Facial Emotion Recognition Using Convolutional Neural Networks for Human-Machine Interaction

Vienna Parnell

I. Introduction

Emotions are critical features in effective communication. The process of expressing and detecting emotional states enhances interpersonal interactions and bolsters strong relationships that are unique to human behavior. In transcending ethical differences and cultural barriers, emotions serve as a universal language in regular conversation. While some signals are more reliable and genuine than others, emotional expression can frequently convey an individual's mental state by nonverbally reflecting his or her intentions (Mehta et al., 2018). Given the rising prominence of artificial intelligence (AI) and affective computing in everyday life, simulation of natural interaction has become particularly valued in computer-based applications. By accurately interpreting emotional states, machines have the potential to provide objective psychological analyses in various fields. Efforts are mainly directed toward the detection and interpretation of human emotions for the purposes of imitation and response. Applications include human-computer interfaces (Cowie et al., 2001), special needs cases (Hantke et al., 2018), diagnosis of mental disorders (Steppan et al., 2020), security systems (Clavel et al., 2008), and animation generation (Deepali, et al., 2016).

There are several methods of performing emotion recognition (ER), including the assessment of tone of voice, body language, written language, and brain activity. Most success can be achieved in facial expression detection, mainly due to its universality, visibility, and availability of datasets. This consensus first emerged in the 1970s, popularized by Professor Albert Mehrabian's studies. Considered a pioneer in the field of nonverbal communication, Mehrabian believed that effective face-to-face communication required three fundamental elements: body language, tone of voice, and the words themselves. He further elaborated that these components varied in relevance, with linguistic language accounting for 7% of communication; intonation, 38%; and body language and facial expressions, 55%. While the 7-38-55% model is decades old, the consensus among psychologists today is still that a significant majority of communication is nonverbal (Mehrabian, 1971). As compared to word choice and tone of voice, visible appearance can be the most reliable, perceptible, and comprehensive of the three elements.

Despite the convenience of interpreting body language, some aspects of facial emotion recognition remain challenging, as an individual may intentionally or inadvertently attempt to disguise his or her emotions with misleading expressions. Interpreting another's intentions based on visible facial expressions alone may be detrimental and lead to a conflictual relationship. In these cases, to reveal an individual's genuine emotional state, micro-expressions should be considered. Unlike their "macro" counterparts, micro-expressions refer to rapid, involuntary muscular movements. Since they are difficult to prevent or manipulate, micro-expressions are particularly expressive in discerning repressed emotions. While micro-expressions are extremely difficult for the naked eye to discern, developments in computer vision techniques have reported promising advances (Xia, et al., 2020).

Recently, convolutional neural networks (CNNs) have been implemented to extract and interpret both overt and elusive facial expressions. This use of deep learning for image classification problems has yielded efficient and accurate detection of facial emotions, though some aspects of conventional frameworks can be optimized (Minaee & Abdolrashidi, 2019). This paper discusses a deep learning approach to efficient emotion detection. Through a joint analysis of images from four facial expression databases–FER-2013, FERG, CASME II, and

SAMM–the extent to which macro- and micro-expression datasets can be cross applied to the same network is explored. This paper also presents an analysis of how the orientation or environment of different headshots affects emotion recognition and, in the process, will demonstrate that introducing complexity to a CNN is not necessary to achieve a high accuracy rate. Finally, potential applications of facial emotion recognition in the field of human-machine interaction (HMI) will be discussed.

II. Background

i. Emotion Classification

Emotion recognition (ER) refers to the process of attributing human emotions to visual and auditory cues. Along with tone of voice and body language, facial expressions are an integral part of non-verbal communication and transcend language barriers and cultural diversity, allowing an individual to convey an emotional state immediately in a universal manner. Dr. Paul Ekman, a prominent 20th century American psychologist, identified six basic universal emotions: joy, anger, fear, surprise, sadness, and disgust (Ekman & Friesen, 1971). Neutrality was also included in later studies, leading to a total of seven universally accepted emotions for facial emotion recognition (FER). Following this revelation, Ekman developed the facial action coding system (FACS), which associated emotions with their corresponding facial muscle movements, known as Action Units (AUs) (Ekman & Friesen, 1978). This technique revolved around identifying relevant regions of interest including important facial components such as the nose, eyes, and mouth.

The most fundamental emotions can be identified by their macro-expressions, which are voluntary facial expressions and are displayed for 0.75s-2s. Macro-expressions are characterized by high intensity movements, and typically, the process of associating them with their corresponding emotion is not challenging. Moreover, macro-expressions are a result of intentional muscle control and can be controlled with ease depending on the sentiment an individual decides to convey. Not all cues are straightforward, however, as an individual might intentionally or inadvertently conceal his or her emotions, as stated by the emotion regulation theory. When such an attempt occurs, micro-expressions, which are subtle facial expressions lasting a duration of 0.04s to 0.2s, are much more reliable than macro-expressions to decipher emotion. Consideration of micro-expressions is particularly relevant in criminal investigation and lie detection, as individuals in these scenarios are inclined to conceal ulterior motives. These movements are challenging for the naked eye to detect; even with the assistance of Ekman's Micro-Expression Training Tool (METT), a program that helps humans learn to detect and classify subtle muscular movements, accuracy of detection by experts is below 50% (Takalkar et al., 2018). As humans are not yet capable of consistently discerning micro-expressions, current efforts are being directed toward developing alternatives using computer vision techniques.

ii. Computer Vision and Facial Feature Detection

Computer vision is a field of artificial intelligence concerning the methods by which computers process and comprehend digital images and videos involving detecting a visual stimulus, interpreting this signal, and producing an output. Aside from facial recognition, computer vision techniques can be implemented in autonomous vehicles and medical diagnostics. With regards to emotion recognition, computer vision techniques can be divided into two main categories, handcrafted features and deep learning, though in recent years, more attention has been directed toward the latter (Li & Deng, 2020). Traditional approaches to emotional recognition are based on a two-step machine learning approach. In the first step, certain features are extracted from the images, while in the second step, a classifier identifies the emotion being expressed. Some techniques for categorization of facial features include local binary patterns (Zhao & Zhang, 2011) and Gabor wavelets (Bartlett et al., 2005), and the histogram of oriented gradients (Chen et al., 2014). While these methods are effective on relatively simple datasets, they are less practical on more challenging datasets in which, for example, some images contain faces that are partially obscured by accessories. This paper will discuss the usage of a Haar cascade classifier which was first introduced in Paul Viola and Michael Jones's paper "Rapid Object Detection using a Boosted Cascade of Simple Features" (Viola & Jones, 2001). For the purposes of this research paper, Haar cascades are useful for capturing images from the video camera.

III. Related Works

i. Emotion Recognition Models and Approaches

Recent studies have explored the application of computer vision techniques in detecting minute fluctuations to reveal genuine emotions, though not without encountering several obstacles. First, Ekman's FACS is not as viable, given that micro-expressions are associated with fewer AUs due to their elusiveness. Also, micro-expression datasets are classified as either spontaneous or posed, though only the former is helpful in recognizing genuine concealed emotions since the latter is contrived and does not contradict the subject's emotional state. Due to the difficulty of capturing spontaneous micro-expression datasets and the novelty of the topic, there is a lack of these data samples. Finally, since micro-expressions are extremely rapid, being able to detect their exact time of occurrence in longer video sequences is a difficult task. While micro-expression recognition has been achieved through the local binary pattern-three orthogonal planes (LBP-TOP) and optical flow (OF) features, deep learning-based methods involving convolutional neural networks (CNNs) have been gaining more traction in recent years (Pan et al., 2021).

The first facial emotion recognition model was developed by Matsugu et al. of the Canon Research Center using a CNN and focused on distinguishing between expressions of happiness, neutrality, and talking (Matsugu et al., 2003). When implemented, the framework managed to achieve an accuracy rate of 97.6% for a dataset consisting of 5600 still images of 10 subjects. Other notable works include that of Khorrami et al. of the Beckman Institute for Advanced Science and Technology; this study focused on the extended Cohn-Kanade database (CK+) and the Toronto Face Dataset (TFD). While previous approaches were limited due to their use of hand-crafted features, Khorrami et al. explored the implementation of a deep learning model consisting of three convolutional layers, each followed by Rectified Linear Unit (ReLU) activation functions. All convolutional layers were followed by a fully-connected layer and a Softmax layer, and the model achieved state-of-the-art results (Khorrami et al., 2015).

Another approach for classifying facial emotion is to identify salient regions, which refer to areas that contain content that is meaningful for a certain scenario. For facial landmark detection, masking the images outside of the salient region can optimize and simplify the deep learning model. Minaee and Abdolrashidi of the University of California, Riverside proposed a deep learning-based framework for recognition of the six cardinal emotions and neutrality that employed the concept of salient regions; relevant areas of the face that were relevant in identifying macro-expressions over multiple datasets were isolated. The study suggests that implementing a convolutional neural network with fewer than 10 layers still yields high accuracies (Minaee & Abdolrashidi, 2019). While this study only accounts for macro-expressions, it does offer a more optimized framework by only scanning the facial regions that are helpful in identifying the emotion, proving its efficacy across four unique datasets: FER-2013, CK+, FERG, and JAFFE.

As stated previously, this study will evaluate two micro-expression datasets and two macro-expression datasets for a total of four facial emotion recognition databases. Although the network discussed does not involve tracking of facial units and is thus more suitable for macro-expression detection, the approach has the potential to serve as the basis for future deep learning implementations for micro-expression classification. This concept was demonstrated by Wang et al. as being a novel mechanism for micro-expression identification known as micro-attention, in which the CNN focuses on regions of the face with certain action units. For instance, the emotion of sadness is associated with movement of the 'inner brow raiser,' the 'brow lower,' and the 'lip corner depressor.' This method is pre-trained on four macro-expression databases before being further trained on three well-established micro-expression databases: CASME II, SAMM, and SMIC (Wang et al., 2019). Successful research on detecting micro-expressions evolves from the fundamental interpretation of macro-expressions.

ii. Emotion Recognition in Human Machine Interaction

Some studies also considered applications of their proposed neural network in human machine interaction (HMI) systems. For example, Mollahosseini et al. from the University of Denver explored automated Facial Expression Recognition (FER), conducting cross-database classification on a deep neural network consisting of two convolution layers, four sub-networks, and one max pooling layer (Mollahosseini et al., 2016). To achieve meaningful interaction with human beings, machines should be programmed to emulate the social and emotional capabilities present in human face-to-face communication.

Advances in artificial intelligence have been making affective computing more achievable in machines. Real-world applications involving facial expression detection include cybersecurity, psychotherapy, and criminology (Takalkar et al., 2018). Affective computing also provides machines with the emotional tools to carry out socially intelligent communication, or human-machine interaction (HMI). With respect to the subfield of human-robot interaction (HRI), robots are trained to possess and exhibit human-like characteristics on three levels: first, in their emotional state; second, in their outward expression; and third, in their ability to infer human emotional state. This third aspect utilizes facial emotion recognition techniques alongside analysis of thermal changes in facial images, body language and kinematics, brain activity, voice, and peripheral physiological responses. The ability to interpret a human user's expressions is crucial in gauging the situation and responding in an adequate manner (Spezialetti et al., 2020).

IV. Experimental Databases, Model Architecture, & Results

This section first provides an overview of the databases used in testing the proposed neural network, including their contents and image descriptions. Next, the architecture of the network used in this paper will be illustrated and discussed. Finally, the experimental results from applying different images from several databases will be presented and summarized.

A. *Databases*

Like the shift of facial emotion detection algorithms from handcrafted to shallow and deep learning, datasets evolved from being small and lab-controlled to being larger scale and in-the-wild. While macro-expression databases typically consist of still images of an exaggerated emotional expression, micro-expression databases are often the often; they tend to consist of short video clips depicting an extremely subtle micro-expression (Xia et al., 2020). Several publicly available datasets were used in this work, and they are described in detail and summarized in Table 1.

Table 1

Descriptions of Micro- and Macro-Expression Databases

Database	Expression	Description	Figure	Reference
FER2013	Macro-	Grayscale images, multiple ethnicities; anger, disgust, fear, happiness, neutrality, sadness, and surprise	Fig. 1	(Zahara et al., 2020)
FERG	Macro-	Stylized characters; anger, disgust, fear, joy, neutrality, sadness, and surprise	Fig. 2	(Aneja et al., 2017)
CASME	Micro-	Chinese ethnicity, colored images; happiness, disgust, surprise, repression, and others	Fig. 3	(Yan et al., 2014)
SAMM	Micro-	Multiple ethnicities, grayscale images; happiness, surprise, anger, disgust, sadness, contempt, and others.	Fig. 4	(Davison et al., 2018)

FER-2013: The Facial Expression Recognition 2013 (FER-2013) database contains 35,887 grayscale images, each of 48 x 48 resolution and categorized as one of seven emotions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. There is variation among the images, including different accessories, partial faces, and face occlusion (Zahara et al., 2020). Fig. 1 depicts several example images from the FER dataset.

Fig. 1: *Sample images from the FER2013 database. From left to right, the following cardinal emotions are displayed: anger, disgust, fear, happiness, neutrality, sadness, and surprise* (Zahara et al., 2020).

FERG: The Facial Expression Research Group 2D (FERG) database contains 55,767 images of six stylized characters with labeled facial expressions: anger, disgust, fear, joy, neutrality, sadness, and surprise. The images were produced using Autodesk Maya software and rendered out in 2D (Aneja et al., 2017). Fig. 2 depicts several example images from the FERG dataset.

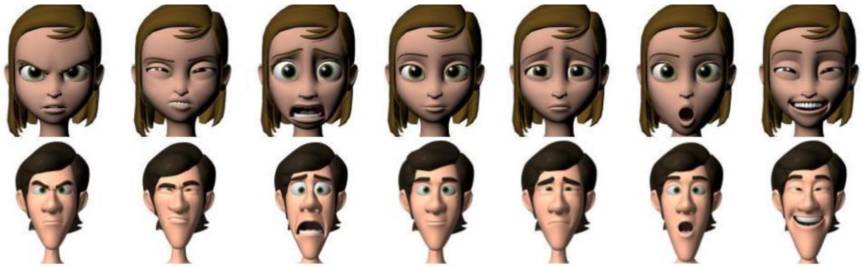

Fig. 2: *Sample images of subjects Aia and Malcolm of the FERG database. From left to right, the following cardinal emotions are displayed: anger, disgust, fear, joy, neutrality, sadness, and surprise* (Aneja et al., 2017).

CASME II: The Chinese Academy of Sciences Micro-Expressions II (CASME II) dataset contains 247 video sequences, recorded using a 200-fps camera with a resolution of 640 x 480. The classes of micro-expression included are as follows: happiness, disgust, surprise, repression, and others. A series of three different CASME databases, CASME II follows CASME and precedes CAS(ME)^2 (Yan et al., 2014). Fig. 3 depicts several example images from the CASME II dataset.

Fig. 3: *Sample images from the CASME II dataset of an expression of happiness* (Yan et al., 2014).

SAMM: The Spontaneous Micro-Facial Movement (SAMM) dataset contains 159 high-speed video sequences of 2040 x 1088 resolution recorded using a 200-fps camera. The emotion category distribution is as follows: Happiness, surprise, anger, disgust, sadness, contempt, and others (Davison et al., 2018). Fig. 4 depicts several example images from the SAMM dataset.

Fig. 4: *Sample images from the SAMM dataset of an expression of anger* (Davison et al., 2018).

B. *Model Architecture and Approach*

The model proposed is a convolutional neural network for image classification, based on existing model architecture developed at Clairvoyant, a data analytics consulting and engineering company. In constructing the model, several following parameters are considered. First, different activation functions are necessary to define the outputs of the units in the network according to their inputs. In this particular network, Rectified Linear Unit (ReLu) and Softmax activation functions are used (Chowdhry, 2021). Another feature is maximum pooling, which is implemented to down sample the dimensionality of each map. In addition, dropout helps prevent a network from overfitting by “dropping out” or omitting certain units to thin the network. Finally, the dense layer, also called a fully connected layer, is a layer in which every neuronal unit is connected to a unit of the preceding layer (Chowdhry, 2021). This facial emotion network in this paper is composed of the parameters discussed and is illustrated in Fig. 5.

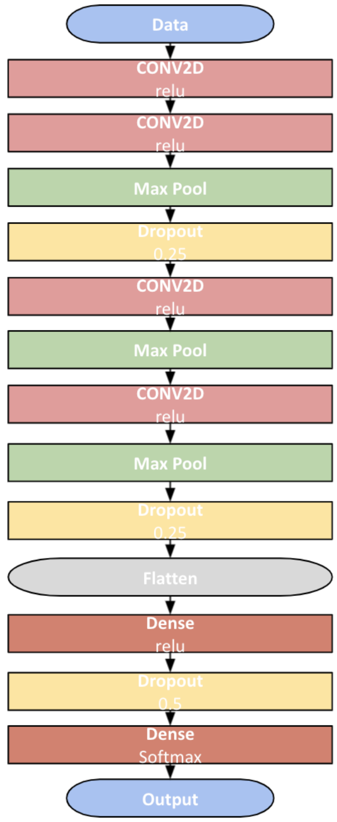

Fig. 5: *Diagram of network, consisting of ReLu and Softmax activation functions, dropouts, maximum pooling, and two dense layers.*

The network described is a sequential model of the Keras Application Programming Interface (API). In total, the network has four convolutional layers, three dropout layers, three maximum pooling layers, a step reserved for flattening to convert the data into a 1-dimensional array, and two dense layers. Next, a brief description of the general logical outline of the program will be provided, as follows:

1. First, the dataset is loaded into the program as a compressed file, from which the directories are extracted, though the dataset may require rearrangement. For instance, the FER-2013 dataset is already divided into its respective train and test data, while the FERG dataset is categorized by the animated characters.
2. After the appropriate APIs are imported, the train and test generators are initialized. The Keras ImageDataGenerator augments the images in the dataset in real-time and outputs

the number of images in the train and test sets. In this network, 28,709 images are in the train set and 7,178 images are in the test set.

3. Next, the CNN itself is built, as illustrated previously in Fig. 5. The dense layer consists of 1000 neurons, and the dropout rates are 0.3, 0.3, and 0.5. Then, the model is then compiled, trained, and saved. The epochs used in this model is 40 and implements Adam optimizer, an algorithm for optimization technique for gradient descent.
4. From here, there are two options to input an image: video camera or static image. A graphical representation of the predicted emotion is displayed as a bar graph. Along the x-axis, the different categories for emotion classification are listed, while the y-axis depicts the total percentage of predictions. The original input image is also outputted along with the bar graph.

C. Presentation of Findings

When implementing the FER-2013 database, this model achieved improvement over a validation set accuracy of 63%, improving from 60% from the first run-through by reducing overfitting. Several examples of test images and their corresponding prediction percentages are depicted in Fig. 6. In addition to running the network on the FER-2013 train and test data, the results of manually inputting data from the FERG macro-expression database were also recorded, as illustrated in Fig. 7.

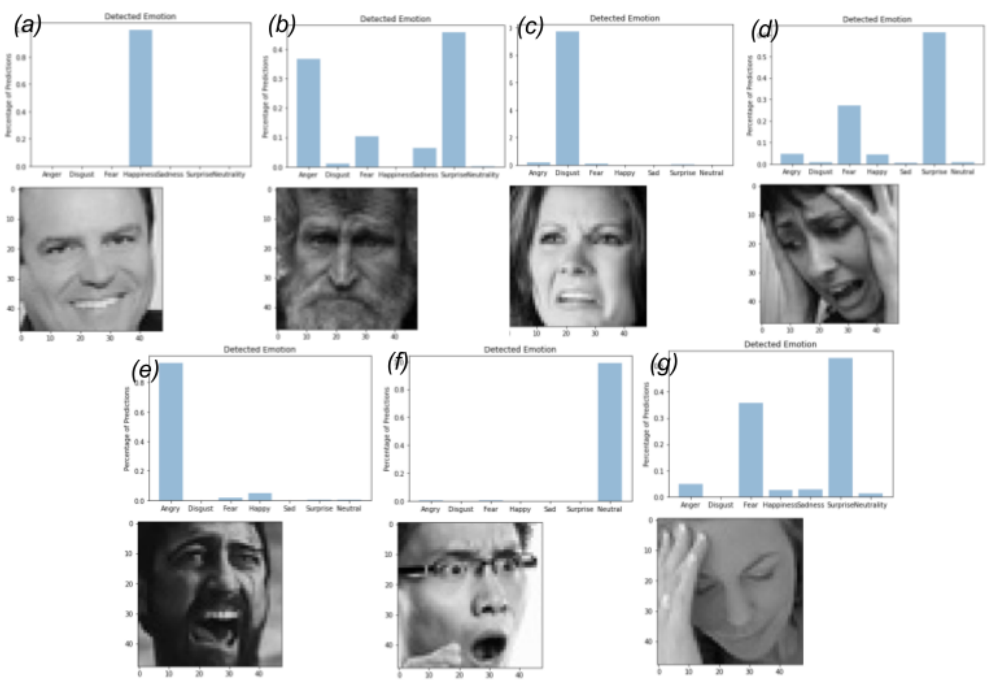

Fig. 6: *This figure depicts seven example images of the FER-2013 dataset for each classified emotion. Along the x-axis, the emotions are listed in the following order: anger, disgust, fear, happiness, sadness, surprise, and neutrality, and the y-axis indicates percentage. The top predicted/actual emotions being expressed are as follows: (a) Happiness/Happiness, (b) Surprise/Sadness, (c) Disgust/Disgust, (d) Surprise/Fear, (e) Anger/Anger, (f) Neutrality/Surprise, and (g) Surprise/Neutrality.*

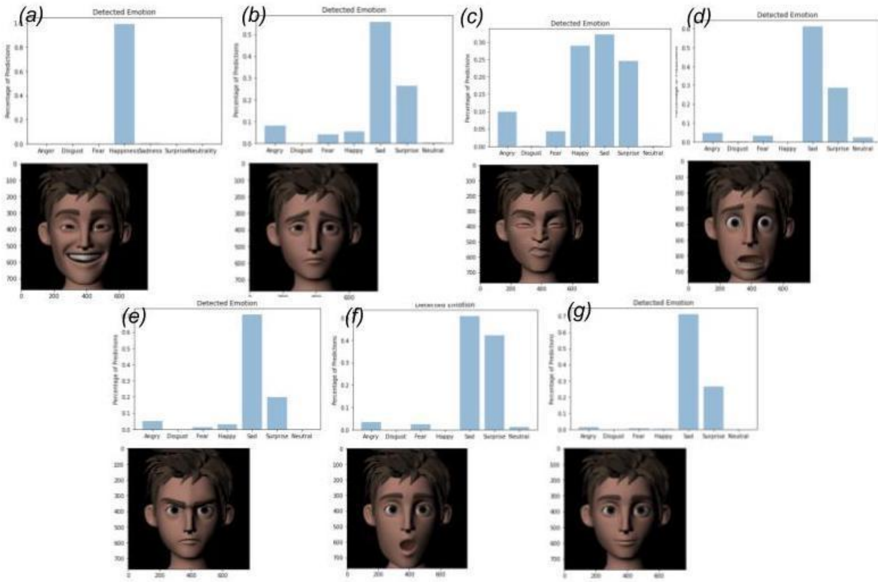

***Fig.** 7: This figure depicts seven example images of the FERG dataset for each classified emotion. Along the x-axis, the emotions are listed in the following order: anger, disgust, fear, happiness, sadness, surprise, and neutrality, and the y-axis indicates percentage. The top predicted/actual emotions being expressed are as follows: (a) Happiness/Happiness, (b) Sadness/Sadness, (c) Sadness/Disgust, (d) Sadness/Fear, (e) Sadness/Anger, (f) Sadness/Surprise, and (g) Sadness/Neutrality.*

In addition to applying an animated macro-expression dataset to evaluate the versatility and accuracy of the network, two examples of micro-expression datasets from CASME II and SAMM were also implemented. Fig. 8 depicts the graphically represented results for CASME II and SAMM, respectively. In addition, all reported findings from the experimental section are summarized in Table 2.

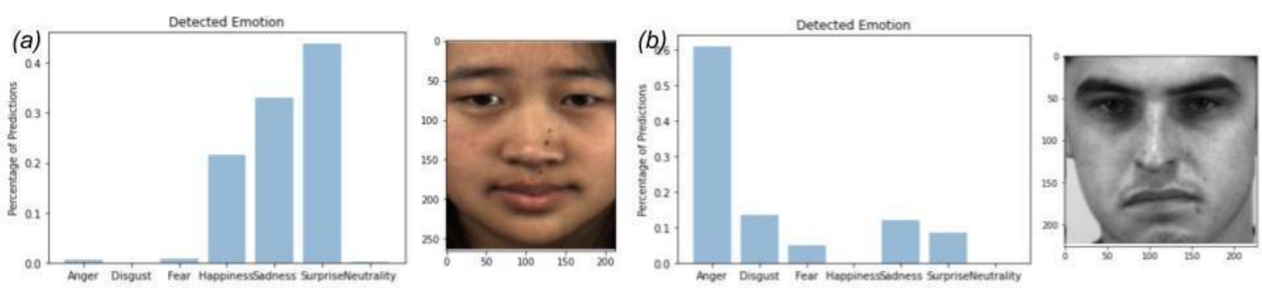

***Fig. 8:** Two example frames from the (a) CASME II and (b) SAMM datasets. While the data inputs themselves are in video form, two frames were inputted into the network. Along the x-axis, the emotions are listed in the following order: anger, disgust, fear, happiness, sadness, surprise, and neutrality, and the y-axis indicates percentage. The top predicted/actual emotions being expressed are as follows: (a) Surprise/Happiness and (b) Anger/Anger.*

Table 2

Summary of Emotion Classification Predictions of Example Images from Databases

Database	Expression	Actual	Predicted (Ranked highest to lowest %)
FER2013	Macro-	**(a) Happiness**	**Happiness**
		(b) Sadness	Surprise, Anger, Fear, Sadness, Disgust
		(c) Disgust	**Disgust**, Anger, Fear, Surprise
		(d) Fear	Surprise, *Fear*, Anger, Happiness, Disgust, Neutrality, Sadness
		(e) Anger	**Anger**, Happiness, Fear, Surprise, Neutrality
		(f) Surprise	Neutrality, Fear
		(g) Neutrality	Surprise, Fear, Anger, Sadness, Happiness, Neutrality
FERG	Macro-	**(a) Happiness**	**Happiness**, Sadness
		(b) Sadness	**Sadness**, Surprise, Anger, Happiness, Fear, Neutrality
		(c) Disgust	Sadness, Happiness, Surprise, Anger, Fear, Neutrality, Disgust
		(d) Fear	Sadness, Surprise, Anger, Fear, Neutrality
		(e) Anger	Sadness, Surprise, Anger, Happiness, Fear
		(f) Surprise	Sadness, *Surprise*, Anger, Neutrality, Fear
		(g) Neutrality	Sadness, Surprise, Anger, Fear
CASME	Micro-	Happiness	Surprise, Sadness, Happiness, Fear, Anger, Neutrality
SAMM	Micro-	**Anger**	**Anger**, Disgust, Sadness, Surprise, Fear

V. Analysis of Findings

This section reviews the graphically represented data presented in the previous section, including discussion of common trends in the data and observations of the various conditions of the datasets. Moreover, this section will also provide insight into the progress of human-machine interaction in achieving human-like communication based on the findings.

When evaluating the FER-2013 dataset, as depicted in Fig. 6, the predictions of happiness, disgust, and anger corresponded with their actual emotional expressions in cases *(a), (c),* and *(e)*. In addition, though fear was not the top prediction for case *(d)*, it the second highest predicted emotion behind surprise. While the selected samples feature individuals differing in ethnicity, gender, and poses, some observations and speculations about the conditions of the datasets can be made that correlate with the data presented. For instance, a potential confounding factor for case *(f)* may be the glasses being a wearable accessory causing a degree of facial occlusion. Similarly, in case *(g)*, the woman's face is partially obscuring the side of her face, and her neck is tilted slightly downward. While for case (*f*), surprise is not a predicted emotion in the first place, for case (*g)*, neutrality is the least predicted emotion. Ideally, all images would be as uniform as possible, such as limited to forward-facing visages, but in the context of human-machine interaction in everyday life, the human user is extremely unlikely to be still or conveniently located. Therefore, frequent gestures and movements that potentially obscure the face for a period of time should be considered when developing deep learning approaches in HMI. Another possible conflicting factor is the subjectivity of different expressions based on an individual's understanding of an emotion. For instance, while the individuals in cases *(c), (d), (e),* and *(f)* all share similar agape expressions, each is intended to reflect an entirely different emotion. In this case, the facial action units approach proposed by Ekman for locating micro-expressions may be misleading, as an opening of facial component representing the mouth may be easily misinterpreted.

Since the FERG database consists of entirely stylized figures, they possess a sense of uniformity that is not possible in a human facial database. For instance, the characters in the FERG database lack skin texture, variability in orientation, and other subjective factors. Regardless, "sadness" was selected as the top prediction for cases *(b)* through *(g)*, suggesting that the model is overfitted and unable to generalize to a larger range of emotions. A potential difficulty in cross-applying the FERG database is that the network can be trained based on the FER-2013 training set to recognize more subtle features such as wrinkles rather than predicting based on movement of the facial components alone. By oversimplifying the facial features, the stylistic characters obscure the finer details that may be needed to classify an expression. As established previously, different individuals' understanding of an emotion is subjective, and relying on obvious macro-expressions can be unreliable.

In addition to implementing the two macro-expression databases, two frames from the CASME II and SAMM datasets were applied. While a significantly more sophisticated algorithmic approach is required to detect subtle muscular movements in a video clip, inputting singular frames from micro-expressions into a generalized macro-expression detection framework is an important starting point. Theoretically, the emotion detected by the network should be a façade if the individual is being disingenuous and intentionally attempting to mask an underlying sentiment. Therefore, it would be important to simultaneously classify both macro-expressions and micro-expressions, compare the predictions, and draw a conclusion about the individual's mental state. This application could be valuable in a wide range of pursuits in the field of HMI, ranging from criminal justice to psychology.

VI. Future Work & Conclusions

This paper reviewed facial emotion recognition, including different methods of classifying emotions, the advantages of the deep learning approach, and potential cross-applications of the FER-2013, FERG, CASME II, and SAMM datasets. As a field that has been experiencing exponential growth during the past several decades, HMI will inevitably become more prevalent in everyday society. Emotion detection algorithms may help to ease this transition by allowing human-like interaction between humans and machine, though not without challenges. This paper highlighted some of the effects of manipulating certain aspects of input images with respect to facial occlusion, animation, color, and positioning, emulating the more unpredictable environment of the real-world and providing insight into the approach required to enable smooth interpersonal interactions.

For future consideration in situations in which a frontal view of an individual's face is not possible, interpreting one's body language may be a possible alternative. Detecting human emotions from facial expressions alone is challenging if an individual is positioned in an obscure area or is facing away from the camera. In these scenarios, scanning and interpreting gestures and posture can be helpful, though this area is less explored and definitive than that of facial emotion recognition, though an individual's posture and body language can also reveal deviations in an individual's state of mind (Santhoshkumar & Geetha, 2019). In combination with body language interpretation and classification of tone of voice, facial emotional recognition is a powerful field that has the potential to revolutionize the field of human-machine interaction.

VII. References

Aneja, D., Colburn, A., Faigin, G., Shapiro, L., & Mones, B. (2017). *Modeling Stylized Character Expressions via Deep Learning BT - Computer Vision – ACCV 2016* (S.-H. Lai, V. Lepetit, K. Nishino, & Y. Sato (eds.); pp. 136–153). Springer International Publishing.

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: machine learning and application to spontaneous behavior. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 568–573 vol. 2. https://doi.org/10.1109/CVPR.2005.297

Chen, J., Chen, Z., Chi, Z., & Fu, H. (2014). *Facial Expression Recognition Based on Facial Components Detection and HOG Features*.

Chowdhry, A. (2021). *Emotion Recognition With Deep Learning On Google Colab*. Clairvoyant. https://blog.clairvoyantsoft.com/emotion-recognition-with-deep-learning-on-google-colab-24ceb015e5

Clavel, C., Vasilescu, I., Devillers, L., Richard, G., & Ehrette, T. (2008). Fear-type emotion recognition for future audio-based surveillance systems. *Speech Communication*, *50*(6), 487–503. https://doi.org/https://doi.org/10.1016/j.specom.2008.03.012

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80. https://doi.org/10.1109/79.911197

Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2018). SAMM: A Spontaneous Micro-Facial Movement Dataset. *IEEE Transactions on Affective Computing*, *9*(1), 116–129. https://doi.org/10.1109/TAFFC.2016.2573832

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/http://dx.doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: a technique for the measurement of facial movement*.

Hantke, S., Cohrs, C., Schmitt, M., Tannert, B., Lütkebohmert, F., Detmers, M., Schelhowe, H., & Schuller, B. (2018). *Introducing an Emotion-Driven Assistance System for Cognitively Impaired Individuals BT - Computers Helping People with Special Needs* (K. Miesenberger & G. Kouroupetroglou (eds.); pp. 486–494). Springer International Publishing.

Khorrami, P., Paine, T. Le, & Huang, T. S. (2015). Do Deep Neural Networks Learn Facial Action Units When Doing Expression Recognition? *CoRR*, *abs/1510.0*. http://arxiv.org/abs/1510.02969

Li, S., & Deng, W. (2020). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 1. https://doi.org/10.1109/TAFFC.2020.2981446

Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks : The Official Journal of the International Neural Network Society*, *16 5-6*, 555–559.

Mehrabian, A. (1971). *Silent Messages*. Wadsworth Publishing Company.

Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2018). Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. *Sensors (Basel)*, *18*(2), 416. https://doi.org/10.3390/s18020416

Minaee, S., & Abdolrashidi, A. (2019). Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *CoRR*, *abs/1902.0*. http://arxiv.org/abs/1902.01019

Mollahosseini, A., Chan, D., & Mahoor, M. (2016). Going deeper in facial expression recognition using deep neural networks. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–10.

Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M., & Tao, J. (2021). Review of micro-expression spotting and recognition in video sequences. *Virtual Reality & Intelligent Hardware*, *3*(1), 1–17. https://doi.org/https://doi.org/10.1016/j.vrih.2020.10.003

Santhoshkumar, R., & Geetha, M. K. (2019). Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks. *Procedia Computer Science*, *152*, 158–165. https://doi.org/10.1016/j.procs.2019.05.038

Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. *Frontiers in Robotics and AI*, *7*, 145. https://doi.org/10.3389/frobt.2020.532279

Steppan, M., Zimmermann, R., Fürer, L., Schenk, N., & Schmeck, K. (2020). *Machine Learning Facial Emotion Recognition in Psychotherapy Research. A useful approach?* https://doi.org/10.31234/osf.io/wpa5e

Takalkar, M., Xu, M., Wu, Q., & Chaczko, Z. (2018). A survey: facial micro-expression recognition. *Multimedia Tools and Applications*, *77*(15), 19301–19325. https://doi.org/10.1007/s11042-017-5317-2

Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. In *IEEE Conf Comput Vis Pattern Recognit* (Vol. 1). https://doi.org/10.1109/CVPR.2001.990517

Wang, C., Peng, M., Bi, T., & Chen, T. (2019). *Micro-Attention for Micro-Expression Recognition*.

Xia, B., Wang, W., Wang, S., & Chen, E. (2020). Learning from Macro-Expression: A Micro-Expression Recognition Framework. *Proceedings of the 28th ACM International Conference on Multimedia*, 2936–2944. https://doi.org/10.1145/3394171.3413774

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–e86041. https://doi.org/10.1371/journal.pone.0086041

Zahara, L., Musa, P., Wibowo, E. P., Karim, I., & Musa, S. B. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–9. https://doi.org/10.1109/ICIC50835.2020.9288560

Zhao, X., & Zhang, S. (2011). Facial Expression Recognition Based on Local Binary Patterns and Kernel Discriminant Isomap. *Sensors (Basel, Switzerland)*, *11*, 9573–9588. https://doi.org/10.3390/s111009573