Erudite Journal of Engineering Technology and Management Sciences

Vol. 6, No. 2, June 2026, pp. 47-55

Journal home page: http://www.ejetms.com

EJETMS

ISSN:2583-2131

Machine Learning-Based Analysis of Facial Expressions and Emotion Regulation through Micro-Expressions

Kalpana. V[1] and K. Arun Kumar[2]

[1] PG Scholar, Dept. of CSE Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India

[2] Professor, Dept. of CSE Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India

Email: kalpanasietk@gmail.com k.arunjoy@gmail.com

Received: 15 January 2026
Accepted: 08 April 2026

Keywords:
Facial micro-expressions, Emotion recognition, Deep learning, HybridMicroNet, CASME-II

ABSTRACT

Emotion Recognition through Facial Expression is an increasingly critical component of Human Computer Interface Technology (HCT), Mental Health Assessments and Remote Patient Monitoring. Traditional macro-expression based emotion recognition systems use facial expressions to assess a persons emotional status; yet, they do not capture the very subtle and involuntary movements of a persons face while experiencing emotion referred to as micro-expressions. Micro-expressions are generally short lived and may offer valuable insight to a person's actual emotional condition. This research proposes HybridMicroNet, a hybrid deep learning model to perform micro-expression detection that incorporates Resnet and Vgg-16 into one model in order to provide a better solution for the classification of images; it has been tested using data samples from many sources such as SAMM and CASME-II that contain micro-expression data. HybridMicroNet has great potential for deployment in numerous real world applications, such as Social Media Video Content and Clinical Interview Recordings, where there are often subtle or unlabeled emotional cues. For example, HybridMicroNet achieved 99.08% accuracy for the CASME-II dataset and 97.62% accuracy for the SAMM dataset. This demonstrates the ability of Machine Learning to recognize Emotions and Detect Mental Illness and highlights the growing interest in developing emotionally responsive systems.

1. INTRODUCTION

AI and machine learning based applications employed to analyze the massive amount of data collected through different resources. These applications can be utilized to evaluate the risk prediction, pattern formation, requirement of customized treatment regimens, and immediate assistance. AI based systems have the potential to analyze both physiological and behavioral data for early identification and relapse of mental health conditions. The technology can be employed to help through chatbots and customize treatment plans generation according to individual preferences and requirements. Human communication involves either verbal and non-verbal indicators or combinations of these that increase the complexity to understand these indicators. The verbal (spoken words, text, speech, audio) and non-verbal (facial expressions, body language & gestures, eye movements) signals allow us to understand the different emotional states, interpersonal relationships, deep conversations, and navigation of social interactions among human. Emotion can be a state of mind or experience that arises without deliberate effort and become reflected in behavioral reactions in human bodies. Human emotion can be recognized through physiological signals, neuroimaging techniques, gestures, speech, text, and facial expressions. According to theory of emotions, discrete and dimensional models are two approaches employed to recognize different emotions. Discrete model specifies the basic emotions in separate and distinct categories like sadness, happiness, disgust, anger, fear, and surprise. Facial expressions are categorized broadly into two distinct two groups: macro-expressions and micro-expressions (MEs) based on their time duration, voluntariness, and intensity. Macro-expressions are most prevalent and instantaneously noticeable facial expressions that usually last among 0.5 to 4 seconds and are clearly visible on the face. These types of expressions can be seen through naked eyes and consciously presented during social interactions to express their emotions. Different characteristics like longer duration, voluntariness or involuntariness, full-facial involvement, and easily perceiving nature of macro-expressions can be detected easily. This expression can be genuine or can be consciously manipulated to hide the actual intentions. Both macro and micro levels of facial expressions provide a lot of insight into the mental wellness of an individual by establishing a strong link between their emotional experience and mental wellness. Disturbances to processing emotions often surface as irregularities or disturbances in facial expressions. The complex neural pathways controlling these facial muscles and emotions are capable of reflecting conscious as well as unconscious and subconscious expression on an individual's face. Therefore, facial expressions have become an increasingly viable biomarker for detecting mental health problems since mental health refers to a person's psychological, emotional, and social wellness. A person's overall mental wellness defines their capacity to manage stress, make decisions, engage in positive behaviors, and interact with other people on a day-to-day basis. Various mental health issues such as stress, depression, anxiety, schizophrenia, and PTSD can manifest in altered facial expressivity, i.e., less frequent emotional displays, suppressed displays of emotion, and/or asymmetrical facial movement. Facial expressions can demonstrate the characteristic patterns of various mental health issues even when an individual intentionally conceals their feelings or is unaware of their actual emotional responses. Mental health professionals and researchers utilize facial expression analysis (specifically ME) to discover an individual's concealed emotions for early identification and development of treatment plans.

The emergence of machine learning has drastically revolutionized facial expression analysis to recognize emotions as compare to previous methods which were expensive, time-consuming, and dependent on human observation. Machine learning based frameworks offer the automatic detection and classification of facial expressions related with different emotions. Automated Facial Expression Recognition (FER) techniques can be utilized in healthcare to diagnose developmental abnormalities, constant monitoring of patients with unstable mental state and other mental illnesses like stress, anxiety, and depression because to their non-invasive nature. It provides the researchers new insights to uncover the subtle connection between emotions and mental health. In most cases, the procedures require a number of different steps, such as face localization, facial component alignment, feature extraction, as well as facial feature categorization. Machine learning based systems are trained on the large datasets to learn the different facial features and their interpretation corresponding to different emotional states. This process of automation can make the analysis more efficient and decrease the limitations that come from human judgment. Facial expression-based emotion recognition is achieved using both supervised and unsupervised machine learning algorithms. Supervised learning algorithms including k-nearest neighbor (knn), naïve bayes (nb), random forests (rf), and support vector machines (svm) utilize datasets that include previously labeled facial expressions with the corresponding emotion category. In supervised learning, algorithms categorize new data according to how they understand facial expression characteristics in relation to specific emotions during the training phase. Unsupervised learning algorithms including clustering, association rule mining, and dimensionality reduction are used to identify patterns within datasets that lack labels. These techniques enable the grouping of similar facial expressions and provide insight into latent emotional characteristics or categories that may not be explicitly defined. These techniques are useful for identifying new emotional expressions through exploratory analysis. Deep learning has provided excellent results in image and video processing. Techniques like long short-term memory (lstm), convolutional neural networks (cnn), and recurrent neural networks (rnn) can automatically derive detailed features from individual frames of video or raw pixel data thereby reducing the need for manually extracting features. Spatial and temporal features of various images can be derived using deep learning-based neural networks. These models have outperformed traditional methods of facial expression analysis and emotion recognition. Automated emotion recognition systems based on machine or deep learning are growing rapidly and will potentially be able to be applied to the assessment of patients' emotional states and mental health conditions through the use of objective measures. It is possible that this technology could be utilized as an extra resource for clinicians to gather additional useful data on the patient's progress through treatment and their present emotional state by being incorporated into either the clinical interview process, or into therapy sessions. This technology can also be paired with wearable devices and/or smart phone apps to monitor a patient's mental health at all times throughout their day. With the combination of machine learning and facial micro-expressions, the identification of mental health is very promising; and various studies have been completed to create an automated framework that detects the first signs of mental illness from minute changes in facial expression. However, while the technology has made considerable advancements, it is necessary to continue researching the many research gaps that exist to enable wider use of this technology. It is therefore imperative to continue researching to improve the reliability, accuracy, and applicability of micro-expression based mental health detection.

2. LITERATURE REVIEW

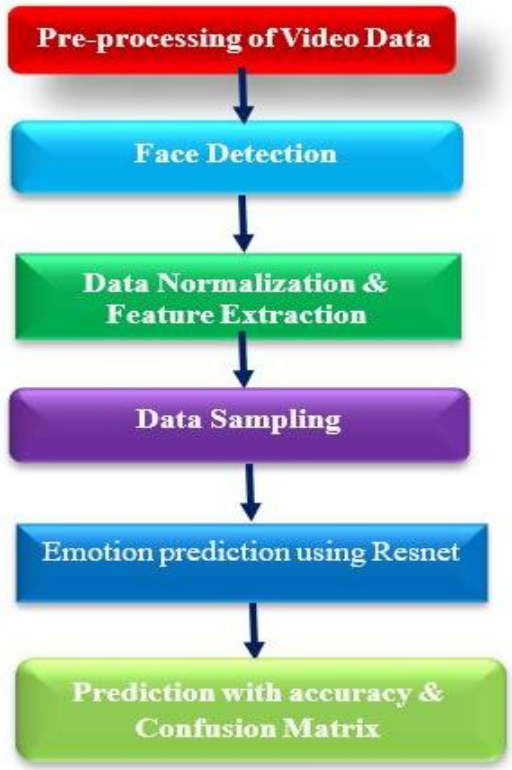
There has been a significant improvement in the application of machine learning to emotion recognition and regulation through facial expressions, because of technology advancements. Building off Paul Ekman's groundbreaking research in the 1970s where he identified universal facial expressions related to primary emotions (sadness, anger, and happiness) and further developed the concept of micro-expressions (short, unconscious facial expressions that indicate an individual's true emotions during attempts to hide them), allows researchers to now understand and identify the emotions of other people. The use of Artificial Intelligence (AI), as well as, Convolutional Neural Networks (CNNs) will allow researchers to better and with a higher degree of accuracy assess and understand the facial expressions of participants than they could before. This is due to the capabilities provided through CNNs allowing researchers to provide greater depth and more accuracy in assessing the details of facial expressions to detect and understand emotions. Additionally, researchers have been studying how machine learning can be used in emotion regulation; emotion regulation is the process of controlling your emotional response to stimuli (for example, suppressing feelings of sadness or "faking" a smile). As researchers begin to study the application of machine learning to emotion regulation, they are beginning to examine how people manage their emotional responses to stimuli using their facial expressions. For example, if someone suppressed anger by displaying a neutral expression, while another person may display a wider-than-average smile to demonstrate happiness. Emotion regulation is important in many social situations (such as a business meeting, public speaking or during therapy); therefore, researchers have determined that Recurrent Neural Networks (RNNs) are useful in examining the dynamic and changes in emotions, as well as, tracking the progression of facial expressions and predicting future emotional states (all of which are key elements in completely understanding how individuals regulate their emotions). Researchers have also started to add other physiological indicators (such as heart rate) to facial expression analyses in an effort to provide a more complete picture of an individual's emotional state. While creating successful emotion recognition systems across cultures is challenging, facial expressions are largely considered to be universal. Nevertheless, studies have shown that the way individuals express emotions is culturally influenced. People from different cultures tend to express similar emotions differently through facial expressions. To accurately identify emotional expressions, therefore, machine learning models need to be trained using large data sets of diverse expressions from various cultures. Additionally, human emotions are complex. Typically, individuals do not experience just one emotion at a time. An individual can concurrently feel happiness and sadness. Therefore, researchers must apply advanced methods to assess and evaluate emotional expressions accurately and comprehensively. FACS (Facial Action Coding System) is a method used to evaluate facial expressions. Finally, the long-term goal is for researchers to develop real-time emotion detection systems so they can be applied practically in fields like therapy and customer service. There are still many barriers to creating effective emotion recognition and regulation systems using machine learning, but the possibilities of being able to study facial expressions and emotion regulation through this type of system are vast. Potential applications of emotion recognition and regulation systems include improving therapist-patient interaction in healthcare settings, enhancing teacher-student relationships in education settings, and improving user experiences in customer service. As machine learning continues to grow and improve, it is anticipated that emotion recognition and regulation systems will play an increasing role in improving human computer interactions and increasing the emotional intelligence of computers. Overall, using machine learning with the study of facial expressions and emotion regulation has greatly increased the capability to detect and understand emotions. Current research continues to present opportunities to increase the effectiveness and applicability of emotion recognition and regulation systems in a wide range of areas.

3. METHODOLOGY

The research aims to clarify the relationship between facial micro-expression (as it relates to) emotional control, and mental health using an in-depth, multi-step methodology for investigation. This was done through use of emotion recognition software and facial expressions. To further understand the way computers can be used to detect emotions through facial expressions, researchers combined knowledge from psychology with computer based methodologies. The paper will discuss all aspects of the study's methodology including the framework development, selecting the dataset for the study, preprocessing the dataset for the study, identifying faces, extracting features from the faces, emotion recognition, and how to implement and evaluate the process used for the study.

3.1 Dataset Description

The two datasets used in this study were the SAMM and CASME-II datasets. These are established data sources for micro-expression research. The SAMM dataset is a relatively new and high-quality collection of 159 naturally-occurring micro-movements that represent a diverse sample of people across different demographics. The SAMM dataset also covers a wide variety of emotional expression types. Participants in the SAMM dataset consisted of 32 people; 16 male, 16 female with a mean age of 34.35 years old. The SAMM dataset included 7 primary emotional induction scenarios, and all videos were recorded at 220 frames per second (fps), and the video resolution for the face was approximately 480 x 480 pixels. The overall frame size was 2028 x 1080 pixels. The micro-expressions were categorized into 7 categories: happiness, fear, anger, contempt, disgust, surprise, and others. The CASME-II dataset is an updated version of the CASME dataset, and it contains dynamic and spontaneous micro-expressions in every recorded instance. The CASME-II dataset contained 248 samples, recorded at 220 fps from 38 subjects (males and females) in the facial region at a resolution of approximately 320 x 384 pixels, and the overall frame resolution was 720 x 540 pixels. The micro-expressions in the CASME-II dataset were categorized into 5 emotional types: repression, surprise, happiness, disgust, and others. Both the SAMM and CASME-II datasets have been used in many studies on emotion recognition and micro-expressions to analyze facial muscle movement, action units, skin deformation, elasticity, and facial landmark features.



3.2 Proposed Framework

The next phase will involve applying the Haar Cascade Classifier to identify the face within an image followed by resizing identified faces to 256 × 256 pixel images. Once resized, these images will be further improved via CLAHE (Contrast-Limited Adaptive Histogram Equalization), to enhance contrast while maintaining image quality. Following this, additional data will be created and features will be extracted through the VGG-16 model developed by the Visual Geometry Group. The VGG-16 utilizes pre-trained weights from ImageNet for transfer learning and is capable of detecting many different types of visual features that are necessary to detect small facial expressions, such as edges, shapes and texture. After the feature extraction process is complete, the dataset will be divided into 70% for training and 30% for testing. Finally, the last phase will be to develop a single, combined deep learning model by combining the VGG-16 model with a lightweight ResNet-based model, to provide emotion prediction and thus increase the ability of the system to accurately detect micro-expressions.ecognition and regulation systems in a wide range of areas.

3.2.1 Data Pre-processing

Converting the micro-expressions video dataset to grayscale is the first processing step. Frames (equal in number) for both the beginning and end of each video are obtained to create a consistent representation of facial features before and after the peak of emotion. Processing the video dataset this way also provides normalized data by creating an environment that has both spatial and temporal equivalence for the frame in which the micro-expression appears; samples of frames were taken at a rate of 10 per second and converted to grayscale images so the facial characteristics could be measured with higher precision and clarity.

3.2.2 Face Detection

The goal of facial detection is to reliably recognize and identify faces in grayscale images using the Haar Cascade Classifier; an advanced computer-based machine learning tool that has been trained to analyze an assortment of images that contain both facial and non-facial examples. The training for the classifier utilizes positive samples which represent facial images; and negative samples that provide examples of objects without faces, so that the model learns to identify images as either having or not having a face. This classification can then be used to determine if an object contains one or more facial features. In addition, the facial recognition system may utilize local histogram equalization, such as CLAHE (Contrast Limited Adaptive Histogram Equalization), to improve image contrast and remove some distortion to make it easier for the computer to track even slight changes in facial expressions and reduce the effects of poor lighting conditions.

3.2.3 Feature Extraction

The proposed framework employs the VGG-16, an effective feature extractor with origins at Oxford. The overall structure consists of 14 convolutional layers combined with four fully connected layers; thus, making it highly suitable for the tasks of image classification and pattern identification. In the proposed framework, the VGG-16 plays a critical role in extracting meaningful characteristics from facial images through the hierarchical representations that are learned. Utilizing the concept of transfer learning, the network effectively identifies several attributes of the facial images (shape, texture, edges, color, and semantics) even though there is little to no training data available. Furthermore, through the application of transfer learning, the training process of the network is significantly reduced and both the computational complexity and required processing power are minimized while maintaining high levels of classification accuracy. Due to the consistent architecture and pre-trained ImageNet weights, the VGG-16 has demonstrated that it can accurately capture minute variations in micro-expressions. As such, the model's three fully connected layers have been manually removed in order to optimize the feature extraction capabilities of the VGG-16. VGG-16 has shown its ability to find important properties within small amounts of data. To expand upon the variety of data in the set while maintaining all of the same classification labels, data augmentation (flipping, rotation, rescale) was performed on each image before applying feature extraction to the data set. By performing the data augmentation process prior to feature extraction the model could determine reliable features using additional training images. This type of augmentation is also especially useful when working with micro-expression analysis since data augmentation allows for the capturing of subtle differences and transitory nature of the facial expressions being analyzed.

3.2.4 Emotion recognition

The Resnet model was modified to include the extraction of features from the VGG-16 model for use in detecting emotions. The data was split prior to training with 75% used for training and 25% for testing to allow for accurate evaluation of the systems' performance. This division ensures that the data will be equally distributed between the data for training the system as well as the data for validating the system during machine learning. The HybridMicroNet model, which was proposed, has two variable parameters; the size of the input data and the number of classes being classified (output). Basic features of the input image are created from the first layer of convolution. The image data has been padded with zeros to ensure that the spatial dimensions of the input and output remain constant throughout the layers of processing. The zero padded data is processed using a 2-D convolutional layer along with batch normalization and ReLU activation to improve the stability of the features and introduce non-linearity. A 7x7 kernel, 16 filters, 2x2 stride, and a padding of 2 were applied to the 224x224 input image. The feature maps produced by this convolution are then processed using a max pooling layer with a 3x3 kernel and a 2x2 stride to reduce the spatial dimension of the feature map and preserve important information.

3.3 Implementation

The original images which were 224 × 224, were expanded by adding a three pixel border to all sides of the image. The images were also manipulated through methods such as resize, rotation, flip, etc., to create diversity within the data set. The model was modified to remove specific layers to aid in feature extraction. A residual attention mechanism was used throughout the training process to allow the model to focus on important facial regions. The training process was completed using a batch size of 65 to improve the efficiency of learning, using a momentum of 0.5 and an initial learning rate for optimization. Categorical cross entropy loss function was used to assess the accuracy of the model and evaluate how well it performed in predicting true class probability vs. the probability that it had predicted. Performance evaluation of the model was completed at the end of each epoch while training in batches of 32 images.

3.4 Assessment Metrics

The ability of the proposed emotion prediction model to accurately predict emotion was assessed on a subset of the test data from the dataset. When the model generated probability scores or continuous values, these were rounded to the closest whole number so that the appropriate category could be assigned. The model's predicted category was selected as the category for which the model had the highest confidence score; this would represent the model's best guess. To evaluate the models' emotion classifications, the following metrics were used: Recall, Accuracy, F-1 Score, and Confusion Matrix were applied to both SAMM and CASME-II Datasets.

4. RESULTS AND DISCUSSIONS

This research study discusses the results from the experimental study as well as an extensive review of the evaluation of the II benchmark data sets. The study determined which deep learning architectures (ResNet & VGG-16) can better find very small and subtle differences in an individuals emotional expression. It also found out whether using cluster analysis (K-Means) along with facial action units would increase overall accuracy for emotion recognition systems. To test the accuracy of the models, they were tested against two data sets (SAMM and CASME-II) and the accuracy was measured using both the accuracy matrix and the confusion matrix. The CASME-II data set consisted of five different emotion labels, including happiness, disgust, surprise, repression, and "all others". The SAMM data set included seven different emotion labels, including: happiness, anger, surprise, sadness, and "all others". The structures of both data sets were similar in that they were labeled in a manner that allowed them to be compared directly. Additionally, it was shown that the expert-created emotion labels for the CASME-II data set were more interpretable than those of the new model because of its ability to utilize facial action units to provide greater emotional granularity.

4.1 Emotion Classification with HybridMicroNet

The Confusion Matrices in Figures 1 and 2 depict the success rate of emotion classification model by the test data, based on CASME-II and SAMM datasets, to illustrate the model's ability to identify a variety of emotions using the Confusion Matrices. The Confusion Matrices are displayed such that the row values represent the true labels (actual emotion), the column values represent the predicted labels (predicted emotion) and the cell values represent the total number of examples that are identified correctly for an actual emotion (true label) and a predicted emotion (predicted label). The confusion matrix allows us to analyze how many correct classifications were achieved for each emotional state, as well as how many incorrect positive classifications were made and how many incorrect negative classifications were made for each emotion state. The color gradient used in conjunction with the confusion matrices provides a visual representation of how well the model has been able to make accurate classifications for each emotion state; with darker shades representing higher accuracy levels. The grouping of darker shaded areas along the diagonal of the matrices indicate high precision and accuracy with which the model was able to classify each emotion state.

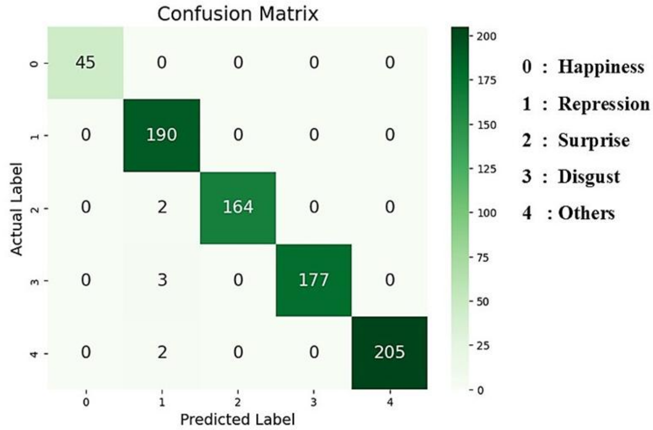


FIGURE 1. Confusion Matrix for CASME-II Dataset

As depicted in Fig. 1 the evaluation results demonstrate the model's ability to classify with excellent results. With regard to the emotion "Happiness", the model correctly identified each of the 45 instances tested. Additionally, the model perfectly classified all 190 instances of the emotion "Repression". For "Surprise," the model correctly classified 164 out of the 166 samples (only 2 were incorrectly classified as "Repression"). For "Disgust," the model correctly classified 177 out of 180 samples; the remaining 3 samples were mislabeled as "Repression." Finally, for "Others," the model correctly classified 205 out of the 207 samples, with only 2 errors that involved incorrectly labeling samples from "Repression." Overall, there were 768 test samples. The above-mentioned results indicate very high precision and low confusion among classes, particularly classes 0, 1, and 4, which further emphasizes the model's superior ability to identify subtle emotional differences. As is typical in micro-expression analysis, the small number of misclassifications that occurred between similar emotions, e.g., between classes 2 and 1, are also typical in this field due to the fact that even minor variations in facial expressions may result in overlapping images.

Figure 2 illustrates a Confusion Matrix that represents the success of emotion classification models, which are designed to categorize emotions into eight categories (including “Others”) and demonstrates the ability of the model to recognize seven emotional expressions (Surprise, Anger, Happiness, Fear, Disgust, Sadness, Contempt, and Others). Overall, the model did an excellent job across all classifications; however, the model did exceptionally well on correctly classifying Disgust, Fear, and Others, and achieved a high degree of correctness when evaluating happiness as 105 correct classifications were produced while only 5 were misclassified as surprise. In addition, there was very minimal crossover in the prediction of Surprise as 113 correct classifications were made. The model also demonstrated good accuracy in its ability to classify sadness, anger, and contempt with the model producing a small amount of error: one misclassified instance of Anger and Sadness, and two misclassified samples of Contempt as being happy. Recall, precision, accuracy, and F1-score were used to evaluate the HybridMicroNet model's performance based upon the results obtained from the CASME-II and SAMM datasets and the results are illustrated in table 4.1.

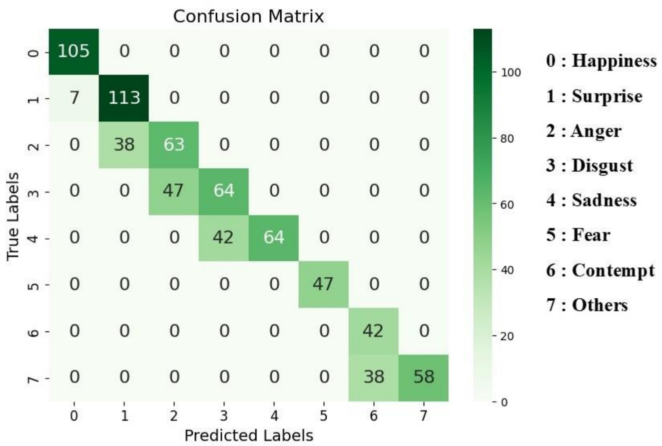


FIGURE 2. Confusion Matrix for SAMM Dataset

Table 1. Comparison of Accuracy Measures between SAMM and CASME-II Datasets

Dataset	Accuracy	Precision	Recall	F1-score
SAMM	97.52	99.99	97.63	0.97
CASME-II	99.07	99.13	99.07	0.98

4.2 Emotion Classification through Enhanced Model

The improved performance of the model was shown via a comparison of Confusion Matrices for both the CASME-II and SAMM Datasets (see Figure 3 and Figure 4). Each confusion matrix demonstrates the overall classification accuracy of the model on each respective emotion category within the two data sets. The confusion matrices showed that the Model performed well with respect to classification accuracy within the CASME-II Dataset (which has 5 emotion categories), as nearly all of the classifications were correct; likewise the model also classified the emotion categories of the SAMM Dataset (7 emotion categories) effectively with few misclassifications. This combination of Action Unit Features with Deep CNN-Based Feature Extraction significantly decreased confusion between visually similar micro-expression classifications. These results show the ability of the model to accurately classify subtle emotional differences and the models ability to perform accurate facial expressions classification.

The results of the Enhanced model's emotion recognition for the CASME-II database is presented in Figure 4.3. The test set was randomly selected from the full CASME-II dataset with 500 samples (5 emotion classes) using 30% of the data. The model was very successful in emotion classification as it correctly classified 92% of the 95 happiness samples and 93% of the 90 surprise samples. The Accuracy for each Emotion category was as follows: "Disgust," 94%; "Repression," 90%; "Sadness," 82%, while the overall results demonstrated a High degree of Reliability in identifying Subtle Variations in Facial Expressions associated with Micro-Expressions. Errors occurred in the Classification of Emotions across Categories; However, the Results indicate that the Enhanced Model has been Developed Robustly and can identify emotions with greater accuracy due to the inclusion of Action Units within its Architecture.

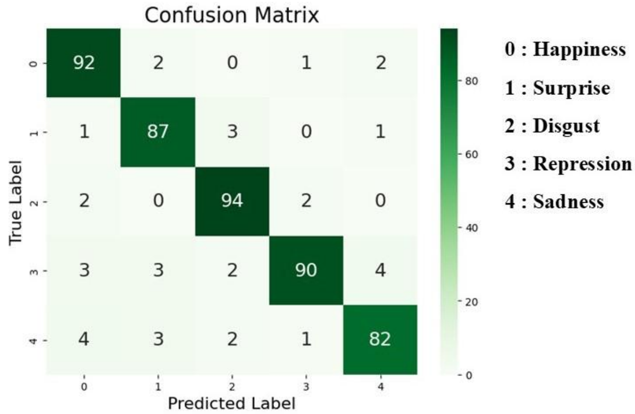


FIGURE3. Confusion Matrix for CASME Dataset

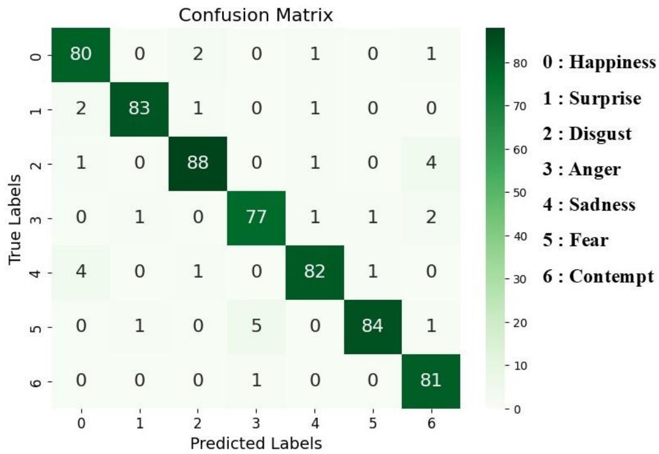


FIGURE 4. Confusion Matrix for SAMM Dataset

In Figure 4, the Confusion Matrix of the Enhanced model indicates how accurately it identified the 7 emotion categories of the SAMM dataset, including surprise, anger, happiness, sadness, disgust, contempt, and fear. As the Enhanced model utilized a subset of the dataset to test its accuracy, the 30% testing segment that included 560 images was used to evaluate the models overall performance. In each category of the Confusion Matrix, the Enhanced model performed well, as evidenced by its ability to correctly classify 80/82 of the happy samples, 83/85 of the surprised samples, and 88/92 of the disgusted samples. The strong classification performance was seen throughout the remaining five categories, including 77/80 classifications of anger with little crossover with other closely related emotions; 82/84 correct classifications of sadness with some degree of overlap between disgust, anger and happiness; and 84/87 correct classifications of fear and 81/83 correct classifications of contempt. Although there were several misclassifications, most were due to similarities or subtle differences between the micro-expressions of the closely related emotions. For example, contempt was confused with disgust four times and there were limited crossovers between fear, anger, and sadness. Overall, the Confusion Matrix clearly shows the dominance of the diagonal line, which demonstrates that the Enhanced Model has a high level of effectiveness in identifying the finer nuances of emotional expressions within the SAMM dataset.

Additionally, the table below compares the performance metrics of the Enhanced Model using both the SAMM and CASME-II datasets.

TABLE 2. Evaluating Performance Metrics in SAMM and CASME-II Datasets

Dataset	Accuracy	Precision	Recall	F1-score
SAMM	93.21	91.45	93.21	89.31
CASME-II	95.62	93.56	95.62	91.58

TABLE 3. Assessing the Improved Model's Performance Against Other Machine Learning Classifiers

Metrix	ANN	CNN	Decision Tree	SVM	Enhanced Model
Accuracy - SAMM	78.23	86.39	15.78	68.98	94.31
Accuracy - CASME II	82.43	88.15	19.65	76.97	96.72

4.3 Enhanced Model – An Ablation Study

The main goal of this research is to compare the performance of the proposed framework with that of several traditional machine learning techniques. An overview of how the proposed framework compares to three common machine learning techniques is shown in table 4.4, including Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Decision Trees (DT), and Support Vector Machines (SVM). The results show the Decision Trees produced the poorest accuracy level for all classification tasks at 18.54% as they are less effective than the other techniques when capturing the subtleties involved in micro-expressions; conversely, the new model performed significantly better and was able to accurately identify micro-expressions using all tested algorithms.

4.4 Results and Interpretation

The Confusion Matrix (Figure 5) is used to evaluate and compare the performance of emotion prediction models that were trained using the SAMM and CASME-II datasets as they were applied to predict emotions from a completely separate group of YouTube video samples. The Confusion Matrix illustrated in this figure shows how well the actual emotions match those that the models predicted (Surprise, Sadness, Fear and Disgust). The diagonal of the matrix illustrates all of the instances where each model correctly predicted the same emotion. The remaining values outside of the diagonal represent those instances in which the models predicted the same emotion differently. Overall, it appears that many of the emotions being studied were similarly identified across the sample set. Surprise was identified correctly 88% of the time; Sadness was identified correctly 83% of the time; Fear was identified correctly 86% of the time; and Disgust was identified correctly 86% of the time. There were some errors made by the models in their identification of certain emotions that have similar visual characteristics and/or features such as Sadness and Anger, or Anger and Disgust, because of the similarity of these expressions, and the overlap of features among them. Overall, the data indicates that the models performed reasonably well, and demonstrated an ability to generalize to new stimuli and conditions and that the method has substantial promise for use in real-world applications of emotion recognition. Additionally, there is evidence of slight variability in the models' performance that may indicate areas for improvement through refinement of the dataset, or optimization of the model.

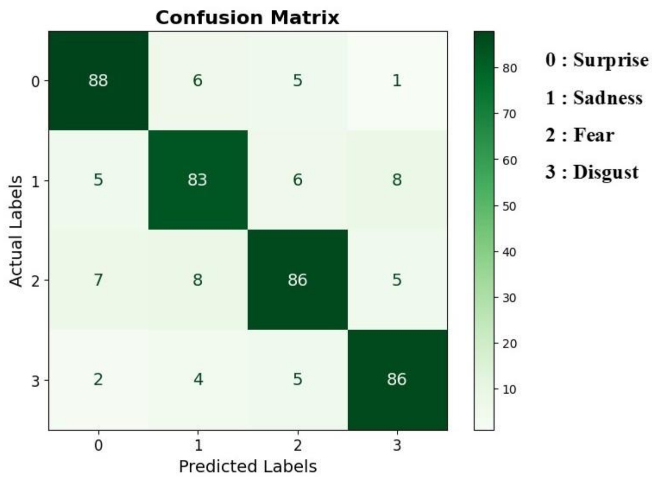


FIGURE 5. Confusion Matrix based on HybridMicroNet Framework

The SAMM and CASME-II models were independently applied to predict emotions from a video data set made up of YouTube videos of spontaneous emotional responses. The CASME-II model had an overall predictive consistency of 99.08% in cross-validation; whereas the SAMM model resulted in 97.62% accuracy when cross-validated. In the absence of ground-truth emotional labels for each of the YouTube video clips, a cross-model validation method was employed to assess the degree of agreement between the two models' predictions regarding emotional responses that are common to each other. The similarity in models is indicative of high reliability and of robust applicability outside of the confines of a controlled laboratory setting.

This case study also examined an advanced emotion recognition framework with a focus on the detection of emotional states within depressed persons. Sadness, disgust, and fear were the emotions identified as being the most frequent. This finding supports the documented research that suggests that individuals with depression tend to express emotions related to withdrawal, negative self-talk and aversive expression.

4. CONCLUSION

Studies using machine learning, along with micro-expression analysis for detecting mental health concerns show a great deal of promise. The facial micro-expressions, which are short-term, unintentional, and unconcealed facial expressions, also known as "true" or "honest" facial expressions, provide an opportunity for this area. HybridMicroNet and other models using action units (AU) for extracting emotional signals from images represent a streamlined and low-cost method of capturing emotional signals; however, both of these models were able to accurately detect very brief emotional expressions with HybridMicroNet achieving 99.08% and 97.62% accuracy on the CASME-II and SAMM databases, respectively, while the AU based model achieved 95.62% and 93.21% accuracy on the CASME-II and SAMM databases, respectively, and were successful in identifying common negative emotions associated with depression including, disgust, sadness, and anger. These studies suggest that models such as HybridMicroNet may be useful for early identification and treatment by providing a more accurate and timely assessment of emotional states.

REFERENCES

1. T. Shuai, “Advances in Facial Micro-Expression Detection and Recognition,” *Information*, vol. 16, no. 10, pp. 876, 2025. DOI: 10.3390/info16100876.
2. R. Jayaswal, “Advances in facial expression recognition technologies for human-machine interaction,” *Int. J. Educ. Technol. High. Ed.*, 2025. DOI: 10.1007/s10791-025-09699-8.
3. H. V. Manalu, et al., “Detection of human emotions through facial expressions using a hybrid CNN-RNN model,” *Computers & Electrical Engineering*, vol. 94, pp. 106378, 2024. DOI: 10.1016/j.compeleceng.2024.106378.
4. P. Malik, et al., “Action unit based micro-expression recognition framework,” *Scientific Reports*, vol. 15, no. 1, pp. 12245, 2025. DOI: 10.1038/s41598-025-12245-7.
5. S. Nikbin, et al., “A study on the accuracy of micro-expression based deception detection using hybrid deep neural networks,” *Eur. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 10201, 2024. DOI: 10.11591/ijece.v10i5.10201.
6. B. K. Durga, R. Vullanki, S. Jagannadham, et al., “Deep Learning Based Micro Facial Expression Recognition Using an Adaptive Tiefes FCNN Model,” *TS*, pp. 215-229, 2023. No DOI.
7. M. Xie, “Deep learning based Facial Micro-Expression Analysis: A Survey of Datasets and Algorithms,” *Curr Trends Biomedical Eng & Biosci.*, vol. 22, no. 3, pp. 100437, 2024. DOI: 10.1016/j.jbt.2024.100437.
8. F. Liu, “Artificial Intelligence in Emotion Quantification: multi-modal sources (facial, speech, text, physiological) and ML methods,” *AIR*, vol. 24, pp. 100120, 2024. DOI: 10.26499/air.2024.100120.
9. Study, “An Automated Human Emotion Detection via Micro-Expression Analysis,” 2025. No DOI.
10. A. Goel, et al., “Exploring automated machine learning to develop facial expression recognition systems,” *Discov Appl Sci*, vol. 14, pp. 106177, 2025. DOI: 10.1016/j.das.2025.106177.
11. “The Development and Applications of Facial Expression Recognition Algorithm Based on Deep Learning,” Master’s thesis, China, 2024. No DOI.
12. “A Survey on Facial Expression Recognition of Static and Video-Based (SFER & DFER),” 2024. No DOI.
13. “A comprehensive bibliometric survey of micro-expression recognition system based on deep learning,” *Heliyon*, vol. 10, no. 5, pp. e27392, 2024. DOI: 10.1016/j.heliyon.2024.e27392.
14. M. Keinert, et al., “Facial Emotion Recognition of 16 Distinct Emotions: A Deep Learning Approach,” *JMIR*, vol. 27, no. 3, pp. 69142, 2025. DOI: 10.2196/jmir.69142.
15. “A Review of Deep Learning Based Micro-Expression Classification,” 2023. No DOI.
16. R. Raj, et al., “An improved facial emotion recognition system using computer vision & machine learning,” *Sci Rep*, vol. 15, pp. 22835, 2025. DOI: 10.1038/s41598-025-22835-0.
17. “Methodical Review of Deep Learning Macro and Micro Face Expression Recognition,” 2025. No DOI.
18. “Facial Expression Analysis and Its Potentials in IoT Systems,” 2025. No DOI.
19. “Systematic Review of Emotion Detection with Computer Vision and Deep Learning,” *Sensors*, vol. 24, no. 1, pp. 276, 2024. DOI: 10.3390/s24010276.
20. “Development and application of emotion recognition technology,” 2023. DOI: 10.1016/j.enggeo.2023.02.013.