations can be explored in different environments
[32]	Micro-expression recognition	CASME II SAMM	5-class, 3-class 83.3, 93.2 79.4, 86.5	AU-class activation map (CAM), area weighted module (AWM)	ME recognition with 5 classes (disgust, happiness, repression, surprise, others) and 3 class categories (negative, positive, surprise).	Graph structure can be used to extract features using facial landmarks, instead of the whole face area.
[30]	Micro-facial movements for objective class AUs	SAMM	Recall 0.91	LBP-TOP, 3D HOG, deformable part models, LOSO	The proposed dataset can be used to train a model for deception detection, emotion recognition, lie detection.	Optical flow and unsupervised clustering techniques can be incorporated to capture micro-movements.
[33]	Micro facial expression recognition	CASME II	70.20	Local motion patterns (LMP)	LMP features were extracted to measure the facial skin elasticity and deformations.	Head movements, non-frontal poses and occlusion can enhance the model accuracy.
[34]	Micro-expressions and emotion recognition	CASME II	73.98	CNN, LSTM, landmark feature map (LFM)	LFM predicted MEs in positive, negative and surprised categories. LFM calculates the proportional distances between landmarks.	Including facial expression intensity, texture features could be employed to achieve high accuracy.

3.2. Proposed Framework

The following procedures are part of the suggested framework for the recognition and classification of micro-expressions utilizing action units: pre-processing of the data, recognition of action units, classification of emotions, and micro-expression subdivision, shown in Figure 2.

3.2.1. Data Pre-Processing

Figure 2 illustrates the proposed framework. Initially, it invokes a two-step data preparation process to pre-process input images for further evaluation. Data pre-processing primarily includes face detection and detected facial image enhancement.

The initial step involves the identification of faces within the given video sequence by employing a pre-trained Haar cascade classifier. Further, the Haar cascade classifier provides bounding boxes and pinpoints the detected faces within an image. This, in turn, facilitates the creation of a uniform data format for subsequent analysis. This uniformity is achieved by enabling the extraction of facial regions with consistent proportions and alignment, as described in [35]. Subsequently, the contrast limited adaptive histogram

equalization (CLAHE) model is invoked for image enhancement. This method augments the contrast within specific facial areas, particularly in images characterized by uneven lighting and low contrast levels. CLAHE operates by dividing the image into smaller sub-regions (tiles) and applying localized histogram equalization to each tile. This approach enhances contrast within each region while minimizing the overall contrast change, thereby mitigating the formation of visual artifacts [36].
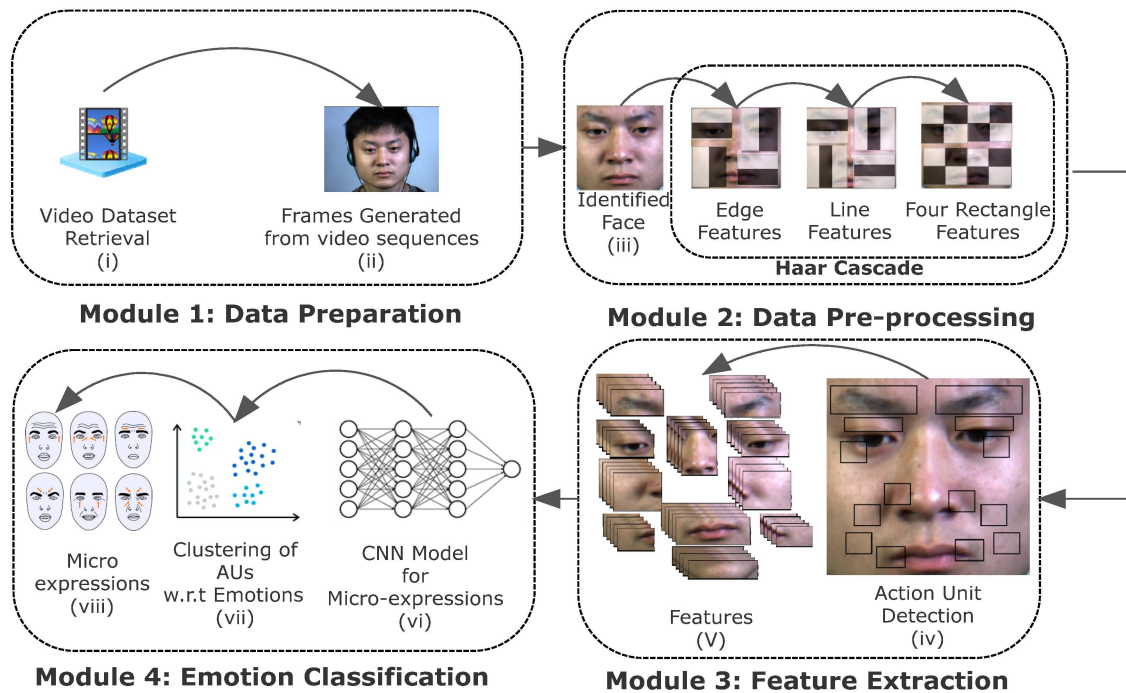
Figure 2. Proposed framework architecture (adapted with permission from ©Xiaolan Fu [31]).

3.2.2. Action Unit Detection

A pre-trained model that has been trained on the facial action coding system (FACS) is used to detect the action units from pre-processed images. During evaluation, the model collects expressions, such as lip configurations, eyebrow movements, and wrinkle patterns that are suggestive of activation of AUs. Embedded computer vision algorithms in the pre-trained model automatically detect the intensity and presence of AUs [26].

3.2.3. Emotion Classification

Comprehensive representations of encountered emotions are provided by the identified AUs. The purpose of this model is to transform the combination of activated AUs into recognizable emotions, such as surprise, happiness, sadness, anger, etc. We use the identified AUs as an abundant collection of characteristics for an independent deep-learning model intended for the categorization of emotions. This model is specially trained to map combinations of AUs to discrete emotion categories. The model intakes a large dataset of annotated facial images during training. Each image has a related active AU (such as happiness, sadness, or anger) that corresponds to that emotion. The program learns the complex links between different AUs and how they are associated with different emotions by analyzing this data. For example, the model may detect that displays of happiness often co-occur with a combination of AU 6 (raised cheeks) and AU 12 (lip corners pushed up). On the other hand, AU 7 (lid tightener) and AU 4 (brow furrow) might represent sadness. During this phase of learning, the model creates a strong association between emotional states and AU pairings.

3.2.4. Action Units Combinations

K-means clustering of the action units is performed by implementing the specific emotion and the associated activated action unit information from the previous step as shown in Figure 2. This will provide the clustering of different combinations of action units as ME1, ME2, to ME_n based on silhouette analysis. It assesses an object's similarity to its cluster concerning other clusters. The silhouette score varies between -1 to 1, which indicates a low value for inadequate match to nearby clusters and a high value for appropriate grouping of its own cluster. The formula used for the calculation of silhouette score at a particulate point is:

$$s(k) = \frac{b(k) - a(k)}{\max\{a(k), b(k)\}}$$

where $s(k) \geq$ silhouette score for a single data point k

a(k) $\geq$ average distance between K to other points (within the same cluster)
b(k) $\geq$ smallest average distance of K (to different clusters)
To calculate the average score for the dataset, calculated by:

$$Average\ Silhouette\ Score = \frac{1}{N} \sum_{k=1}^{N} =s(k)$$

where N $\geq$ total number of points in the dataset.

3.3. Model Architecture

The proposed framework used a customized CNN architecture that emerged especially to detect action units (AUs) in facial images. This model maintains superior results while giving priority to lightweight design.

The input layer that the model receives has a predetermined form, such as $8 \times 8 \times 512$, where 8 stands for the width and height of resized frames in pixels. The images are pre-processed to generate a high-dimensional feature vector having 512 channels [37]. The benefit of a high-dimensional feature vector is that it captures the essential information from the video frames. A zero-padding layer adds padding pixels around the borders of the input image. This helps prevent information loss during the convolution operation. To extract initial features from the padded picture, a 7×7 convolutional layer with 16 filters and a stride of 2 pixels is used. Stride 2 reduces the dimensionality of the image by having the filter move by 2 pixels in both the horizontal and vertical directions following each convolution process.

By stabilizing the neuronal activations, batch normalization contributes to the stabilization of the learning process. By introducing non-linearity into the network using a ReLU (rectified linear unit) activation function, the model can learn more intricate correlations between features. The most important characteristics are retained while the picture dimensionality is further reduced by a subsequent 3×3 max pooling layer with a stride of 2 pixels. Max pooling takes the maximum value from a defined window (here, 3×3) and moves the window across the image with a stride of 2.

The RepeatLayers function uses multiple convolutional layers, which enhances the feature extraction at each layer. The main motive for adding multiple layers is that the usage of features extracted at the initial layer will be added to subsequent layers with high-dimensional features. Using this approach, the network will gradually learn more intricate visualizations of the input [38,39]. This function stacks two convolutional layers with the same number of output filters but different filter sizes (e.g., 3×3). After every convolutional layer, batch normalization, and ReLU activation are used. Including a shortcut connection is an important feature. Because of this link, the model will immediately add the first convolutional layer's output to the second convolutional layer's output inside the RepeatLayers block. Particularly in deeper networks, this aids in gradient flow and learning. To gradually capture increasingly complicated characteristics, the design uses the

RepeatLayers function twice, with progressively more filters, such as 16 to 16, then 32 to 32. In addition, this function projects the shortcut connection that ensures the dimensions are accurately aligned. In case the number of filters in the shortcut connection differs from the number of filters in the output of the convolutional layers, a 1×1 convolutional layer is applied to the shortcut to bring the number of filters in line.

The collected features are combined over the feature maps' whole spatial dimension using a global average pooling layer. In other words, it creates a single feature vector for each image by averaging the values from each channel throughout the feature map's width and height. Finally, a dense layer that is completely linked and has as many neurons as the number of AUs being targeted is applied. The output layer generates the final output probabilities for each AU using a softmax activation function. The output values of the softmax function are guaranteed to total up to 1 and reflect probabilities between 0 and 1. Table 3 presents the network configuration of the proposed model.

Table 3. Proposed model's network configuration.

Layer Type	Output Shape	# of Parameters
Conv2D (7×7, 16 filters)	(64, 64, 16)	2368
BatchNormalization	(64, 64, 16)	64
MaxPooling2D	(31, 31, 16)	0
Conv2D (3×3, 16 filters)	(31, 31, 16)	2320
BatchNormalization	(31, 31, 16)	64
Conv2D (3×3, 16 filters)	(31, 31, 16)	2320
BatchNormalization	(31, 31, 16)	64
Add	(31, 31, 16)	0
Conv2D (3×3, 32 filters)	(31, 31, 16)	4640
BatchNormalization	(31, 31, 32)	128
Conv2D (3×3, 32 filters)	(31, 31, 32)	9248
BatchNormalization	(31, 31, 32)	128
Con2d (1×1, 32 filters)	(31, 31, 32)	544
BatchNormalization	(31, 31, 32)	128
Add	(31, 31, 32)	0
GlobalAveragePooling2D	(32,)	0
Dense (output layer)	(7,)	231
Total	**-**	**22,247**

Performance Metrics

The accuracy (ACC) measure shows the proportion of test dataset emotions that were properly identified overall. It provides a general understanding of the model's ability to map detected AUs to pre-defined emotional categories. The mathematical expression for this is:

$$ACC = (TP + TN)/(TP + TN + FP + FN)$$

where TP is true positive, the total number of emotions that are accurately identified, TN is true negatives, the total number of accurately categorized non-emotions, FP is false positives, the number of emotions that are misclassified (e.g., disgust when contempt), and FN is false negatives, the number of emotions that the model failed to recognize (missing a fear expression).

Precision (P) quantifies the proportion of true positives, or emotions that are accurately categorized, to all positive predictions, or images that the model recognized as correspond-

ing to a certain emotion. The model's ability to prevent false positives—the inaccurate classification of emotions—is shown by a high accuracy. One way to calculate it is:

$$P = TP/(TP + FP)$$

Recall (R) is the percentage of accurately identified emotions, or true positives, among all true positive examples (images that are in the test set containing the intended emotion). Most of the pertinent emotions found in the data can be identified by the model with a high recall. It is comprehensible as:

$$R = TP/(TP + FN)$$

F1-score (F1) offers a balanced perspective on recall and precision by taking the harmonic mean of both variables. It considers the model's capability to detect true positives as well as its ability to prevent false positives. It is computed as follows:

$$F1 = 2 \times (P \times R)/(P + R)$$

3.4. Micro-Expression Sub-Division

To make clusters (sub-divisions) of the different micro-expressions, such as happiness, sadness, anger, etc., the K-means clustering algorithm is employed. The K-means clustering technique is widely used to categorize unlabeled data elements into discrete groups. The data elements are progressively assigned to the cluster corresponding to the closest mean as a centroid in an iterative manner. The objective of this method is to form clusters that are both well-separated from one another as well as internally homogenous.

4. Result and Discussion

4.1. Results

4.1.1. Emotion Classification

A customized CNN model architecture is used for emotion classification, employing the datasets CASME II, with 5 classes, and SAMM, with 7 classes. To evaluate the model's performance to classify emotions, a confusion matrix is employed, which provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. Further, the model is assessed using standard metrics like accuracy, precision, recall, and F1-score to determine emotions based on the observed AUs. Figures 3 and 4 show the confusion matrix for CASME and SAMM datasets:
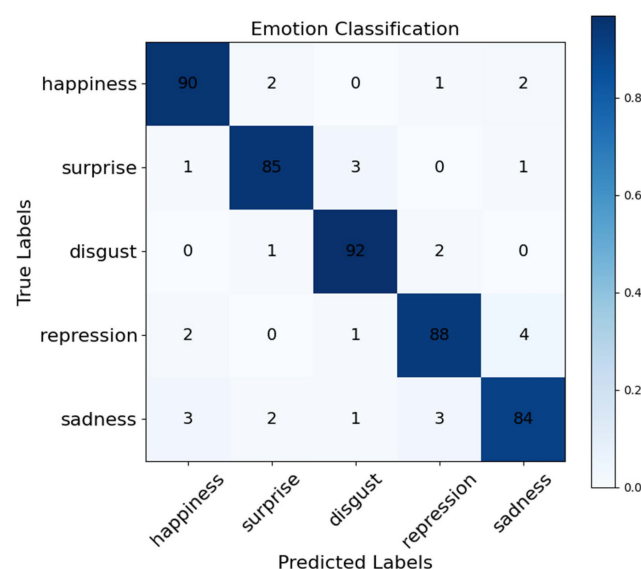
Figure 3. Confusion matrix for CASME dataset.
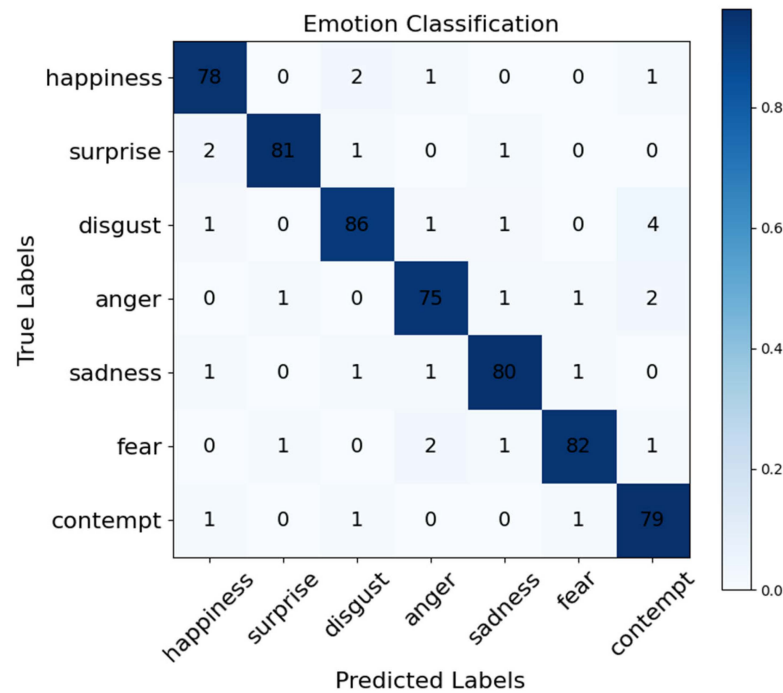
Figure 4. Confusion matrix for SAMM dataset.

The model achieved an accuracy of 95.62% on the CASME II dataset and 93.21% on the SAMM dataset. The CASME II dataset achieved a precision of 93.56, recall of 95.62, and F1-score of 91.78, and for the SAMM dataset these values are 91.45, 93.21, and 89.31. The attained precision signifies the model's ability to categorize well-known emotions according to the identified AUs. Due to its high accuracy value, the model is thought to identify emotions effectively and prevent misclassifications with a minimal number of false positives. The recall number indicates how well the model can represent most of the real emotions seen in the test data. Ultimately, the F1-score supports the model's balanced performance by taking both recall and accuracy into account.

4.1.2. Micro-Expression Sub-Division Based on Action Units

To make clusters (sub-divisions) of the different micro-expressions, such as happiness, sadness, anger, etc., the K-means clustering algorithm is employed. The K-means clustering technique is widely used to categorize unlabeled data elements into discrete groups. The data elements are progressively assigned to the cluster corresponding to the closest mean as a centroid in an iterative manner. The objective of this method is to form clusters that are both well-separated from one another as well as internally homogenous. The effectiveness of data points' clustering inside their designated groupings in comparison to points in other clusters is evaluated by silhouette score. As mentioned above in Section 3.4, scores nearest to 1 indicate well-formed clusters, 0 for overlapping, and -1 for weak cluster formation. The results of this experiment have obtained significant information on the ideal number of clusters and the overall efficacy of the K-means clustering for the SAMM dataset by examining the silhouette score for various K combinations for distinct emotions, as shown in Figure 5:
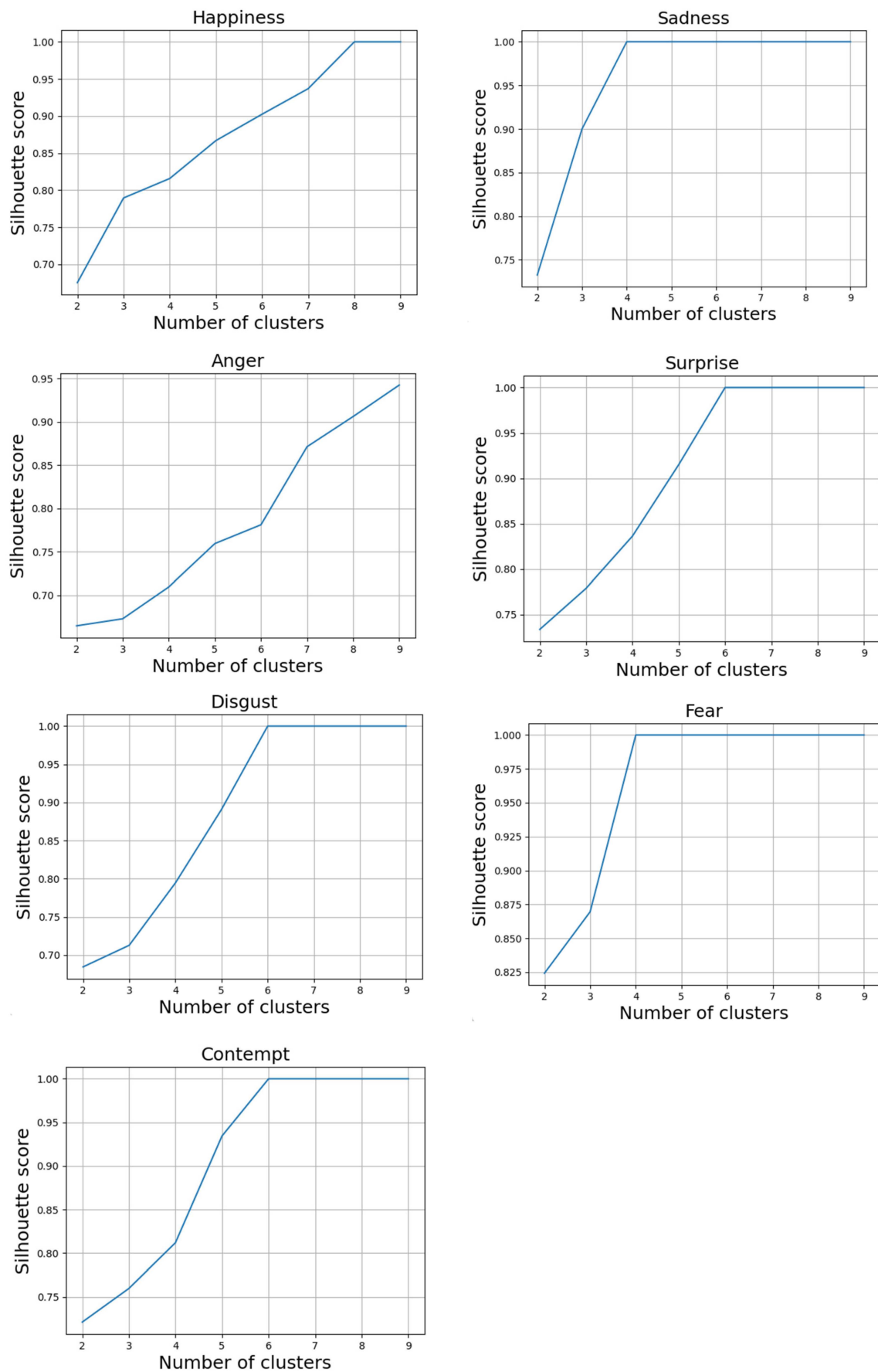
Figure 5. Silhouette score for combination of AUs for distinct micro-expressions.

4.1.3. Ablation Study

This ablation study aims to analyze the impact of the proposed framework to measure its performance among other machine-learning algorithms. The comparison between the proposed model and other algorithms, such as SVM, DT, ANN, and conventional CNN, is presented in Table 4. The decision tree has achieved a very low accuracy of 18.54% in micro-expression recognition. A decision tree works on clear decision boundaries and micro-expressions change with complex patterns. This might be the reason that the decision tree failed to capture these subtle changes with complex patterns. The results show that the proposed model has achieved remarkable accuracy among others.

Table 4. Evaluation of the proposed model among other ML classifiers.

Metrix	CNN	ANN	SVM	Decision Tree	Proposed Model
Accuracy-CASME II	87.14	81.32	75.96	18.54	**95.62**
Accuracy-SAMM	85.29	79.12	69.89	14.67	**93.21**

To assess the performance of the proposed model through varying the numbers of convolutional layers, epochs, batch size, optimizer, batch normalization, etc., these are presented in Tables 5–8.

Table 5. Evaluation of the proposed model depending on a variable number of parameters.

Number of Layers	Optimizer	Accuracy
1	sgd	0.7056
1	adam	0.7104
3	sgd	0.7230
3	adam	0.8268
5	sgd	0.8815
5	adam	**0.9562**

Table 6. Evaluation of the proposed model depending on epochs and batch size.

Epochs	Batch Size	Accuracy
32	16	0.7230
32	32	0.8041
45	16	0.8376
45	32	**0.9562**

Table 7. Evaluation of the proposed model with and without batch normalization layers.

Batch Normalization	Accuracy
No	0.738
Yes	**0.956**

Table 8. Evaluation of the proposed model depending on the activation function.

Activation Function	Accuracy
ReLu	**0.956**
sigmoid	0.912
tanh	0.947
elu	0.913

The results shown in Table 5 indicate that both the number of layers and the optimizer play a significant role in neural networks. The increment in the number of layers led to improved accuracy irrespective of the optimizer used. The Adam optimizer consistently achieved higher accuracy compared to SGD. In our case, the best combination was found with a 5-layer structure with the Adam optimizer. Table 6 suggests that the model benefits from both more training time to learn the intricate patterns and more data points per update to improve gradient estimation. If we increase the epochs or batch size, it can lead to overfitting. In that case, the model can learn from training data but performs poorly on unseen examples. Table 7 shows that batch normalization plays an important function in stabilizing the training process, which makes the model learn effectively. The selection of the activation function has a significant influence on the correctness of the proposed model. The activation functions explored during the study in Table 8 indicate that the elu and sigmoid functions result in low performance. The activation function tanh performed better. Due to the high computational efficiency, as well as to prevent the vanishing gradients, the relu function outperformed the other functions. It enables the model to learn effectively from variations throughout the whole network.

4.2. Discussion

The earlier sections examined the possibilities of computational learning and action units in identifying mental health conditions. This section explores the significance and ramifications of the results through a comparison between state-of-the-art techniques. Further, micro-expression sub-division is discussed based on different combinations of action units related to specific emotions. Subsequently, a case study is performed to evaluate the performance of the model by analyzing interview videos of depressed patients. Furthermore, this section explores the promising possibilities of ME utilization in applied, experimental, clinical, and assistive scenarios.

4.2.1. Comparison with State-of-the-Art Techniques

Micro-expressions—brief facial movements that convey real emotions—have enormous possibilities for use in automated emotion evaluation. However, it is still difficult to discern emotions through such nuanced visuals. This section compares the outcomes of the proposed approach to existing state-of-the-art methods in terms of datasets, methods, and accuracy percentage in Table 9.

Table 9. Comparison of state-of-the-art techniques with the proposed technique.

Year/Ref	Dataset	Method	Accuracy
2023 [40]	CASME, CAS(ME)2	Regional feature module (Reg), 3DCNN	F1-score (0.786)
2024 [41]	SMIC[1] CASME-II SAMM	3DCNN AU graph convolutional networks (GCN)	81.85% F1-score (0.7760)
2021 [42]	CASME-II SAMM	Depth-wise conv AU GCN	80.80%
2021 [43]	CK+[2]	Transfer learning ResNet50, VGG16 Inception V3, Mobile Net	96%
2020 [34]	SMIC[1] CASME-II	CNN, LSTM (LFM, CLFM)	71.34% 73.98%
2020 [23]	CASME-I CASME-II CAS(ME)2 SMIC[1]	CNN	76.57%
Proposed Method	CASME-II SAMM	CNN K-means	95.62% 93.56%

SMIC[1]: Spontaneous Micro-expression Database, CK+[2]: Extended Cohn–Kanade.

4.2.2. Micro-Expression Sub-Division

Table 10 lists the grouping of action units according to emotions. This demonstrates that an emotion and the action units that go along with it can have one or many combinations. Some combinations of AUs with 12A or 12B AUs represent the eye blinking and left or right eye movements. The AUs with L or R represent the activation on the left side or right side of the face.

Table 10. Micro-expressions of different emotions.

| Emotion	Action Units	MEs Sub-Division
Happiness	{AU7, AU12}	Happiness_ME1
	{AU12}	Happiness_ME2
	{AU12A}	Happiness_ME3
	{AU6, AU12, AU15}	Happiness_ME4
	{AU6, AU12}	Happiness_ME5
	{AU12A, AU24}	Happiness_ME6
	{AU12B}	Happiness_ME7
	{AUL12A, AU25}	Happiness_ME8
Surprise	{AU5A}	Surprise_ME1
	{AU5B, AU24}	Surprise_ME2
	{AU25, AU26}	Surprise_ME3
	{AU1, AU2}	Surprise_ME4
	{AU5}	Surprise_ME5
	{AU1A, AU2B, AU14C}	Surprise_ME6
Anger	{AU4}	Anger_ME1
	{AU4, AU7}	Anger_ME2
	{AU4, AU7, AU43}	Anger_ME3
	{AU7B, AU43E}	Anger_ME4
	{AU4, AU6, AU7, AU43}	Anger_ME5
	{AU7B}	Anger_ME6
	{AU7C}	Anger_ME7
	{AU7}	Anger_ME8
	{AU7A}	Anger_ME9
Sadness	{AU15, AU17}	Sadness_ME1
	{AU1}	Sadness_ME2
	{AU17}	Sadness_ME3
	{AU12, AU15}	Sadness_ME4
Disgust	{AU10}	Disgust_ME1
	{AU4, AU9}	Disgust_ME2
	{AU9}	Disgust_ME3
	{AU9, AU12}	Disgust_ME4
	{AU9, AU10}	Disgust_ME5
	{AU10, AU25, AU26}	Disgust_ME6
Fear	{AU20}	Fear_ME1
	{AU7, AU20, AU26}	Fear_ME2
	{AUL20, AU21}	Fear_ME3
	{AU20C, AU25, AU26}	Fear_ME4
Contempt	{AUL14}	Contempt_ME1
	{AUR14}	Contempt_ME2
	{AUR14, AUR17}	Contempt_ME3
	{AUL12, AUL14}	Contempt_ME4
	{AU14, AU25, AU26}	Contempt_ME5
	{AUR12, AUR14}	Contempt_ME6

4.2.3. Mental Health Assessment

Depression, stress, and anxiety are common mental health conditions that are frequently characterized by an all-encompassing sense of melancholy, disinterest, or adjust-

ments to sleep and eating patterns. One of the major obstacles to treating these kinds of mental health issues is that they can appear subtly, especially in those who are good at hiding their emotions. Considering such conditions, facial micro-expressions present a promising way to identify depression, stress, anxiety, or other mental health issues because they are brief, spontaneous expressions that might disclose underlying emotions. These un-noticed, millisecond-long micro-expressions can reveal a lot about an individual's genuine emotional state and can be evaded consciously. However, these fast and nuanced emotions are difficult for the human eye to pick up on.

Case Study

The proposed model is implemented on a video dataset of clinical interactions with patients suffering from mental health issues, which is publicly available on YouTube. These videos are the psychiatric interviews uploaded by the University of Nottingham with 480×360 resolution. There was no predetermined structure for these clinical interviews, so participants conversed freely and answered standard questions on mental health without being prompted. Initially, the pre-processing of these videos was done for face detection. After that, the pre-processed data were passed to the trained model. Figure 6 shows the resultant confusion matrix of the experiment, which has achieved an accuracy of 92.99% for emotion recognition.
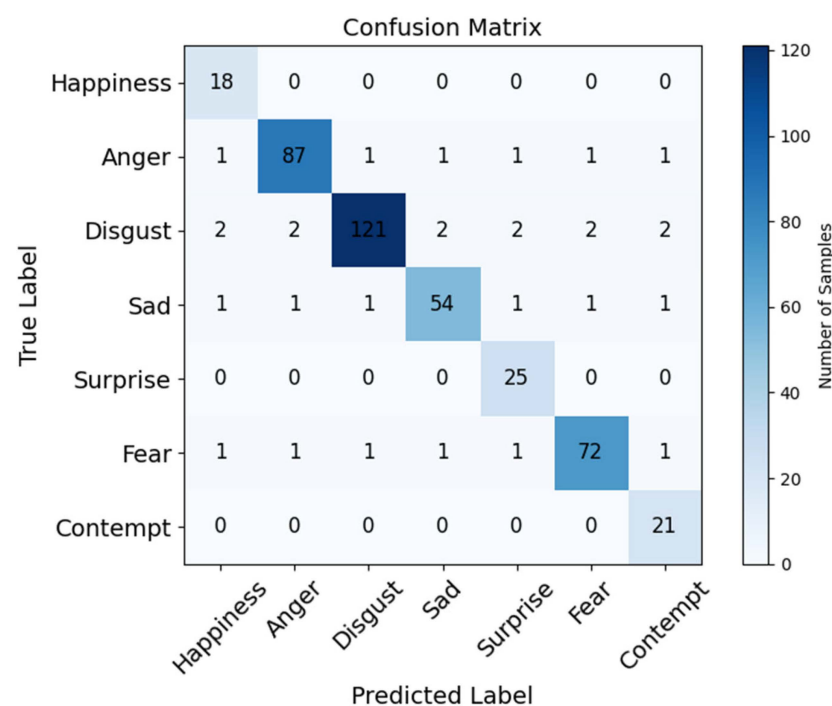
Figure 6. Confusion matrix for the case study.

This confusion matrix reveals that most depressed individuals experience disgust, anger, sadness, fear, and surprise emotions. The clustering model gives the predicted action units to determine under which emotion group they will be classified. The results show that most of these images have the following emotions: disgust, anger, and sad-ness with activated AU10 (upper lip raiser), AU4 (brow lowerer), AU1(inner brow raiser), AU17 (chin raiser), AU15 (lip corner depressed), and AU7 (lid tightener) with AU4, AU6 and AU9 combination. Therefore, the predicted cluster values fall under Disgust_ME1, Disgust_ME2, Sadness_ME1, Sadness_ME2, Anger_ME1, Anger_ME2, Anger_ME5, Sur-prise_ME4, and Disgust_ME8.

The proposed emotion recognition technique based on facial micro-expressions has the potential as a beneficial tool in numerous healthcare settings. It can facilitate the diagnoses

of mental health conditions, such as stress, anxiety, depression, autism spectrum disorder, etc., while interacting with the patients [44]. Also, it can be used to enhance the accuracy of pain analysis in individuals with impaired communication abilities. Emotion recognition remains an area for improvement in human–computer interaction (HCI) applications. Therefore, it can enhance the HCI in medical settings by enabling machines to have a deeper knowledge of user emotions [45]. Though these applications indicate the potential, additional investigations are required to validate the proposed model's accuracy in real-world settings. Moreover, ethical considerations are the major challenge related to the confidentiality and information security of the patients.

4.2.4. Application Scenarios

Emotion recognition through micro-expressions has the potential to provide substantial contributions to various real-world applications. This technology can be employed in assistive environments as well as in controlled clinical research. ME-based emotion recognition can be used in numerous areas, such as healthcare, neuromarketing, marketing, gaming, UI and UX Design, human–computer interaction (HCI), computer vision (CV) simulation, entertainment, and education. In healthcare monitoring systems, facial expression recognition is used for behavior analysis [46], pain detection [47], patient monitoring [48], and developing assistive technologies [49]. In HCI, assistive techniques for emotion recognition can provide a better understanding of users' emotional reactions to content moderation and sentiment analysis [45,50].

4.2.5. Experimental Scenarios

The experimental scenarios of this work could be in mental health assessment, healthcare monitoring, driver drowsiness detection, flight simulation, user interface designing, education, human resource management, and game design. The effectiveness of the method can be measured by using different stimuli, such as video clips, images, and audio. The subtle facial muscle movements vary, correlated to physiological changes, giving a better understanding of emotions comprehensively. Also, the experiment could be employed for developing a cross-cultural facial expression recognition model [51–53].

Emotions are strongly influenced by cultural differences and indirectly affect the recognition and production of MEs. Cultural influence on micro-expressions can be divided into people from Eastern cultures, with low-arousal, and people from Western cultures, with high-arousal emotional expressions. The neural indication, such as functional magnetic resonance imaging (fMRI), has shown that Chinese people value a state of calmness and European/American people value excitement more than calmness. people from Eastern cultures express more around the eye regions and people from Western cultures express their emotions through the full-face [54–56]. The considerate intensity of an identified emotion is the most important determinant for ME-based emotion recognition. Previous research has shown that MEs appear with very low intensity and macro facial expressions appear with high intensity. Pronounced intensities of macro-expressions can be captured through facial texture deformations and facial movements due to their presence on a large facial region [29].

4.2.6. Clinical Scenarios

As mentioned in Section 4.2.3, this technique can be used for mental health assessment and pain management. Early identification of different mental illness signs, such as depression, anxiety, or stress might be helpful to the clinician as well as patients for proper diagnosis and treatment. It can also be used to monitor the emotional states and improvement during rehabilitation, which enables the therapists to manage the treatment approach. Patients having impaired communication skills or difficulty in verbal communication during pain management can be monitored and helped with critical situations [53,57–59]. Facial MEs have the potential for utilization in mental health analysis, despite the existence

of substantial challenges. Some of the prominent areas for mental health detection are discussed below:

(a)   Depression detection:

Non-verbal clues, such as facial expressions, might serve as symptoms of depressive conditions in mental health analysis. As in clinical settings, depression diagnosis can be time-consuming and highly dependent upon expert observation. Automated detection of depressive disorders must be taken into consideration for early and effective treatment. It can be used as a screening tool for depression detection and treatment progress monitoring [60–63].

(b)   Anxiety and stress detection:

ME-based stress or anxiety detection methods have great potential in mental health interventions. Their goal is to identify the subtle symptoms of repression, discomfort, or worry, which could reveal the hidden stress. This technique has potential for applications in healthcare monitoring, workplace stress assessment, and in psychology, due to its effective and non-invasive methodology. Anxiety levels can also be measured through monitoring the subtle changes in facial muscles, like twitching under the eye and lip corner areas [7,8,64]

(c)   Autism spectrum disorder (ASD):

ASD is a neurodevelopmental disorder that affects the social engagement, communication skills, and behavior of an individual. People with ASD cannot express their emotions and face difficulty in managing their anxiety or stress. ME analysis of the ASD-affected persons could offer a direction to understand their emotional states and behavior that can be used to diagnose and their communication training [65,66].

(d)   Mood disorder:

ME-based applications can be explored to monitor the individual's overall mood variations with mental health conditions like bipolar disorder. This could be helpful for tracking the treatment progress throughout the time period where the continuous monitoring of mood variations is required [67].

(e)   Post-traumatic stress disorder (PTSD):

People with mental health conditions such as PTSD are prone to experiencing emotional trauma and flashbacks related to a horrifying event they experienced before. Micro-expression-based emotion analysis could potentially be used to identify the PTSD triggers and help to diagnose and treatment [14,68].

4.2.7. Assistive Scenarios

To enhance communication for individuals who have difficulty speaking, augmentative and alternative communication (AAC) technology might be upgraded by incorporating this technique [69]. Also, in educational scenarios, it could help the educator to change their teaching techniques by reviewing the student's emotional involvement during class activities from real-time captured emotional information [70]. The proposed technique can be explored in a virtual reality (VR)-based therapy system for mental health assessment by adapting the individual's emotional state [71–73]. This could also be beneficial for robots utilized in marketing, companionship, or in healthcare [49].

5. Conclusions

This study intended to determine the feasibility of assigning micro-expression labels to indicate possible emotional variances seen in facial expressions using facial action units. The proposed method incorporates the CASME II and SAMM datasets, which provide information regarding the emotion, action units related to the facial muscle movements, objective classes, and significant parameters. The model trained on these datasets achieved 95.62% and 93.21% accuracy and was trialed for the assessment of depressive disorder

from annotated video data. The findings show that majority of the depressed individuals exhibit disgust, sadness, and anger emotions. This has great potential in the discipline of mental health, as early intervention and therapy can greatly benefit from an accurate assessment of emotional states. In the future, other significant feature extraction techniques will be used to attain higher accuracy for mental health assessment and to assign the labels for this clinical interview data as ground truth. The suggested framework provides an intriguing approach for emotion recognition using micro-expressions, through action units with CNN and K-means clustering algorithms for ME label assignment. Future avenues for investigation also include finding the intensity level of low, medium, or high for micro-expression sub-division.

Author Contributions: Conceptualization, D.S. and J.S.; Formal analysis, S.S.S. and S.K.S.; Methodology, D.S., J.S. and S.S.S.; Project administration, D.S. and J.S.; Writing – original draft, D.S. and J.S.; Writing—review & editing, S.S.S. and S.K.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The dataset used in this study is fully available as an open-source resource.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Galderisi, S.; Heinz, A.; Kastrup, M.; Beezhold, J.; Sartorius, N. Toward a new definition of mental health. *World Psychiatry* **2015**, *14*, 231–233. [CrossRef] [PubMed]
2. Rayan, A.; Alanazi, S. A novel approach to forecasting the mental well-being using machine learning. *Alex. Eng. J.* **2023**, *84*, 175–183. [CrossRef]
3. Pise, A.A.; Alqahtani, M.A.; Verma, P. K, P.; Karras, D.A.; Halifa, A. Methods for Facial Expression Recognition with Applications in Challenging Situations. *Comput. Intell. Neurosci.* **2022**, *2022*, 9261438. [CrossRef] [PubMed]
4. Feng, Y.; Zhou, Y.; Zeng, S.; Pan, B. Facial expression recognition based on convolutional neural network. In Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 18–20 October 2019; pp. 410–413. [CrossRef]
5. Shen, X.B.; Wu, Q.; Fu, X.L. Effects of the duration of expressions on the recognition of microexpressions. *J. Zhejiang Univ. Sci. B* **2012**, *13*, 221–230. [CrossRef] [PubMed]
6. Farnsworth, B. Facial Action Coding System (FACS)—A Visual Guidebook—iMotions. *Web* **2019**, 1–26. Available online: https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/ (accessed on 2 July 2024).
7. Gavrilescu, M.; Vizireanu, N. Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors* **2019**, *19*, 3693. [CrossRef]
8. Giannakakis, G.; Koujan, M.R.; Roussos, A. Automatic stress detection evaluating models of facial action units. In Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020; pp. 728–733.
9. Martinez, B.; Valstar, M.F.; Jiang, B.; Pantic, M. Automatic analysis of facial actions: A survey. *IEEE Trans. Affect. Comput.* **2019**, *10*, 325–347. [CrossRef]
10. Friesen, W.V. Nonverbal Leakage and Clues to Deception. *Psychiatry* **1969**, *32*, 88–106. [CrossRef]
11. O'Sullivan, M.; Frank, M.G.; Hurley, C.M.; Tiwana, J. Police Lie Detection Accuracy: The Effect of Lie Scenario. *Law Hum. Behav.* **2009**, *33*, 530–538. [CrossRef]
12. Hussein, S.A.; Bayoumi, A.E.R.S.; Soliman, A.M. Automated detection of human mental disorder. *J. Electr. Syst. Inf. Technol.* **2023**, *10*, 9. [CrossRef]
13. La Monica, L.; Cenerini, C.; Vollero, L.; Pennazza, G.; Santonico, M.; Keller, F. Development of a Universal Validation Protocol and an Open-Source Database for Multi-Contextual Facial Expression Recognition. *Sensors* **2023**, *23*, 8376. [CrossRef]
14. Mahayossanunt, Y.; Nupairoj, N.; Hemrungrojn, S.; Vateekul, P. Explainable Depression Detection Based on Facial Expression Using LSTM on Attentional Intermediate Feature Fusion with Label Smoothing. *Sensors* **2023**, *23*, 9402. [CrossRef] [PubMed]
15. Zhao, G.; Pietik, M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [CrossRef] [PubMed]
16. Davison, A.K.; Merghani, W.; Yap, M.H. Objective classes for micro-facial expression recognition. *J. Imaging* **2018**, *4*, 119. [CrossRef]
17. Liu, Y.J.; Zhang, J.K.; Yan, W.J.; Wang, S.J.; Zhao, G.; Fu, X. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition. *IEEE Trans. Affect. Comput.* **2016**, *7*, 299–310. [CrossRef]

18. Liong, S.T.; See, J.; Wong, K.S.; Phan, R.C.W. Less is more: Micro-expression recognition from video using apex frame. *Signal Process. Image Commun.* **2018**, *62*, 82–92. [CrossRef]

19. Duque, C.A.; Alata, O.; Emonet, R.; Konik, H.; Legrand, A.C. Mean oriented Riesz features for micro expression classification. *Pattern Recognit. Lett.* **2020**, *135*, 382–389. [CrossRef]

20. Polikovsky, S.; Kameda, Y.; Ohta, Y. Facial micro-expressions recognition using high speed camera and 3D-Gradient descriptor. In Proceedings of the IET Seminar Digest, London, UK, 3 December 2009; Volume 2009.

21. Patel, D.; Hong, X.; Zhao, G. Selective deep features for micro-expression recognition. In Proceedings of the International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 2258–2263.

22. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, CVPRW 2010, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.

23. Verma, M.; Vipparthi, S.K.; Singh, G.; Murala, S. LEARNet: Dynamic Imaging Network for Micro Expression Recognition. *IEEE Trans. Image Process.* **2020**, *29*, 1618–1627. [CrossRef] [PubMed]

24. Kim, D.H.; Baddar, W.J.; Ro, Y.M. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In *MM 2016—Proceedings of the 2016 ACM Multimedia Conference, Amsterdam, The Netherlands, 15–19 October 2016*; Association for Computing Machinery, Inc.: New York, NY, USA, 2016; pp. 382–386.

25. Khor, H.Q.; See, J.; Phan, R.C.W.; Lin, W. Enriched long-term recurrent convolutional network for facial micro-expression recognition. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 667–674. [CrossRef]

26. Lewinski, P.; Den Uyl, T.M.; Butler, C. Automated facial coding: Validation of basic emotions and FACS AUs in facereader. *J. Neurosci. Psychol. Econ.* **2014**, *7*, 227–236. [CrossRef]

27. Ekman, P.; Friesen, W.V. Facial Action Coding System. In *Facial Action Coding System (FACS)*; APA PsycTests: Washington, DC, USA, 1978.

28. Zhao, Y.; Xu, J. Necessary morphological patches extraction for automatic micro-expression recognition. *Appl. Sci.* **2018**, *8*, 1811. [CrossRef]

29. Zhao, Y.; Xu, J. An improved micro-expression recognition method based on necessary morphological patches. *Symmetry* **2019**, *11*, 497. [CrossRef]

30. Davison, A.K.; Lansley, C.; Costen, N.; Tan, K.; Yap, M.H. SAMM: A Spontaneous Micro-Facial Movement Dataset. *IEEE Trans. Affect. Comput.* **2018**, *9*, 116–129. [CrossRef]

31. Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* **2014**, *9*, e86041. [CrossRef]

32. Ruan, B.K.; Lo, L.; Shuai, H.H.; Cheng, W.H. Mimicking the Annotation Process for Recognizing the Micro Expressions. In *MM 22: The 30th ACM International Conference on Multimedia, Lisbon, Portugal, 10–14 October 2022*; Association for Computing Machinery: New York, NY, USA, 2022; Volume 1, ISBN 9781450392037.

33. Allaert, B.; Bilasco, I.M.; Djeraba, C. Micro and Macro Facial Expression Recognition Using Advanced Local Motion Patterns. *IEEE Trans. Affect. Comput.* **2022**, *13*, 147–158. [CrossRef]

34. Choi, D.Y.; Song, B.C. Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps. *IEEE Access* **2020**, *8*, 121549–121563. [CrossRef]

35. Li, C.; Qi, Z.; Jia, N.; Wu, J. Human face detection algorithm via Haar cascade classifier combined with three additional classifiers. In Proceedings of the ICEMI 2017—IEEE 13th International Conference on Electronic Measurement and Instruments, Yangzhou, China, 20–22 October 2017; pp. 483–487.

36. Benitez-Garcia, G.; Olivares-Mercado, J.; Aguilar-Torres, G.; Sanchez-Perez, G.; Perez-Meana, H. Face identification based on Contrast Limited Adaptive Histogram Equalization (CLAHE). In Proceedings of the 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2011, Washington, DC, USA, 20–25 June 2011; Volume 1, pp. 363–369.

37. Irawan, B.; Utama, N.P.; Munir, R.; Purwarianti, A. Spontaneous Micro-Expression Recognition Using 3DCNN on Long Videos for Emotion Analysis. In Proceedings of the 2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA), Lombok, Indonesia, 7–9 October 2023; pp. 1–6. [CrossRef]

38. Chowanda, A. Separable convolutional neural networks for facial expressions recognition. *J. Big Data* **2021**, *8*, 132. [CrossRef]

39. Hashmi, M.F.; Kiran Kumar Ashish, B.; Sharma, V.; Keskar, A.G.; Bokde, N.D.; Yoon, J.H.; Geem, Z.W. Larnet: Real-time detection of facial micro expression using lossless attention residual network. *Sensors* **2021**, *21*, 1098. [CrossRef] [PubMed]

40. Zhou, G.; Yuan, S.; Xing, H.; Jiang, Y.; Geng, P.; Cao, Y.; Ben, X. Micro-expression action unit recognition based on dynamic image and spatial pyramid. *J. Supercomput.* **2023**, *79*, 19879–19902. [CrossRef]

41. Wang, L.; Cai, W. Micro-expression recognition by fusing action unit detection and spatio-temporal features. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 5595–5599. [CrossRef]

42. Lei, L.; Chen, T.; Li, S.; Li, J. Micro-expression recognition based on facial graph representation learning and facial action unit fusion. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 1571–1580. [CrossRef]

43. Chowdary, M.K.; Nguyen, T.N.; Hemanth, D.J. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Comput. Appl.* **2023**, *35*, 23311–23328. [CrossRef]

44. Banskota, N.; Alsadoon, A.; Prasad, P.W.C.; Dawoud, A.; Rashid, T.A.; Alsadoon, O.H. A novel enhanced convolution neural network with extreme learning machine: Facial emotional recognition in psychology practices. *Multimed. Tools Appl.* **2022**, *82*, 6479–6503. [CrossRef]

45. Wang, H.H.; Gu, J.W. The applications of facial expression recognition in human-computer interaction. In Proceedings of the 2018 IEEE International Conference on Advanced Manufacturing (ICAM), Yunlin, Taiwan, 16–18 November 2018; pp. 288–291. [CrossRef]

46. Healy, M.; Walsh, P. Detecting demeanor for healthcare with machine learning. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 2015–2019. [CrossRef]

47. Bargshady, G.; Zhou, X.; Deo, R.C.; Soar, J.; Whittaker, F.; Wang, H. Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Syst. Appl.* **2020**, *149*, 113305. [CrossRef]

48. Onyema, E.M.; Shukla, P.K.; Dalal, S.; Mathur, M.N.; Zakariah, M.; Tiwari, B. Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet. *J. Healthc. Eng.* **2021**, *2021*, 5196000. [CrossRef] [PubMed]

49. Bisogni, C.; Castiglione, A.; Hossain, S.; Narducci, F.; Umer, S. Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5619–5627. [CrossRef]

50. Wei, H.; Zhang, Z. A survey of facial expression recognition based on deep learning. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9–13 November 2020; pp. 90–94. [CrossRef]

51. Saffaryazdi, N.; Wasim, S.T.; Dileep, K.; Nia, A.F.; Nanayakkara, S.; Broadbent, E.; Billinghurst, M. Using Facial Micro-Expressions in Combination With EEG and Physiological Signals for Emotion Recognition. *Front. Psychol.* **2022**, *13*, 864047. [CrossRef]

52. Huang, Y.; Yang, J.; Liu, S.; Pan, J. Combining facial expressions and electroencephalography to enhance emotion recognition. *Futur. Internet* **2019**, *11*, 105. [CrossRef]

53. Lu, S.; Li, J.; Wang, Y.; Dong, Z.; Wang, S.J.; Fu, X. A More Objective Quantification of Micro-Expression Intensity through Facial Electromyography. In Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis, Lisbon, Portugal, 14 October 2022; pp. 11–17. [CrossRef]

54. Park, B.K.; Tsai, J.L.; Chim, L.; Blevins, E.; Knutson, B. Neural evidence for cultural differences in the valuation of positive facial expressions. *Soc. Cogn. Affect. Neurosci.* **2015**, *11*, 243–252. [CrossRef] [PubMed]

55. Lim, N. Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integr. Med. Res.* **2016**, *5*, 105–109. [CrossRef] [PubMed]

56. Yang, Y. Micro-expressions: A Study of Basic Reading and The Influencing Factors on Production and Recognition. *J. Educ. Humanit. Soc. Sci.* **2024**, *26*, 1048–1053. [CrossRef]

57. Carneiro de Melo, W.; Granger, E.; Hadid, A. A Deep Multiscale Spatiotemporal Network for Assessing Depression from Facial Dynamics. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1581–1592. [CrossRef]

58. Fei, Z.; Yang, E.; Li, D.D.U.; Butler, S.; Ijomah, W.; Li, X.; Zhou, H. Deep convolution network based emotion analysis towards mental health care. *Neurocomputing* **2020**, *388*, 212–227. [CrossRef]

59. De Sario, G.D.; Haider, C.R.; Maita, K.C.; Torres-Guzman, R.A.; Emam, O.S.; Avila, F.R.; Garcia, J.P.; Borna, S.; McLeod, C.J.; Bruce, C.J.; et al. Using AI to Detect Pain through Facial Expressions: A Review. *Bioengineering* **2023**, *10*, 548. [CrossRef]

60. Gorbova, J.; Colovic, M.; Marjanovic, M.; Njegus, A.; Anbarjafari, G. Going deeper in hidden sadness recognition using spontaneous micro expressions database. *Multimed. Tools Appl.* **2019**, *78*, 23161–23178. [CrossRef]

61. Chahar, R.; Dubey, A.K.; Narang, S.K. A review and meta-analysis of machine intelligence approaches for mental health issues and depression detection. *Int. J. Adv. Technol. Eng. Explor.* **2021**, *8*, 1279–1314. [CrossRef]

62. Huang, W. Elderly depression recognition based on facial micro-expression extraction. *Trait. Signal* **2021**, *38*, 1123–1130. [CrossRef]

63. He, L.; Jiang, D.; Sahli, H. Automatic Depression Analysis Using Dynamic Facial Appearance Descriptor and Dirichlet Process Fisher Encoding. *IEEE Trans. Multimed.* **2019**, *21*, 1476–1486. [CrossRef]

64. Zhang, J.; Yin, H.; Zhang, J.; Yang, G.; Qin, J.; He, L. Real-time mental stress detection using multimodality expressions with a deep learning framework. *Front. Neurosci.* **2022**, *16*, 947168. [CrossRef] [PubMed]

65. Munsif, M.; Ullah, M.; Ahmad, B.; Sajjad, M.; Cheikh, F.A. *Monitoring Neurological Disorder Patients via Deep Learning Based Facial Expressions Analysis*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; Volume 652, ISBN 9783031083402.

66. Beibin, L.; Mehta, S.; Aneja, D.; Foster, C.; Ventola, P.; Shic, F. A facial affect analysis system for autism spectrum disorder. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 4549–4553.

67. Gilanie, G.; Ul Hassan, M.; Asghar, M.; Qamar, A.M.; Ullah, H.; Khan, R.U.; Aslam, N.; Khan, I.U. An Automated and Real-time Approach of Depression Detection from Facial Micro-expressions. *Comput. Mater. Contin.* **2022**, *73*, 2513–2528. [CrossRef]

68. Wu, Y.; Mao, K.; Dennett, L.; Zhang, Y.; Chen, J. Systematic review of machine learning in PTSD studies for automated diagnosis evaluation. *npj Ment. Health Res.* **2023**, *2*, 16. [CrossRef]

69. Elsahar, Y.; Hu, S.; Bouazza-Marouf, K.; Kerr, D.; Mansor, A. Augmentative and Alternative Communication (AAC) Advances: A Review of Configurations for Individuals with a Speech Disability. *Sensors* **2019**, *19*, 1911. [CrossRef]

70. Tonguç, G.; Ozaydın Ozkara, B. Automatic recognition of student emotions from facial expressions during a lecture. *Comput. Educ.* **2020**, *148*, 103797. [CrossRef]

71. Papoutsi, C.; Drigas, A.; Skianis, C. Virtual and Augmented Reality for Developing Emotional Intelligence Skills. *Int. J. Recent Contrib. Eng. Sci. IT* **2021**, *9*, 35. [CrossRef]
72. Sharma, B.; Mantri, A. Augmented reality underpinned instructional design (ARUIDS) for cogno-orchestrative load. *J. Comput. Theor. Nanosci.* **2019**, *16*, 4379–4388. [CrossRef]
73. Dawn, S. Virtual reality and augmented reality based affective computing applications in healthcare, challenges, and its future direction. In *Affective Computing in Healthcare: Applications Based on Biosignals and Artificial Intelligence*; IOP Publishing Ltd.: Bristol, UK, 2023. [CrossRef]