

RESEARCH ARTICLE



Micro-expression and Masked Expression Classification Using Neural Networks

Karan Suresh¹ , Shamyuktha Ramachandran Sugumaran¹ , Suja Palaniswamy^{1,*} , and Manju Priya Arthanarisamy Ramaswamy¹

¹ Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Bengaluru, India

Abstract: Micro-expressions are short, involuntary facial expressions unnoticeably occurring for a fraction of a second. They are highly valuable in various real-life situations since the facial muscles associated with micro-expressions cannot be consciously controlled. They can also aid psychiatrists in diagnosing patients or police officers in detecting the true emotions of suspects. In contrast, masked expressions are intended to hide genuine emotions, and these occur when someone uses a falsified expression to conceal their true feeling. In this study, a custom Micro Mask Emotion Dataset-Face Speech Text was developed for both micro-expressions and masked expressions using questionnaires and images as stimuli. The study participants were 18–25 years old and did not wear any face mask. The dataset was compiled and evaluated using various neural network models. The results indicate that females are more expressive and the questionnaire-induced dataset showed better prediction ability than the image-induced dataset on a similar set of neural network models.

Keywords: micro-expression, deep learning, masked expression, neural networks, hidden emotion

1. Introduction

Micro-expressions are fleeting facial expressions that emerge when an individual attempts to hide or restrain an emotion. These expressions are typically spontaneous and unfold rapidly, usually spanning from 1/25th to 1/5th of a second. Moreover, they can unveil a person's authentic feelings, often followed by subsequent displays of feigned or controlled emotions. The amygdala, a region in the brain responsible for processing emotions, plays a pivotal role in the manifestation of micro-expressions. When the amygdala reacts to a stimulus, it engenders an emotional response that can be simultaneously involuntary and voluntary and hence an internal conflict within the individual, leading to the brief emergence of a micro-expression that discloses their true emotional state before they conceal it with an alternative emotion. Despite conscious attempts to mask their genuine feelings, micro-expressions can be vexing to suppress or govern entirely. They unfold too swiftly for most individuals to consciously manipulate, rendering them challenging to discern without adequate training and observational acumen.

Leveraging datasets incorporating images of micro-expressions can confer advantages in comprehending an individual's thoughts, disposition, and character. By scrutinizing micro-expressions, researchers and experts can amass valuable insights into an individual's underlying emotions and motivations. For instance, if someone ruminates on something delightful or pleasing, they might momentarily exhibit an unconscious smile or a tranquil countenance. Conversely, if someone grapples with a disquieting matter, they might involuntarily expose expressions of anger, sadness, or apprehension. In particular scenarios, such as law enforcement inquiries, classification models

that recognize micro-expressions can be deployed to aid in gauging the veracity or potential deception of an individual. By dissecting the micro-expressions manifested during interrogations or interviews, investigators can glean insights into the individual's genuine emotions and evaluate whether they might be concealing information or engaging in deception. When deciphering a person's emotional state and conduct, micro-expressions should be considered and analyzed in conjunction with other contextual clues and behavioral manifestations for a more precise assessment. Concealed emotions represent a substantial concern in psychology. When individuals outwardly manifest one emotion while concurrently experiencing a different one internally, it can give rise to a multitude of issues, including potential harm to their well-being. The act of suppressing emotions transpires when individuals consciously or unconsciously stifle or obscure their genuine emotional responses for various reasons, such as societal expectations, apprehension of judgment or rejection, or a belief that the expression of emotions signify weakness or societal unacceptability. Nevertheless, protracted emotional suppression can exert adverse effects on mental and emotional well-being. It is imperative for individuals to cultivate emotional awareness and acquire healthy coping mechanisms for the expression and management of their emotions. This process may encompass seeking guidance from mental health professionals, participating in therapy or counselling, engaging in self-care practices, and acquiring techniques like mindfulness and emotional regulation. Thus, recognition of masked expression is useful in the treatment of people suffering from mental health issues and social anxiety.

Micro Mask Emotion Dataset-Face Speech Text (MMED-FST) is a custom in-house dataset that stores data for two forms of hidden emotions – micro-expression and masked expressions. MMED-FST is generated by recording videos using two distinct methods. In one method, participants are asked questions relating to anger, sadness, happiness, disgust, and contempt. Their facial expressions and speech invoked as a reply for the questions are recorded in video. In the other

*Corresponding author: Suja Palaniswamy, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Bengaluru, India. Email: p_suja@blr.amrita.edu

method, the participants are shown images to provoke emotional responses, which are recorded in video. Consequently, datasets are formed independently for both questionnaire-triggered emotions and image-induced emotions, thereby creating separate datasets for micro-expressions and masked expressions and forming four subsets of MMED-FST. Deep learning classification algorithms are then employed to train and evaluate the model's performance.

In this study, we aimed to investigate how emotions are triggered when questions are asked and images are shown. In the first method, there was no visual cue and the participants tended to relate the question to their experience or in general and try to answer by kindling their memories or their thoughts on the question. In the second method, the images shown to the participants triggered without any delay the emotion that was directly connected to their personal life and they respond very quickly through facial expressions and speech. If the emotion triggered was a negative emotion and if the participants did not let the observer to know it, they attempted to hide the real emotions, which still manifested but for a microsecond either in their face or through voice. The novelty and contribution of this work are as follows:

- 1) Custom in-house questionnaire- and image-induced dataset for micro-expression and masked expressions
- 2) Deep learning models to recognize micro expressions in genuine and masked expressions

2. Related Works

Limited work has been carried out in micro-expressions and masked expression domain through different modalities. In their comprehensive literature review of previous studies on automatic micro-expressions, Zhang and Arandjelović [1] emphasized the scarcity of accessible micro-expression dataset and summarized the neural network models that have been applied to the existing micro-expressions datasets.

Choi and Song [2] recognized facial micro-expressions using two-dimensional landmark feature maps, thereby demonstrating the advantage of tracking facial muscle changes. Notably, their approach combines two neural network algorithms: convolutional neural network (CNN) and long short-term memory (LSTM). Their research employed the SMIC and CASME datasets and achieved accuracy rates of 71% and 74%, respectively. For facial expression recognition, Shi et al. [3] presented a method based on a multiple branch cross-connected convolutional neural network (MBCC-CNN).

Recent work by Li et al. [4] proposed the fusion of hierarchical CNN and the robust principal component analysis (PCA)-based recurrent neural network (RNN) algorithm leveraging hierarchical structure of subnetworks. An extensive analysis of micro-expressions with an improved recognition accuracy is possible by extraction of space and time domain information using neural networks.

Ayyalasomayajula et al. [5] addressed the challenges of low intensity and short durations associated with micro-expressions by amplifying the micro-expressions to a predefined threshold by applying Eulerian video magnification (EVM) and CNN model. Takalkar and Xu [6] augmented the original dataset with CASME and CASME II and explored multiple neural network algorithms for training a micro-expression classification model. Kadakia et al. [7] evaluated the performance of pre-trained neural network models, ResNet50, VGG16, InceptionV3, MobileNet, and Xception, with the MobileNet algorithm achieving 99% accuracy.

Raval et al. [8] contributed to the growing body of research on emotion recognition of criminals and detect crime patterns with potential applications in judiciary decisions. Gupta et al. [9] and Mokshit et al. [10] proposed techniques for emotion recognition

using several neural network models, BERT and XAI tools. With these tools, important features that contribute to emotion recognition are extracted. Arthanarisamy Ramaswamy and Palaniswamy [11] compared the prediction ability of various physiological signals in subject independent environment and studied the influence of gender and modeling approach on emotion recognition using EOG and PPG. They concluded that modeling arousal and valence independently is better than joint modeling and inclusion of gender improves few of the model evaluation metrics [12].

Siddiqui et al. [13] developed a multimodal ER dataset that includes facial expressions, speech, physiological signals, body movements, gestures, and lexical features; images are taken using visible and infrared camera. Lian et al. [14] established recognition methods using face recognition, speech, and text. Das and Singh [15] presented a survey on methods, trends, and challenges on multimodal sentiment analysis. Ezzameli and Mahersia [16] carried out an exhaustive survey on multimodal datasets, deep learning methods, and fusion techniques and compared different modalities of emotion recognition and the datasets for the same. He et al. [17] reviewed the three types of multimodal affective BCI along with design, algorithms, and advances for the same. A recent review on multimodal emotion recognition also reiterated the limited availability of multimodal micro-expression datasets [18]. The existing micro-expression dataset, the stimulus, and emotion elicitation strategy are detailed in Table 1.

Through literature survey the following gaps have been identified: There is limited access to micro-expression dataset and the emotion categories of micro-expressions. Few studies on masked expression was conducted, which motivated the development of MMED-FST dataset that would facilitate various experiments on this area of research.

3. Dataset Description

A total of 21 college students (11 female and 10 male) in the age group of 18–22 years volunteered to participate in the video recording procedure. Each video is the spontaneous expression of the participants in 8–12 minutes. The distance between the camera and the participant was nearly 2 feet, and the illumination was set bright for the emotions to be captured. All four datasets as mentioned in Section 1 consist of five classes such as anger, sad, happiness, disgust, and contempt. The videos used were converted into frames using Jupyter Notebook (Python) where for every 0.5 second, a frame would be saved and labeled into appropriate classes. The difference between micro-expression and masked expression are as follows: for micro-expression, all facial muscles involved can be used as part of observation, whereas for

Table 1
Existing micro-expression datasets and the stimuli used to elicit emotion

Dataset	Emotion elicitation strategy	Stimuli
USD-HD [19]	Posed	Video
Polikovskiy's database [20]	Posed	Instruction
SMIC [21]	Spontaneous	Video
SMIC extension [22]	Spontaneous	Video
CASME [23]	Spontaneous	Video
CASME II [24]	Spontaneous	Video
MFED [25]	Posed	Video
Proposed MMED-FST*	Spontaneous	Images & Questionnaire

masked expression, the mouth region is ignored and more importance is given to eye region as the participant could be smiling but the eye and eyebrows could be signifying the real expression. To record the micro-expressions and masked expressions from the participants, the questionnaire and images for each of five classes (anger, sad, happiness, disgust, and contempt) were prepared. The participants were informed in detail about the whole process, and their consent was obtained prior to video recording. The expression of the participants for the questions and images was captured, reviewed, curated, and labeled to create the datasets for micro-expression and masked expression. A total of 2096 image frames for all emotion classes were extracted.

3.1. Questionnaire-induced datasets for micro-expression and masked expression

To create questionnaire-induced datasets, questions were posed such that the participants would answer the question that makes them express subconsciously, which was captured in the video even if the expressions last for a shorter duration. Sample questions asked to induce the emotions are illustrated in Figure 1. The questions were prepared on a lighter note and not critical. In their responses, some participants shared about their personal life, which helped in capturing genuine expressions.

3.1.1. Questionnaire-induced micro-expression dataset

The micro-expression dataset was created by observing the frames created from the video using Python code. For every 0.5 second, there was a frame that means that the 30th frame was the 15th second of the video. This way, mapping of frames with videos was performed to understand the type of questions that were asked during that particular time frame. Micro-expression lasts for milliseconds, which makes it harder to observe in video but gets perfectly captured in the frames. Such observation was true for all images, and the images indicating micro-

expressions were classified accordingly. Sample images are shown in Figure 2. Further, this dataset was divided into male and female groups to further identify the effect of questions. In Figure 2, subtle expressions can be observed. The most expressive expression is happiness.

3.1.2. Questionnaire-induced masked expression dataset

The masked expression dataset was created by carefully analyzing the recorded video. Moments when the participants appeared to be smiling were scrutinized to determine whether they were concealing either anger or sadness. Snapshots capturing such deceptive expressions were collected as integral components of the dataset. Creating the masked expression dataset proved to be a formidable challenge due to the inherent difficulty in deciphering concealed emotions. Humans often mask their true feelings, especially when they have no intention of expressing those emotions outwardly. Instead, they frequently resort to smiles or laughter as a means of veiling their genuine emotions. Recognizing these concealed emotional states holds significant potential, as it could aid psychiatrists in diagnosing patients grappling with depression and identifying distinct personality traits, among other applications.

Within the masked expression dataset, only two distinct classes have been established. The first class pertains to instances categorized as "Happy but actually Sad," signifying that the outward expression appears cheerful, yet the underlying emotion is sadness. The second class is designated as "Happy but actually Angry," encapsulating situations where individuals exhibit happiness on the surface but harbor concealed anger. Figure 3 provides a sample visual representation of the two masked expression classes.

Figure 1

Sample questions used to create the questionnaire-induced micro-expression and masked expression datasets

- Who is your hero? - What are you most proud of? - What is your favorite place you've ever visited?
Sad - How do you deal with failure? - What is your greatest personal obstacle? - How do you offer sympathy to your friends when they're experiencing life challenges?
Anger - What do you think is the most significant issue society is currently facing? - What makes you angry? - How do you feel if your friend is super late for your party?
Disgust - Would you rather go without shampoo for the rest of your life or toothpaste for the rest of your life? - Would you rather be extremely allergic to your favorite food or forced to eat your least favorite food once a week?
Contempt - What word would people close to you use to describe you? - Which question people ask you often? - What do you consider your greatest accomplishment?

Figure 2

Sample images of micro-expression dataset

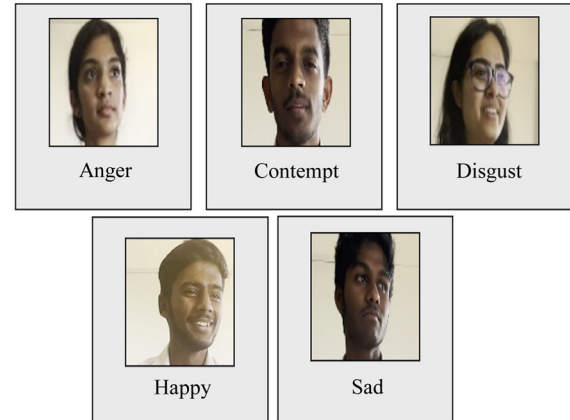


Figure 3

Sample images of masked expression dataset



3.2. Image-induced datasets for micro-expression and masked expression

Both micro-expression and masked expression datasets were created by selecting and organizing frames in a way that closely resembles that of the question-based dataset. However, instead of questions, images from the same categories were employed to elicit facial expressions. Importantly, these images were intentionally chosen to avoid any intense or potentially distressing stimuli for the participants. All images used to provoke emotional responses were of a lighter nature; sample images are shown in Figure 4.

In contrast to the question-based approach, using images to elicit expressions from participants was more intricate owing to the need for the participants to connect with the images in order to trigger observable expressions. As the process advanced, discernible expressions did emerge.

3.2.1. Image-induced micro-expression dataset

The images were similar to the questionnaire-based micro-expression dataset, but the process of identifying micro-expression from image-based videos was more complex as expressions were less observed, resulting a dataset smaller than the dataset for question-based micro-expression. Further, this dataset was divided according to the participant's gender to determine the effect of images on each gender. Sample images from image-induced micro-expression dataset are shown in Figure 5.

3.2.2. Image-induced masked expression dataset

In the process of creating masked expression dataset by showing images, scenarios were considered to identify masked expressions where the participant was actually talking about something sad or angry

but smiled, which is indicated as happy emotion. Sample images from image induced masked expression dataset are shown in Figure 6.

4. Proposed Framework

The proposed concept involves the utilization of sub-datasets within the MMED-FST framework to evaluate the effectiveness of different neural network models. Specifically, the micro-expressions dataset was partitioned based on gender in order to assess how questions influence male and female participants differently and determine which gender displays more pronounced expressions in response. Various models used in this study were custom CNN, VGG-19, RESNET-50, INCEPTION V3, VGG-16, and EfficientNet B0. Apart from the CNN, the remaining models fall under the category of transfer learning models, implying that they have been pre-trained with a default dataset; however, their parameters were adjusted according to the specifics of the custom datasets.

The proposed framework is shown in Figure 7. The initial step involved segregating the datasets into training and testing subsets. For training and testing, 80% and 20% of the samples was used, respectively. The model was trained for the specified number of epochs using the pre-processed training dataset, and the model performance was evaluated using the testing dataset. Image dimension configuration and grayscale conversion were the pre-processing steps applied on the training set.

4.1. Custom convolutional neural network

Table 2 shows the layers and parameters of the custom CNN model used in this study. The image was converted to vectors, and maximum length padding was performed as pre-processing steps. The processed vector was passed as input to the model. The different layers of CNN were convolution, pooling, and fully connected layer. Features were extracted in the convolution layers, which were then sub-sampled in pooling layers. The resulting array of the pooling layers was then flattened, and highest values were extracted in fully connected layers, which classified the output using Softmax function. The parameters were selected on experimental basis, and those that yielded the highest accuracy were mentioned in the Results section.

The sparse categorical entropy and Adam were used as loss function and optimizer, respectively. The Adam optimizer was applied for its bias correction, dynamic learning rates, and the momentum that accelerates the training process. Sparse categorical cross-entropy

Figure 4
Sample images used as stimuli to induce micro-expression and masked expression



Figure 5
Sample images of micro-expression dataset

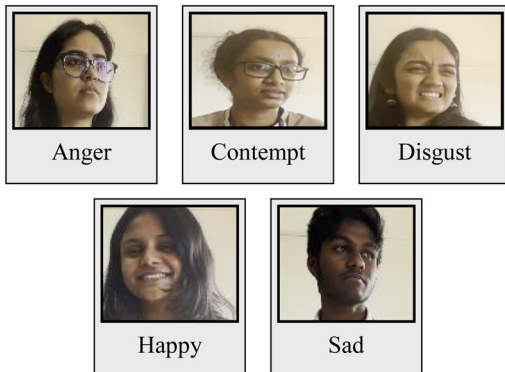


Figure 6
Sample images of masked expression dataset

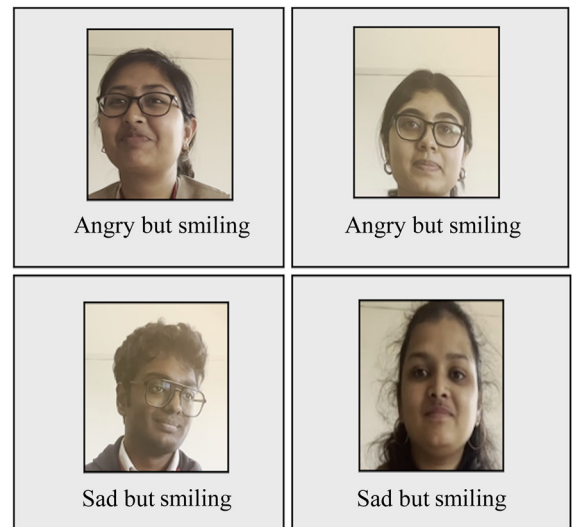


Figure 7
Proposed framework

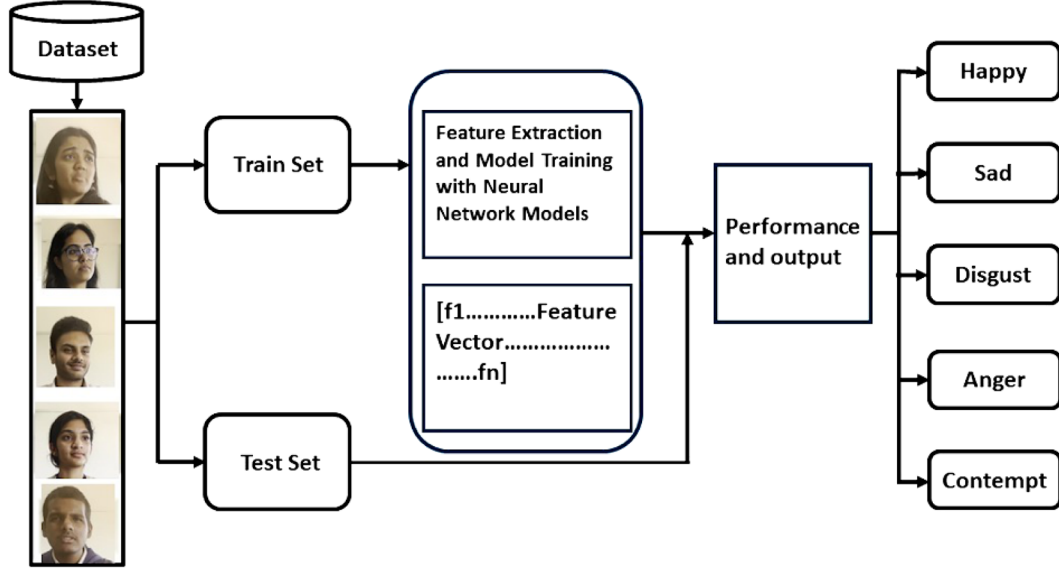


Table 2
Architecture of custom CNN

Layers (in the starting order)	Parameters for each layer
Convolution	Filter =16, Kernel size=3, activation = relu
Max Pooling	Pool Size = (2,2), strides = None, padding = valid
Convolution	Filter =32, Kernel size=3, activation = relu
Max Pooling 2D	Pool Size = (2,2), strides = None, padding = valid
Convolution	Filter =64, Kernel size=3, activation = relu
Max Pooling 2D	Pool Size = (2,2), strides = None, padding = valid
Dense	Units = 128, activation = relu
Dense	Units = number of classes, activation = relu

computed the cross-entropy loss by measuring dissimilarity among the predicted probability distribution and the target class label. This allowed the model to optimize the parameters by reducing the difference among predicted probabilities and the ground truth class label.

4.2. Pre-trained neural network models

The various pre-trained deep neural network models employed were RESNET-50, VGG-16, INCEPTION V3, VGG-19, and EfficientNet B0. EfficientNet B0 [26] applies a compound scaling to modify the depth, width, and resolution. All dimensions have fixed scaling coefficients and scaling of those aid the model to determine fairness among model capacity and computational efficiency. This assists in the extraction of intricate characteristics present in the datasets, which leads to better accuracy. In addition, EfficientNet B0 includes different regularization techniques like drop-out, which is part of the loss function that aided to minimize overfitting, thus preparing the model to generalize the custom dataset and increase the accuracy on the two datasets considered in this study.

VGG-16 and VGG-19 [27] consist of important deeper architectures but less convolutional filters that enhanced the capturing of local details of the dataset are used in this study. This eliminated

overfitting and resulted in better accuracy. RESNET-50 [28] was used to skip connections that solved the vanishing gradient issue. The 50 convolutional layers extracted information from samples. Default parameters were applied on custom datasets and found to be collaborative with the dataset. The bottleneck part of RESNET is well suited for extracting low and high-level features. Inception V3 [29] consists of several filters of sizes 1×1 , 3×3 , and 5×5 , which enables it to extract information of different resolution that will aid in improving the results. Default parameters were applied in the models used as they are in tune with the custom dataset MMED-FST and are shown in Table 3.

4.3. Performance metrics used

To measure the performance of the models, Accuracy, F1-score, Precision, and Recall were used. Accuracy is a basic metric used for evaluation. F1-score is considered to be a better metric for multiclass classification because it takes the values of precision and recall, which are focused on true positive, true negative, false positive, and false negative.

4.3.1. Accuracy

Accuracy is one of the most common measure employed and the formula for the same is provided in equation (1).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (1)$$

4.3.2. F1-score

F1 score is used to combine the precision and recall to evaluate the performance accuracy of the model as shown in equation (2).

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2)$$

Table 3
Parameters used in pre-trained architectures

Parameters	Default
Weight	ImageNet
Optimizer	Adam
Loss function	Sparse categorical cross entropy

4.3.3. Precision

Precision elucidates the model's performance on number of positive class predictions that are actually positive model as shown in equation (3).

$$Precision = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Positives}} \quad (3)$$

4.3.4. Recall

Recall explicates the model's performance on number of positive class predictions from all positive classes as shown in equation (4).

$$Recall = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Negatives}} \quad (4)$$

5. Results and Analysis

Multiple neural network models mentioned in earlier section were trained and evaluated on the test data. Different studies conducted to understand the impact of the questions and image based induced emotions on the subjects are discussed in this section.

5.1. Questionnaire- and image-induced micro-expression dataset

Table 4 details the average accuracy and F1-Score for the questionnaire and image-induced micro-expression datasets for the various models used in this study. Resnet-50 and VGG-16 performed well on the questionnaire-induced micro-expression dataset, which could be attributed to the depth of its network, simplicity of architecture, and generalization capabilities. The utilization of VGG-16 has demonstrated its advantages in smaller datasets, potentially contributing to its enhanced learning performance. For image-induced dataset, ResNet-50 and F1-score showed the highest accuracy at 77%, which could be attributed to the characteristics of the dataset that made the model perform better comparatively. In some cases, a deeper and more complex model like ResNet-50 can extract more intricate patterns and features, tending toward performance improvement. The questionnaire-induced dataset outperformed the image-induced dataset, indicating that the questions had more impact on the participants expressing their emotions, resulting in more identifiable expressions than image-based approach.

Figures 8–13 represent the classification analysis for a few of the models for micro-expression dataset. Figures 14 and 15 represent the classification analysis for masked expression dataset. In Figure 8,

Table 4

Performance of questionnaire induced and image induced micro expression dataset

Model	Questionnaire-induced dataset		Image-induced dataset	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
CNN	75	74	74	74
RESNET-50	80	77	77	77
VGG-19	76	76	75	74
VGG-16	79	79	74	74
INCEPTION V3	20	14	25	10
EfficientNet B0	78	77	75	74

Figure 8

Classification analysis of VGG-16 for question-induced dataset

	precision	recall	f1- score	support
Anger	0.68	0.74	0.71	88
Contempt	0.77	0.80	0.78	86
Disgust	0.80	0.92	0.85	72
Happiness	0.95	0.85	0.90	89
Sad	0.76	0.64	0.70	84
accuracy			0.79	419
macro avg	0.79	0.79	0.79	419
weighted avg	0.79	0.79	0.79	419

Figure 9

Classification analysis of RESNET-50 for image based dataset

	precision	recall	f1- score	support
Anger	0.67	0.69	0.68	70
Contempt	0.79	0.80	0.79	88
Disgust	0.75	0.71	0.73	66
Happiness	0.81	0.91	0.86	87
Sad	0.84	0.72	0.78	72
accuracy			0.77	383
macro avg	0.77	0.76	0.77	383
weighted avg	0.77	0.77	0.77	383

Figure 10

Classification analysis of VGG-19 for question-induced dataset

	precision	recall	f1- score	support
Anger	0.95	0.72	0.82	50
Contempt	0.90	0.75	0.82	48
Disgust	0.98	0.84	0.90	50
Happiness	0.84	1.00	0.91	48
Sad	0.62	0.92	0.74	38
accuracy			0.84	234
macro avg	0.86	0.85	0.84	234
weighted avg	0.87	0.84	0.84	234

Figure 11

Classification analysis of RESNET-50 for question-induced dataset

	precision	recall	f1- score	support
Anger	0.62	0.81	0.71	43
Contempt	0.85	0.79	0.81	42
Disgust	0.84	0.71	0.77	52
Happiness	0.87	0.98	0.92	41
Sad	0.90	0.72	0.80	36
accuracy			0.80	214
macro avg	0.82	0.80	0.80	214
weighted avg	0.81	0.80	0.80	214

Figure 12

Classification analysis of EfficientNet for question-induced dataset

	precision	recall	f1- score	support
Anger	0.54	0.88	0.67	43
Contempt	0.87	0.69	0.77	39
Disgust	1.00	0.83	0.91	36
Happiness	0.85	1.00	0.92	29
Sad	0.80	0.42	0.55	38
accuracy			0.76	185
macro avg	0.81	0.77	0.77	185
weighted avg	0.80	0.76	0.75	185

Figure 13

Classification analysis of CNN for image-induced dataset

	precision	recall	f1- score	support
Anger	0.81	0.68	0.74	38
Contempt	0.78	0.78	0.78	32
Disgust	0.76	0.74	0.75	38
Happiness	0.73	0.88	0.80	24
Sad	0.73	0.73	0.73	26
accuracy			0.76	168
macro avg	0.76	0.76	0.76	168
weighted avg	0.76	0.76	0.76	168

Figure 14

Classification analysis of VGG-16 for question-induced dataset (masked expressions)

	precision	recall	f1- score	support
Happy anger	0.77	0.94	0.85	54
Happy sad	0.87	0.57	0.69	35
accuracy			0.80	89
macro avg	0.82	0.76	0.77	89
weighted avg	0.81	0.80	0.79	89

Figure 15

Classification analysis of VGG-16 for image-induced dataset (masked expressions)

	precision	recall	f1- score	support
Happy anger	0.78	0.76	0.77	33
Happy sad	0.68	0.71	0.69	24
accuracy			0.74	57
macro avg	0.73	0.73	0.73	57
weighted avg	0.74	0.74	0.74	57

happiness was predicted most accurately, while anger was the least well predicted; happiness micro-expression was expressed more than the other classes, which helped the model to learn better given improved results than other classes. From the classification analysis in Figure 9, it is evident that the predictions for Sad achieved the highest accuracy, suggesting that the images depicting sadness had a more pronounced impact on eliciting sad expressions than the other images, for both male and female participants.

5.2. Questionnaire- and image-induced micro-expression dataset comprising only female participants

Table 5 shows that VGG-19 resulted in a high average accuracy and F1-score of 86% and 84%, respectively, for the questionnaire-based dataset, which is comprehensively more than the other models. The reasons for VGG-19 performance could be the same reasons as mentioned for VGG-16, but for the smaller custom dataset used in this study, the additional three layers in VGG-19 could have enabled better learning. For image-based dataset, RESNET-50 performance was maximum with an accuracy and F1-score of 82% and 80% respectively. The superior performance of ResNet-50 compared to other models could be attributed to the dataset characteristics and skip connections that favored its performance. ResNet-50, being a shallow model compared to others, has a more intricate architecture, which enables ResNet-50 to capture finer patterns and features, potentially resulting in enhanced performance. Moreover, the question-based dataset provides better results like the previous case.

Anger and Disgust among female participants were the most accurately predicted by the model, whereas Sad was not predicted very well. This indicates that the micro-expressions for anger and disgust given by female participants are more expressive than the other classes.

From Figure 11, the observations reveal that the model accurately predicted emotions such as sadness, disgust, and happiness in female participants, while anger was not predicted as effectively. The micro-expressions for sadness, happiness, and disgust exhibited by female participants are more pronounced than for other expressions. Additionally, it indicates that female participants in the age group 18–22 years tend to express their emotions more through sadness or happiness rather than anger. Such insights can be valuable for psychiatrists in assessing and treating individuals experiencing specific emotional states.

5.3. Questionnaire- and image-induced micro-expression datasets comprising only male participants

Table 6 shows that EfficientNet B0 performs best among all models with an average accuracy and F1-score of 81% and 78% for the questionnaire based dataset. This male only dataset is the smallest among all. The advantage of the EfficientNet is the compound scaling parameters that helped in better feature learning and is more suitable for this dataset. For image-induced dataset, CNN performed best among all models with an accuracy and F1-score of 76%. The combinations of conv2d layer and max pooling followed by two dense layers helped in filtering out the noise in the feature leading to superior performance of this model compared to the pre-trained model.

Table 5
Performance of questionnaire- and image-induced micro-expression dataset comprising only female participants

Model	Questionnaire-induced dataset		Image-induced dataset	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
CNN	81	80	77	76
RESNET-50	81	81	82	80
VGG-19	86	84	79	78
VGG-16	82	79	80	79
INCEPTION V3	30	28	33	19
EfficientNet B0	77	77	67	66

Table 6
Performance of questionnaire- and image-induced micro-expression dataset comprising only male participants

Model	Questionnaire-induced dataset		Image-induced dataset	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
CNN	78	78	76	76
RESNET-50	78	76	76	72
VGG-19	77	77	72	71
VGG-16	78	74	73	72
INCEPTION V3	58	47	50	42
EfficientNet B0	81	78	70	68

From the classification analysis of questionnaire-induced dataset (Figure 12), disgust was perfectly predicted, whereas anger was not predicted well, which indicates that male participants are not expressive of anger.

From the classification analysis of the image-induced dataset (Figure 13), anger was well predicted and the prediction rate for other classes was not significantly different. This could indicate that all emotions triggered by the images of different classes are almost equally expressive for men. Like the previous cases, results are better for questionnaire-induced dataset, indicating that questions have more impact on the participants than images.

5.4. Questionnaire- and image-induced masked expression dataset

Table 7 shows that VGG-16 performed best with average accuracy and F1-score of 82% and 77%, respectively, in questionnaire-based masked emotion dataset with accuracy and F1-score of 73% in image-based masked emotion dataset. Its superior performance can be attributed to several factors. Firstly, the depth of its network, simplicity of architecture, and generalization capabilities contribute to its high accuracy. The deep structure of VGG-16 allows it to capture intricate features, while its straightforward architecture facilitates effective training and understanding. Additionally, VGG-16 has demonstrated the ability to generalize well, making it suitable for various tasks. Moreover, the transfer learning aspect of VGG-16, leveraging its pre-trained weights on large-scale datasets like ImageNet, proves advantageous when dealing with smaller datasets, enhancing its learning capabilities.

Table 7
Performance of questionnaire- and image-induced masked expression dataset

Model	Questionnaire-induced dataset		Image-induced dataset	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
CNN	72	68	69	68
RESNET-50	78	73	68	68
VGG-19	79	71	73	72
VGG-16	82	77	73	73
INCEPTION V3	68	66	70	70
EfficientNet B0	75	76	60	60

From classification report of questionnaire-induced dataset based on F1-score (Figure 14), it is observed that suppressing sadness while showing happy emotion is more obvious than the one with suppressing anger while showing happy emotion. The reason could be that the sad expressions suppressed by happiness are more expressed than the suppressed anger, enabling the model to learn better.

From classification analysis based on F1-score for image-based dataset (Figure 15), it is observed that suppressing sadness while showing happy emotion is less than the one with suppressing anger while showing a happy emotion. The model training and performance are done on various cases. Among male and female participants, the female participants were more expressive and less inclined to suppression than the male participants.

6. Conclusion

The MMED-FST has been developed by asking questions and showing images to induce the emotions of the individual. The objective of developing this dataset was to study the process of how emotions are triggered spontaneously, whether people are comfortable in expressing negative emotions or do they try to hide their emotions, to detect subtle micro-expressions and which gender is more expressive. To the best of the knowledge of the authors, this is the first dataset of this kind. Various neural network algorithms were trained and assessed for their performance. The results indicate that RESNET-50 performed better in terms of accuracy for micro-expression dataset, and VGG-16 performed better for masked expression, demonstrating the influence of questions and images on individuals. Additionally, models were trained based on gender to determine how images shown to induce emotions affect male and female individuals, further showing the differing reactions of male and female to specific questions and images, as well as which gender tends to exhibit greater expressiveness. The results of this study revealed similarities in the expression of certain emotions, such as anger and sadness, among male individuals, highlighting commonalities in their emotional expression patterns. The dataset that focuses on the emotions, which are suppressed while expressing different emotions, are difficult to produce, but once done, it will be beneficial in treating patients going through depression or anxiety. Future work could be to extend the size of the existing dataset to make the neural network models learn better. The audio and text from the videos could be used to create speech and text datasets for further emotion classifications using speech and text.

Ethical Statement

All subjects provided informed consent for inclusion before participating in the study. The study was conducted in accordance with the Declaration of Helsinki.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Karan Suresh: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft. **Shamyuktha Ramachandran Sugumaran:** Conceptualization, Methodology, Software, Validation, Formal

analysis, Investigation, Resources, Data curation, Writing – original draft. **Suja Palaniswamy**: Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Manju Priya Arthanarisamy Ramaswamy**: Writing – review & editing, Visualization.

References

- [1] Zhang, L., & Arandjelović, O. (2021). Review of automatic microexpression recognition in the past decade. *Machine Learning and Knowledge Extraction*, 3(2), 414–434. <https://doi.org/10.3390/make3020021>
- [2] Choi, D. Y., & Song, B. C. (2020). Facial micro-expression recognition using two-dimensional landmark feature maps. *IEEE Access*, 8, 121549–121563. <https://doi.org/10.1109/ACCESS.2020.3006958>
- [3] Shi, C., Tan, C., & Wang, L. (2021). A facial expression recognition method based on a multibranch cross-connection convolutional neural network. *IEEE Access*, 9, 39255–39274. <https://doi.org/10.1109/ACCESS.2021.3063493>
- [4] Li, J., Wang, T., & Wang, S.-J. (2022). Facial micro-expression recognition based on deep local-holistic network. *Applied Sciences*, 12(9), 4643. <https://doi.org/10.3390/app12094643>
- [5] Ayyalasomayajula, S. C., Ionescu, B., & Ionescu, D. (2021). A CNN approach to micro-expressions detection. In *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics*, 345–350. <https://doi.org/10.1109/SACI51354.2021.9465542>
- [6] Takalkar, M. A., & Xu, M. (2017). Image based facial micro-expression recognition using deep learning on small datasets. In *2017 International Conference on Digital Image Computing: Techniques and Applications*, 1–7. <https://doi.org/10.1109/DICTA.2017.8227443>
- [7] Kadakia, R., Kalkotwar, P., Jhaveri, P., Patanwadia, R., & Srivastava, K. (2021). Comparative analysis of micro expression recognition using deep learning and transfer learning. In *2021 2nd Global Conference for Advancement in Technology*, 1–7. <https://doi.org/10.1109/GCAT52182.2021.9587731>
- [8] Raval, J., Jadav, N. K., Tanwar, S., Pau, G., Alqahtani, F., & Tolba, A. (2025). Criminal emotion detection framework using convolutional neural network for public safety. *Scientific Reports*, 15(1), 15279. <https://doi.org/10.1038/s41598-025-97879-3>
- [9] Gupta, P. K., Varadharajan, N., Ajith, K., Singh, T., & Patra, P. (2024). Facial emotion recognition using computer vision techniques. In *2024 15th International Conference on Computing Communication and Networking Technologies*, 1–7. <https://doi.org/10.1109/ICCCNT61001.2024.10725699>
- [10] Mokshit, P., Prasad, P. S., Reddy, N. S., & Singh, T. (2024). Deep emotion analysis: “Enhancing emotion classification with transformer model”. In *2024 15th International Conference on Computing Communication and Networking Technologies*, 1–7. <https://doi.org/10.1109/ICCCNT61001.2024.10723323>
- [11] Arthanarisamy Ramaswamy, M. P., & Palaniswamy, S. (2022). Subject independent emotion recognition using EEG and physiological signals—A comparative study. *Applied Computing and Informatics*. <https://doi.org/10.1108/ACI-03-2022-0080>
- [12] Ramaswamy, M. P. A., & Palaniswamy, S. (2024). Unveiling the influence of modelling approach and gender in subject independent multimodal emotion recognition using EOG and PPG. *IEEE Access*, 12, 177342–177354. <https://doi.org/10.1109/ACCESS.2024.3506157>
- [13] Siddiqui, M. F. H., Dhakal, P., Yang, X., & Javaid, A. Y. (2022). A survey on databases for multimodal emotion recognition and an introduction to the VIRI (visible and InfraRed image) database. *Multimodal Technologies and Interaction*, 6(6), 47. <https://doi.org/10.3390/mti6060047>
- [14] Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C., & Zong, Y. (2023). A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. *Entropy*, 25(10), 1440. <https://doi.org/10.3390/e25101440>
- [15] Das, R., & Singh, T. D. (2023). Multimodal sentiment analysis: A survey of methods, trends, and challenges. *ACM Computing Surveys*, 55, 270. <https://doi.org/10.1145/3586075>
- [16] Ezzameli, K., & Mahersia, H. (2023). Emotion recognition from unimodal to multimodal analysis: A review. *Information Fusion*, 99, 101847. <https://doi.org/10.1016/j.inffus.2023.101847>
- [17] He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain–computer interfaces. *Brain Sciences*, 10(10), 687. <https://doi.org/10.3390/brainsci10100687>
- [18] Ramaswamy, M. P. A., & Palaniswamy, S. (2024). Multimodal emotion recognition: A comprehensive review, trends, and challenges. *WIREs: Data Mining and Knowledge Discovery*, 14(6), e1563. <https://doi.org/10.1002/widm.1563>
- [19] Shreve, M., Godavarthy, S., Goldgof, D., & Sarkar, S. (2011). Macro-and micro-expression spotting in long videos using spatio-temporal strain. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition*, 51–56. <https://doi.org/10.1109/FG.2011.5771451>
- [20] Polikovsky, S., Kameda, Y., & Ohta, Y. (2009). Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In *3rd International Conference on Imaging for Crime Detection and Prevention*, 1–6. <https://doi.org/10.1049/ic.2009.0244>
- [21] Pfister, T., Li, X., Zhao, G., & Pietikäinen, M. (2011). Recognising spontaneous facial micro-expressions. In *2011 International Conference on Computer Vision*, 1449–1456. <https://doi.org/10.1109/ICCV.2011.6126401>
- [22] Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 1–6. IEEE. <https://doi.org/10.1109/FG.2013.6553717>
- [23] Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., & Fu, X. (2013). CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 1–7. <https://doi.org/10.1109/FG.2013.6553799>
- [24] Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE*, 9(1), e86041. <https://doi.org/10.1371/journal.pone.0086041>
- [25] Mo, F., Zhang, Z., Chen, T., Zhao, K., & Fu, X. (2021). MFED: A database for masked facial expression. *IEEE Access*, 9, 96279–96287. <https://doi.org/10.1109/ACCESS.2021.3091289>
- [26] Hoang, V.-T., & Jo, K.-H. (2021). Practical analysis on architecture of EfficientNet. In *2021 14th International Conference on Human System Interaction*, 1–4. <https://doi.org/10.1109/HSI52170.2021.9538782>
- [27] Bagaskara, A., & Suryanegara, M. (2021). Evaluation of VGG-16 and VGG-19 deep learning architecture for classifying dementia people. In *2021 4th International Con-*

- ference of Computer and Informatics Engineering*, 1–4. <https://doi.org/10.1109/IC2IE53219.2021.9649132>
- [28] Zakaria, N., Mohamed, F., Abdelghani, R., & Sundaraj, K. (2021). VGG16, ResNet-50, and GoogLeNet deep learning architecture for breathing sound classification: A comparative study. In *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy*, 1–6. <https://doi.org/10.1109/AI-CSP52968.2021.9671124>
- [29] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>

How to Cite: Suresh, K., Ramachandran Sugumaran, S., Palaniswamy, S., & Arthanarisamy Ramaswamy, M. P. (2025). Micro-expression and Masked Expression Classification Using Neural Networks. *Journal of Computational and Cognitive Engineering*. <https://doi.org/10.47852/bonviewJCCE52026148>