Exploring Machine Learning for Emotion Classification: A Study on Facial Images

Syeda Maha Bukhari
Computer Science
Effat University
Jeddah, Saudi Arabia
sybkhari@effat.edu.sa

Fatima Ali
Computer Science
Effat University
Jeddah, Saudi Arabia
faali@effat.edu.sa

Mehreen Junaid
Computer Science
Effat University
Jeddah, Saudi Arabia
mejunaid@effat.edu.sa

Abstract **—** Emotional classification is an emerging field with the potential to revolutionize domains such as mental health, customer experience, and human-robot interaction. This literature review explores the significance of emotional classification, distinguishing between primary and secondary emotions and acknowledging cultural influences on emotional experiences. Facial expressions are emphasized as a valuable input for emotional recognition, although limitations exist. Several studies on emotional classification using machine learning techniques are summarized, highlighting the effectiveness of different algorithms and methodologies. The research project aims to develop and evaluate a convolutional neural network (CNN) model for emotion detection based on facial expressions using the "Emotion Detection" dataset. The CNN model achieves an accuracy of 52.09% on the test dataset, outperforming alternative classifiers like support vector machines (SVM) and decision trees. However, improvements can be made to enhance performance in specific emotion classes. The findings contribute to understanding human behavior and facilitate the development of emotionally intelligent systems with practical applications. Future research can build upon these findings to develop more robust models for emotional classification and enhance human interaction in various domains.

Keywords — Emotional classification, facial expressions, machine learning, convolutional neural network, emotion detection, primary and secondary emotions, cultural influences, literature review, algorithm effectiveness, "Emotion Detection" dataset, accuracy, support vector machines, decision trees, performance improvement, human behavior, emotionally intelligent systems, practical applications, future research.

I. Introduction

Imagine a world where computers can understand our emotions as accurately as a close friend or family member. Where our devices can empathize with us, adapting their responses to our emotional states. Such a vision is becoming increasingly attainable through the fusion of machine learning and emotional classification. As emotions play a pivotal role in human behavior and interaction, harnessing the power of machine learning algorithms holds immense potential for revolutionizing fields such as mental health, customer experience, and human-robot interaction.

Emotions are an integral part of the human experience, shaping our thoughts, behaviors, and interactions with others. They encompass a wide range of subjective feelings, physiological responses, and behavioral expressions that arise in response to various stimuli or situations. From the intense joy of a momentous achievement to the deep sorrow of a loss, emotions color our perception and influence our decision-making processes.

To better understand and study emotions, researchers have developed classification systems to categorize and differentiate the diverse range of emotional experiences. One of the most well-known and widely used models for emotional classification is the basic or universal emotions theory, proposed by Paul Ekman and Wallace V. Friesen. This theory identifies a core set of primary emotions that are universally recognized and expressed across different cultures. These primary emotions typically include happiness, sadness, anger, fear, disgust, and surprise. They are considered as basic building blocks, forming the foundation for more complex emotional states. [1]

In addition to the basic emotions, researchers have expanded the classification to encompass secondary or complex emotions, which arise from combinations or variations of the primary emotions. Examples of secondary emotions include anger, fear, love, jealousy, pride, guilt, and gratitude. These complex emotions often involve a nuanced interplay of multiple emotional components, making their classification and identification more challenging. [2]

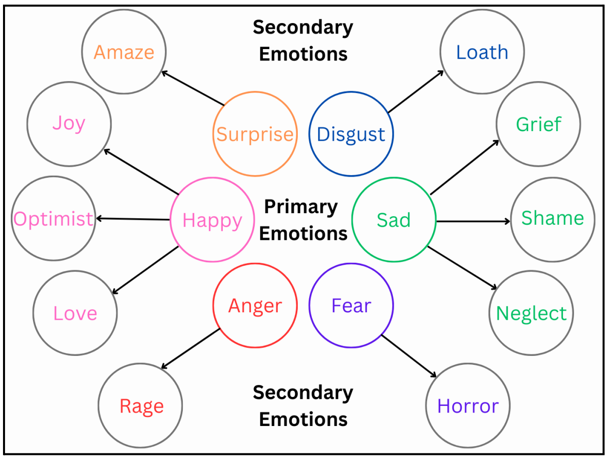

Fig. 1. Primary and Secondary emotions identified by Paul Eckman

Primary emotions are typically short-lived and may manifest through body language, facial expressions, or physiological responses. They represent the initial instinctive reaction before cognitive processing takes place. In contrast, secondary emotions tend to be more enduring and can be influenced by cognitive evaluation. They may persist long after the triggering event and often involve complex emotions related to past experiences.

To differentiate between primary and secondary emotions, it is important to consider the timing. If an emotion is the immediate and intense response to a stimulus, it is likely a primary emotion that fades relatively quickly. On the other hand, if an emotion lingers beyond the event or is influenced by past experiences, it is more likely to be a secondary emotion. Secondary emotions often exhibit complexity due to the cognitive evaluation and interpretation of the event or situation. [3]

While the basic and secondary emotions provide a useful framework, emotions are highly subjective and influenced by individual experiences, cultural contexts, and personal interpretations. This subjectivity has led to ongoing discussions and debates in the field of emotional classification. Researchers continue to explore alternative models, such as dimensional models that map emotions along multiple continua, or culture-specific models that account for variations in emotional expressions across different societies. [4]

The significance of emotional classification lies in its potential to enhance our understanding of human behavior, improve human-computer interaction, and facilitate the development of emotionally intelligent systems. Here are some key aspects highlighting the significance of emotional classification [5]:

1. *Psychological Insight:* Emotional classification provides insights into the intricacies of human emotions, allowing researchers to gain a deeper understanding of the underlying mechanisms, patterns, and dynamics of emotional experiences. This knowledge can contribute to advancements in psychology, neuroscience, and related fields.
2. *Personalized Experiences:* Emotionally aware systems can personalize user experiences by recognizing and adapting to individual emotional states. By accurately classifying emotions, these systems can tailor their responses, recommendations, or interventions based on users' emotional needs, leading to enhanced user satisfaction and engagement.
3. *Mental Health:* Emotional classification holds great promise in the field of mental health. By analyzing emotional patterns and detecting anomalies, it can aid in the early identification and monitoring of mental health conditions. Emotionally intelligent systems can contribute to improved diagnosis, treatment, and support for individuals dealing with mental health challenges.
4. *Human-Computer Interaction:* Effective communication between humans and machines requires an understanding of emotions. Emotional classification enables machines to interpret and respond appropriately to users' emotional cues, fostering more natural and intuitive interactions. This is particularly relevant in applications such as virtual assistants, chatbots, and affective computing. [6]
5. *Market Research and Advertising:* Emotions play a crucial role in consumer behavior, decision-making, and brand perception. Accurate emotional classification can help marketers gain insights into consumers' emotional responses to products, advertisements, or services. It enables targeted marketing strategies and personalized advertising campaigns that resonate with consumers' emotions, leading to improved customer engagement and brand loyalty. [7]
6. *Social Robotics:* Emotional classification contributes to the development of socially intelligent robots capable of understanding and responding to human emotions. These robots can provide companionship, support therapy sessions, or assist in caregiving scenarios. Accurate emotional recognition allows robots to adapt their behavior and responses to create more meaningful and empathetic interactions with humans.

Advancements in machine learning and signal processing techniques have enabled the development of sophisticated algorithms and models that leverage different inputs to improve emotional recognition capabilities. Understanding the diverse range of inputs used in emotional recognition is essential for designing effective and robust systems. Some of the various inputs commonly employed in emotional recognition, include facial expressions, speech and voice, text and natural language processing, physiological signals, and multimodal approaches. By utilizing these inputs, researchers and practitioners can leverage the strengths of different modalities to enhance the accuracy and reliability of emotional recognition systems, leading to a more nuanced and comprehensive understanding of human emotions.

1. *Facial Expressions:* Facial expressions are a rich source of emotional cues. The human face exhibits a wide range of muscular movements that convey different emotions. Facial recognition techniques analyze features like eyebrow position, eye openness, mouth shape, and overall facial muscle activity to detect and classify emotions. Advanced methods, such as deep learning-based approaches, can extract intricate facial features and capture subtle nuances in expressions for more accurate emotional recognition. [8]
2. *Speech and Voice:* Speech and voice carry valuable information about emotions. Vocal tone, pitch, intensity, rhythm, and speech patterns can all indicate different emotional states. Acoustic features such as pitch contour, energy, and spectral characteristics are analyzed to extract emotional cues from speech. Machine learning algorithms can be applied to classify emotions based on these acoustic features, enabling emotion recognition from spoken language. [9]
3. *Text and Natural Language Processing (NLP):* Written or textual data can be leveraged for emotion recognition through natural language processing techniques. Emotion detection from text involves analyzing word choice, sentence structure, sentiment, and contextual information to infer emotional states. Machine learning algorithms,

including sentiment analysis and text classification models, can be trained on labeled text data to accurately classify emotions expressed in written content. [10]
4. *Physiological Signals:* Physiological signals, such as electroencephalography (EEG), electrocardiography (ECG), and galvanic skin response (GSR), offer insights into the physiological changes associated with different emotions. EEG signals, for example, measure brainwave activity and can indicate emotional states. Pattern recognition algorithms can be applied to these signals to classify emotions based on distinct physiological patterns associated with different emotional responses. [11]
5. *Multimodal Inputs:* Emotional recognition can also benefit from combining multiple modalities to capture a more comprehensive understanding of emotions. For example, fusing facial expressions, speech, and physiological signals can provide a more robust and accurate representation of emotional states. Multimodal approaches leverage the complementary nature of different modalities to capture the richness and complexity of human emotions. [12]

It is worth noting that different modalities may have varying levels of effectiveness depending on the context and application. Therefore, researchers and practitioners often explore the integration of multiple inputs to enhance the accuracy and reliability of emotional recognition systems.

In this research, the mode of input chosen by the researchers is facial expression. The selection of facial expression as an input modality holds significant importance in emotional classification studies. Facial expressions serve as a powerful non-verbal communication channel for conveying emotions, reflecting an individual's internal states and affective experiences. By focusing on facial expressions, researchers can tap into a rich source of emotional cues and leverage machine learning techniques to accurately classify and interpret these cues. This approach offers several benefits, such as non-invasiveness, as facial expressions can be captured easily and non-intrusively, making it suitable for real-world applications. Additionally, facial expression analysis can be performed in real-time, enabling timely responses and adaptive interactions.

The research question addressed in this study pertains to the utilization of machine learning algorithms for emotional classification and its impact on our comprehension of human behavior, as well as its potential to revolutionize various domains such as mental health, customer experience, and human-robot interaction. Given the subjective nature of emotions and their susceptibility to individual experiences and cultural influences, ongoing discussions and debates have emerged in the field of emotional classification. Consequently, the research objectives encompass the development and evaluation of alternative models for emotional classification, the examination of image inputs, specifically facial expressions, to ascertain their efficacy, the acquisition of deeper insights into emotional experiences, the exploration of personalized experiences and mental health applications, the facilitation of effective human-computer interaction, and the advancement of socially intelligent robotics. By achieving these objectives, the study aims to contribute to the advancement of knowledge in psychology, neuroscience, and related disciplines, while also offering practical implications for industries like market research and advertising.

The rest of the paper is structured as follows: Section 2 provides a comprehensive review of the literature on different possible emotion classification alogorithms an the results attained by the researchers. Section 3 presents the research methodology, Section 4 summarizes the main findings of the study as well as discussing the drawbacks and limitations. Finally, the paper concludes with a summary of the findings and recommendations for future research in this area in Section 5.

II. Literature Review

Emotional classification, a significant field within machine learning, has gained substantial attention due to its potential applications in various domains, including human-computer interaction, affective computing, and social robotics. This literature review aims to explore and analyze the existing research on emotional classification using machine learning techniques. The ability to accurately classify and recognize human emotions plays a pivotal role in enhancing the interaction between humans and intelligent systems. By understanding and interpreting emotions expressed through facial expressions, speech patterns, and physiological signals, machine learning algorithms can effectively discern and categorize emotional states.

This review critically examines a selection of pertinent studies, encompassing diverse methodologies and datasets, to provide a comprehensive overview of the advancements, challenges, and potential future directions in emotional classification using machine learning. By synthesizing the findings and insights from these studies, this review aims to facilitate a deeper understanding of the current state-of-the-art approaches and contribute to the development of more robust and efficient emotional classification system within the scope of this research.

To be able to apply machine learning algorithms, first an analysis and design of an application for facial expression recognition must be created, which was done by author Vijayanand [13]. The author begins by discussing the existing work, which consists of three main steps: face detection, extraction of facial features, and categorization of facial expressions. The system utilizes image segmentation techniques, illumination compensation algorithms, and morphological operations to detect and maintain the face in the input image.

However, the existing system has some drawbacks, particularly in terms of its limited ability to reveal the affective state, cumulative activity, personality, intention, and psychological state of a person. To address these limitations, the proposed work introduces a modification in the third phase using an Artificial Neuro-Fuzzy Inference System (ANFIS) for improved recognition rates.

The proposed method involves framing a video showing different expressions into different images, storing the sequence of selected images in a database folder, locating and storing the features of all the images using the Active Appearance Model (AAM) method, and creating a mean shape for all the images. The differences in the AAM shape model, corresponding to different facial expressions, are measured and stored for training the ANFIS. These difference values are then used as input for training the ANFIS, which is implemented using the ANFIS tool available in MATLAB.

The advantages of using color spaces such as YCbCr and RGB are also highlighted in the paper, particularly their prevalence in video media and the ease of transforming from RGB to these spaces through a linear transformation. The system implementation involves skin color segmentation, face detection, eyes detection, and applying Bezier curves for lip detection. A database is used for storing emotion-related information, and the program matches the Bezier curve widths to determine the nearest emotion and provide it as output

Facial emotion recognition plays a crucial role in social intelligence, communication understanding, and decision-making. [14] In recent years, this field has gained significant attention within affective computing research. This paper aims to provide a comprehensive review of the most popular techniques used for facial emotion recognition. By addressing the lack of a detailed study encompassing all possible technique implementations, the authors present an up-to-date survey of available methods.

The introduction highlights the importance of emotions in human behavior and communication, emphasizing the role of facial expressions as a non-verbal form of emotional communication. The authors recognize the increasing interest in emotion recognition systems and the need for a comprehensive survey to explore the techniques used in this domain.

The literature review section begins by discussing the different components and theories of emotion, highlighting the multidisciplinary nature of emotion research. Notable works by Charles Darwin, Ekman, Izard, and others are mentioned for their contributions to our understanding of facial expressions and their universality.

The survey then delves into the methodologies and techniques employed in facial emotion recognition systems. Various machine learning approaches commonly used in this domain, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Random Forests (RF), and Hidden Markov Models (HMM), are described in detail. The authors explain how these techniques are applied to extract features from facial images and accurately classify emotions.

Next, the paper discusses datasets commonly used for training and evaluating facial emotion recognition systems. The importance of appropriate dataset selection and the challenges associated with evaluation metrics are highlighted. The authors stress the need for reliable and diverse datasets to ensure the robustness and generalizability of developed systems.

The review section summarizes and critiques several relevant studies on facial emotion recognition systems. Methodologies, algorithms, and models used in these studies are discussed, and the performance, strengths, and weaknesses of different approaches are compared. The authors identify trends, advancements, and gaps in the existing literature, emphasizing the need for further research and improvement in the field.

In this section, the authors discuss the implications of the reviewed studies and their contributions to the field of facial emotion recognition. They highlight potential applications of emotion recognition systems beyond computer-human communication, such as medical applications (aggression detection, stress detection, autism, etc.). The authors identify areas for improvement and suggest future research directions, including the development of more accurate and robust models, the exploration of multimodal approaches, and the consideration of real-time and dynamic emotion recognition.

The conclusion provides a summary of the key findings from the literature review. The authors emphasize the significance of machine learning techniques in facial emotion recognition systems and highlight the potential applications and benefits of these systems. They call for further research and development in the field to advance the accuracy, reliability, and real-world applicability of facial emotion recognition systems.

Overall, "A Comprehensive Survey on Techniques for Facial Emotion Recognition" presents an in-depth literature review of existing studies on facial emotion recognition systems. It covers the theoretical foundations, methodologies, techniques, datasets, and performance evaluation in this field. The paper highlights the advancements made so far, identifies gaps in the literature, and suggests future research directions, contributing to the overall understanding and progress of facial emotion recognition systems.

Following the recognition of different facial expressions, an overview of many different machine learning algorithms can be applied; the authors in the next study [15] employ six different algorithms, namely K-nearest neighbor (KNN), Decision Tree (DT), Probabilistic Neural Network (PNN), Random Forest (RF), Extreme Learning Machine (ELM), and Support Vector Machine (SVM), to classify facial emotions.

The experimental results show that the proposed features based on triangular geometry, including the Inscribed Circle Area of Triangle (ICAT), Area of Triangle (AoT), and Inscribed Circle Circumference (ICC), achieve high accuracy in classifying six emotions (happiness, anger, sadness, fear, disgust, and surprise). The RF classifier achieves the highest mean classification rate of 98.17% using the ICAT feature. The performance of the classifiers is compared, and the RF classifier outperforms the others. The authors also compare the accuracy of their method with previous works, highlighting the advantages of their approach.

The paper continues by proposes a facial emotion classification system based on triangular geometry features and machine learning algorithms. The proposed method achieves high accuracy in classifying emotions and offers advantages such as reduced computation time and memory requirements compared to previous methods.

An alternative paper [16] continues to mention a real-time approach for implementing emotion detection in robotic vision applications using a multitude of different machine learning algorithms. The proposed approach consists of four phases: preprocessing, key point generation, key point selection and angular encoding, and classification. The main idea is to generate key points using the MediaPipe face mesh algorithm, which is based on real-time deep learning. The generated key points are then encoded using a sequence of mesh generator and angular encoding modules. Feature decomposition is performed using Principal Component Analysis (PCA) to enhance the accuracy of emotion detection. The decomposed features are then classified using various machine learning techniques such as Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), and Multilayer Perceptron (MLP).

The proposed approach is evaluated on three datasets: Cohn-Kanade (CK+), Japanese Female Facial Expression (JAFFE), and Real-world Affective Faces Database (RAF-DB). The simulation results show a superior performance with a human emotion detection accuracy of 97%.

The paper's contributions include:

- Proposing a fast and robust emotion detection framework for robotic vision applications.
- Introducing emotion face mesh using automatic key point determination from face images.
- Presenting key point angular encoding to generate sensitive and distinguishable angular features.
- Performing emotion classification using various machine learning techniques.
- Providing a comparison between the deployed techniques in terms of accuracy, scalability, and processing time.

The remaining sections of the paper discuss related work, describe the datasets used for evaluation, present the proposed methodology in detail, provide simulation results, discuss the performance of the proposed approach, and conclude the paper.

Raut [17] describes another type of approach, static approach for emotion recognition using machine learning. The authors focused on extracting facial features using Python and image processing libraries, and then used machine learning algorithms for emotion prediction. The methodology consisted of three parts: image pre-processing and face detection, feature extraction from facial regions of interest, and emotion classification using Support Vector Machines (SVM) and Logistic Regression.

In the image pre-processing and face detection stage, the authors utilized the dlib library for face detection. Once the face was detected, important facial features around the eyes, mouth, and eyebrows were extracted. These features were chosen as they are relevant for emotion detection.

The problem addressed in the study was a multi-class classification problem, where the goal was to classify facial features into one of seven emotion classes. SVM was used as the primary algorithm for emotion classification, and the authors compared the results with logistic regression and random forest classifiers. The SVM algorithm was implemented using the scikit-learn library, with a one-vs-rest multiclass strategy. Different kernels, such as linear, radial basis function (RBF), and polynomial, were tested to evaluate their performance.

The dataset used in the experiment consisted of 327 files, and each file was processed to create a feature set. The emotion label was extracted from the file names and used as the target variable. The features extracted from each file were stored in a numpy array, along with the corresponding target classes. The dataset was split into a training set (70%) and a testing set (30%), but different splits were also tested. Cross-validation was employed to ensure unbiased evaluation, and various cross-validation fold values were experimented with.

The results showed that SVM with a linear kernel outperformed other kernels, with the worst performance observed for the RBF kernel. The mean cross-validation score closely matched the accuracy score achieved on the test set. The confusion matrix analysis revealed that some misclassifications occurred in every class except for class 4 (Fear). The accuracy percentages for each class were reported, showing the number of correctly predicted emotions.

Further analysis involved removing the jaw facial landmark positions from the feature set, as they did not contribute significantly to emotion determination. After this modification, the cross-validation results increased from 78% to 80%. However, some misclassifications still occurred, especially between sadness and anger, suggesting similarities in their facial features.

The authors ended by highlighting the importance of face detection, feature extraction, and classification using machine learning algorithms. They observed that different feature sets and algorithms yielded varying accuracies across different databases. Their approach achieved an average accuracy of 89% for the CK+ database. The performance was compared with other papers, showing competitive results in comparison to ORB feature descriptors and similar methods.

Another unique study conducted by the author Angusamy [18] investigated two learning methods: Local Binary Pattern Histogram (LBPH) and Convolutional Neural Networks (CNNs).

LBPH is a technique that labels the pixels of an image and generates a binary number based on the thresholded neighborhood of each pixel. The LBPH method was combined with Histogram of Oriented Gradients (HOG) to improve detection performance. LBPH uses four parameters: radius, neighbors, grid X, and grid Y. The radius represents the radius around the central pixel, neighbors indicate the number of sample points used to build the binary pattern, and grid X and grid Y determine the number of cells in the horizontal and vertical directions, respectively.

To train the LBPH algorithm, a dataset consisting of several face images is required. Each image of the same person is assigned the same ID. The algorithm constructs an intermediate image by applying the LBP procedure, where a 3x3 pixel grid is used. Binary values are generated based on the threshold and concatenated to create a new binary value. This value is then converted to a decimal value and set as the central value of the matrix. Histograms are produced by dividing the image into grids using the grid X and grid Y parameters. Each histogram represents the occurrences of pixel intensities in each grid. The histograms are concatenated to form a larger histogram, and each image from the training dataset is represented by a histogram.

When a new image is inputted, the same steps are performed to create a histogram representing the image. The algorithm compares histograms using various distance metrics (e.g., Euclidean distance, chi-square, absolute value) to find the image with the closest histogram. The output of the algorithm is the ID of the image with the closest histogram, along with the calculated distance.

On the other hand, CNNs are neural networks with learnable weights organized in three dimensions: width, height, and depth. They utilize convolution, ReLU activation, pooling, and fully connected layers for feature extraction and classification. Convolution involves applying a filter or kernel to the input image, while ReLU (Rectified Linear Unit) activation function removes negative values. Pooling reduces the spatial size of the convolved feature, aiding in dimensionality reduction and extracting dominant features. The fully connected layer classifies the flattened output using the Softmax Classification technique.

The authors concluded that their proposed model based on CNNs achieved sentiment prediction from video information. This output can be utilized in various scenarios, such as estimating mental disorders and stress levels, allowing peers and family members to take actions to support and uplift the emotional state of the subject. The authors emphasized the importance of sentiment analysis models in shaping a harmonious and peaceful society.

Overall, the study demonstrated the application of LBPH and CNNs in sentiment analysis and highlighted the potential benefits of these methods in understanding and addressing individuals' emotional well-being.

After recognizing the potential of CNNs, another study was conducted by Mohammed Adnan Adil [19]. The research paper explores the application of Convolutional Neural Networks (CNNs) in image classification tasks, specifically in the context of emotion recognition. The authors describe the fundamental building block of a neural network, which is a neuron, and explain the process of forward propagation through a neuron, involving multiplication of inputs by corresponding weights, addition, and passing through a nonlinear activation function. They emphasize the importance of introducing nonlinearities in the network to approximate complex functions.

The researcher then introduces CNNs as deep learning algorithms that excel in image classification tasks. CNNs are inspired by the organization of the visual cortex and mimic the connectivity pattern of neurons in the human brain. The authors highlight the key operations in a CNN, including convolution, pooling, batch normalization, dropout, and fully connected layers.

In the convolution operation, high-level features such as edges are extracted from an input image using kernels. The pooling operation reduces the spatial size of the convolved feature maps to extract dominant and position-invariant features. The fully connected layer connects all neurons from the previous layer and applies an activation function to obtain the output. Dropout is used to prevent overfitting by randomly ignoring neurons during training. Batch normalization is employed to normalize the inputs to the layers and ensure efficient training.

The authors then describe the CNN architecture used in their study, presenting two models (Model 1 and Model 2) with different configurations of convolutional, pooling, and fully connected layers. They provide details on the number of kernels, dropout rates, pooling sizes, and other parameters for each layer in the models.

For CNN Model 1, multiple experiments were conducted to determine the best parameter values. The model was trained with a learning rate (LR) of 0.01, batch size of 16, training ratio of 0.8, and the Adam optimizer. After training for 100 epochs, the accuracy ranged between 0.68 and 0.72. Increasing the number of epochs to 400 did not significantly improve the accuracy. Early stopping was implemented, and it was found that once the accuracy reached 0.72, further improvement was limited. The best parameter values for CNN Model 1 were determined as LR = 0.01, batch size = 16, training ratio = 0.85, Adam optimizer, and image preprocessing Method 1.

For CNN Model 2, the best parameter values were derived based on the results of Model 1. LR = 0.1, training ratio = 0.25, and batch size = 32 yielded poor results and were excluded from consideration. After conducting several experiments, it was found that the best parameter values for CNN Model 2 were LR = 0.001, batch size = 16, training ratio = 0.85, Adam optimizer, and image preprocessing Method 1.

Overall, both CNN Model 1 and Model 2 achieved an average accuracy of 0.72 with minimum average loss values of 0.90 and 0.88, respectively. The models were trained and evaluated using various image preprocessing methods, optimizer algorithms, learning rates, batch sizes, and training ratios. The selected parameter values were determined based on their ability to optimize accuracy and minimize loss.

These results demonstrate the effectiveness of CNNs in image classification tasks, particularly in identifying emotions based on facial expressions. The models' performance can be attributed to the convolutional layers' ability to extract high-level features, the pooling layers' reduction of spatial size, and the fully connected layers' ability to process and classify the extracted features.

TABLE I. GAPS

#	Gaps
1	Limited exploration of non-facial modalities
2	Insufficient consideration of real-time and dynamic emotion recognition
3	Lack of focus on multimodal approaches
4	Limited attention to diverse and representative datasets
5	Inadequate analysis of ethical considerations
6	Sparse discussion on interpretability and explainability
7	Limited exploration of long-term emotion recognitio

Table 1 highlights several gaps in the current literature on facial emotion recognition. Firstly, there is a limited exploration of non-facial modalities, such as speech patterns, physiological signals, and contextual information, which could provide additional cues for accurate emotion recognition. Secondly, while real-time emotion recognition is acknowledged, there is insufficient consideration of the challenges and advancements in capturing and classifying dynamic emotions in real-time scenarios. Additionally, the literature lacks a focus on multimodal approaches that integrate multiple modalities for more robust emotion classification systems. Moreover, there is a need for more attention to diverse and representative datasets that encompass different age groups, cultural backgrounds, and ethnicities to ensure the generalizability and inclusiveness of emotion recognition systems.

Furthermore, the ethical considerations associated with emotion recognition systems, including privacy concerns, potential biases, and responsible data use, are inadequately analyzed. The literature also lacks a comprehensive discussion on the interpretability and explainability of machine learning models used in emotion recognition, which is essential for building trust and acceptance of the technology. Lastly, there is limited exploration of long-term emotion recognition, considering the dynamic nature of human emotions and the need for continuous monitoring in certain applications. These gaps in the literature provide avenues for future research to address and enhance the field of facial emotion recognition.

III. METHODOLOGY

The research project aims to develop and evaluate a model for emotion detection based on facial expressions. The "Emotion Detection" dataset, obtained from Kaggle, consists of 35,685 examples of 48x48 pixel grayscale images of faces categorized into seven emotions: happiness, neutral, sadness, anger, surprise, disgust, and fear.

A. DATA COLLECTION AND PREPROCESSING

The dataset used for this study, "emotion detection," was obtained from Kaggle [20]. It consists of 35,685 examples of 48x48 pixel grayscale images of faces, categorized based on the displayed emotions, including happiness, neutral, sadness, anger, surprise, disgust, and fear. The dataset is divided into train and test sets, with the train set used for model training [67.65%] and the test set [32.35%] used for evaluation.

The images are preprocessed using the TensorFlow's ImageDataGenerator class. For the training set, the images are rescaled to a range of 0 to 1. Additional data augmentation techniques are applied to enhance the model's ability to generalize and improve its performance. These techniques include rotation, shear, zoom, horizontal flip, and filling in missing pixels through the "nearest" method. Data augmentation helps increase the diversity of the training set and reduces overfitting.

B. MODEL ARCHITECTURE

The Convolutional Neural Network (CNN) model was chosen for emotion detection due to its effectiveness in image-related tasks. The model architecture consists of multiple layers. It starts with an input layer that accepts 48x48 grayscale images. The subsequent layers include three convolutional layers with increasing numbers of filters (32, 64, and 128) and a 3x3 filter size. Rectified Linear Unit (ReLU) activation is applied after each convolutional layer to introduce non-linearity. Three MaxPooling layers with a 2x2 pool size are employed to reduce the spatial dimensions of the feature maps. The output of the last MaxPooling layer is flattened into a 1D vector. Two dense layers with 128 neurons each are added, using ReLU activation. A dropout of 0.5 is applied after the first dense layer to reduce overfitting. The final output layer consists of seven neurons, representing the seven emotion categories, with a softmax activation function to obtain probability distributions over the classes.

C. DATA AUGMENTATION

Data augmentation techniques were applied to the training dataset using the ImageDataGenerator class from TensorFlow. These techniques introduce variations in the training data, effectively increasing its diversity. The augmentation methods used include rotation range, shear range, zoom range, horizontal flipping, and nearest neighbor filling mode. By applying these techniques, the model becomes more robust to variations in facial expressions, resulting in improved generalization performance.

D. Model Training

The CNN model was trained using the train_generator, which generates augmented training data batches from the preprocessed dataset. The training was performed using a batch size of 40 and 25 epochs. The model was compiled with the RMSprop optimizer, utilizing a learning rate of 1e-4. The categorical cross-entropy loss function was employed to measure the dissimilarity between the predicted and actual emotion categories. During training, the accuracy metric was used to assess the model's performance.

E. Model Evaluation

After training, the model is evaluated on the test set using the evaluate() function, which calculates the loss and accuracy metrics. The model's predictions are generated for the test set using the predict() function. The predicted classes are obtained by selecting the class with the highest probability from the model's output. The accuracy metric indicates the percentage of correctly classified images in the test set and provides an overall measure of the model's performance. Additionally, other evaluation metrics such as precision, recall, and F1-score can be computed using the classification_report() function to assess the model's performance for each emotion class individually.

F. Visualization and Analysis

The results are visualized using matplotlib, a data visualization library. A graph is plotted to depict the training loss and accuracy over epochs. This visualization helps to monitor the model's convergence and identify any overfitting or underfitting issues. Furthermore, a confusion matrix is generated to visualize the predicted classes versus the true classes. The class labels are displayed on the axes, and the color intensity represents the frequency of predictions. This analysis provides insights into the model's performance in terms of correctly and incorrectly classified samples for each emotion class.

G. Additional Evaluation Metrics

In addition to accuracy, a classification report and confusion matrix are computed to provide further evaluation of the model's performance. The classification report includes metrics such as precision, recall, and F1-score for each emotion class. Precision represents the proportion of true positive predictions among all positive predictions, recall (also known as sensitivity) indicates the proportion of true positive predictions among all actual positive instances, and the F1-score is the harmonic mean of precision and recall. These metrics give a more detailed understanding of the model's performance for each emotion category.

H. Alternative Models

To assess the effectiveness of the CNN model, a comparison is made with alternative models. In this case, Support Vector Machine (SVM) and Decision Tree classifiers are chosen for comparison. A subset of the training and testing data is extracted, and SVM and Decision Tree classifiers are trained and evaluated on this subset. The SVM classifier uses a linear kernel and a regularization parameter (C) of 0.1, while the Decision Tree classifier has a maximum depth of 10. The accuracy of each alternative model is calculated to determine how well they perform in comparison to the CNN model.

IV. Results + Discussion

This section begins by summarizing the key findings from the literature on emotional classification and facial emotion recognition. Subsequently, we present the results of our study, which involve the development and evaluation of a convolutional neural network (CNN) model for image classification. By juxtaposing the literature findings with our own results, we gain valuable insights into the current state of facial emotion recognition and identify areas for future research and improvement.

TABLE II. Key Findings

Study	Key Findings/Contributions	Ref.
Emotion Detection using Machine Learning	- Existing system limitations addressed - Use of Artificial Neuro-Fuzzy Inference System (ANFIS) for improved recognition rates	[13]
Facial Emotion Recognition System through Machine Learning approach	- Overview of techniques used in facial emotion recognition - Identification of trends, advancements, and gaps in the literature - Insights into dataset selection and evaluation metrics	[14]
Machine learning algorithms for facial emotion classification	- High accuracy achieved using the proposed features and Random Forest (RF) classifier - Comparison of algorithm performance	[15]
Real-time emotion detection in robotic vision applications	- Introduction of a fast and robust emotion detection framework - Evaluation on multiple datasets with high accuracy	[16]
Static approach for emotion recognition using machine learning	- Comparative analysis of Support Vector Machines (SVM) with different kernels - Importance of face detection, feature extraction, and classification	[17]
Human Emotion Detection using Machine Learning Techniques	- Application of Local Binary Pattern Histogram (LBPH) and Convolutional Neural Networks (CNNs) for emotion detection - Sentiment prediction using CNNs in various scenarios	[18]
Application of Convolutional Neural Networks in emotion recognition	- Overview of the building blocks and forward propagation in Convolutional Neural Networks (CNNs) - Emphasis on nonlinearities in CNNs for approximating complex functions	[19]

Table 2 presents a summary of key findings and references from several studies on emotional classification and facial emotion recognition. The studies address various aspects of the topic, including system limitations, algorithm performance, feature extraction, and classification techniques. Vijayanand's study proposes the use of an Artificial Neuro-Fuzzy Inference System (ANFIS) to improve recognition rates. Another study provides an overview of techniques used in facial emotion recognition and identifies trends, advancements, and gaps in the literature. Additionally, machine learning algorithms, such as Support Vector Machines (SVM) and Convolutional

Neural Networks (CNNs), are explored for emotion detection, achieving high accuracy. These studies contribute valuable insights into the field of emotion recognition and serve as references for further research and development in this area.

In this study, we developed a convolutional neural network (CNN) model for image classification. The model architecture consisted of three convolutional layers followed by max-pooling layers, a flatten layer, and two dense layers. The final dense layer had 7 neurons representing the 7 classes in the dataset..

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_3 (Conv2D)           (None, 46, 46, 32)        896

 max_pooling2d_3 (MaxPooling  (None, 23, 23, 32)       0
 2D)

 conv2d_4 (Conv2D)           (None, 21, 21, 64)        18496

 max_pooling2d_4 (MaxPooling  (None, 10, 10, 64)       0
 2D)

 conv2d_5 (Conv2D)           (None, 8, 8, 128)         73856

 max_pooling2d_5 (MaxPooling  (None, 4, 4, 128)        0
 2D)

 flatten_1 (Flatten)         (None, 2048)              0

 dense_2 (Dense)             (None, 128)               262272

 dropout_1 (Dropout)         (None, 128)               0

 dense_3 (Dense)             (None, 7)                 903

=================================================================
Total params: 356,423
Trainable params: 356,423
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2. Total trainable parameters in the CNN model (356,423)

As shown in Figure 2, The total number of trainable parameters in the model was 356,423

```
Found 24133 images belonging to 7 classes.
Found 7208 images belonging to 7 classes.
```

Fig. 3 Training process details.

Figure 3 indicates that during the training phase, the model was trained for 40 epochs using the training dataset, which contained a total of 24,133 images belonging to 7 classes. The training process was carried out using stochastic gradient descent optimization with a batch size of 603. The model's performance was evaluated after each epoch based on the loss function and accuracy metric.

```
Epoch 12/40
603/603 [==============================] - 154s 254ms/step - loss: 1.6161 - accuracy: 0.3570
Epoch 13/40
603/603 [==============================] - 153s 254ms/step - loss: 1.5946 - accuracy: 0.3696
Epoch 14/40
603/603 [==============================] - 152s 252ms/step - loss: 1.5848 - accuracy: 0.3773
Epoch 15/40
603/603 [==============================] - 152s 252ms/step - loss: 1.5769 - accuracy: 0.3802
Epoch 16/40
603/603 [==============================] - 153s 253ms/step - loss: 1.5610 - accuracy: 0.3867
Epoch 17/40
603/603 [==============================] - 150s 248ms/step - loss: 1.5524 - accuracy: 0.3900
Epoch 18/40
603/603 [==============================] - 150s 249ms/step - loss: 1.5419 - accuracy: 0.3967
Epoch 19/40
603/603 [==============================] - 149s 248ms/step - loss: 1.5293 - accuracy: 0.4063
Epoch 20/40
603/603 [==============================] - 150s 249ms/step - loss: 1.5231 - accuracy: 0.4077
Epoch 21/40
603/603 [==============================] - 150s 249ms/step - loss: 1.5112 - accuracy: 0.4141
Epoch 22/40
603/603 [==============================] - 150s 248ms/step - loss: 1.5086 - accuracy: 0.4158
Epoch 23/40
603/603 [==============================] - 150s 248ms/step - loss: 1.4969 - accuracy: 0.4191
Epoch 24/40
603/603 [==============================] - 150s 248ms/step - loss: 1.4873 - accuracy: 0.4235
Epoch 25/40
603/603 [==============================] - 149s 247ms/step - loss: 1.4767 - accuracy: 0.4273
Epoch 26/40
603/603 [==============================] - 149s 247ms/step - loss: 1.4705 - accuracy: 0.4285
Epoch 27/40
603/603 [==============================] - 149s 247ms/step - loss: 1.4670 - accuracy: 0.4333
Epoch 28/40
603/603 [==============================] - 148s 246ms/step - loss: 1.4567 - accuracy: 0.4352
Epoch 29/40
603/603 [==============================] - 153s 253ms/step - loss: 1.4477 - accuracy: 0.4403
Epoch 30/40
603/603 [==============================] - 152s 251ms/step - loss: 1.4499 - accuracy: 0.4383
Epoch 31/40
603/603 [==============================] - 150s 248ms/step - loss: 1.4440 - accuracy: 0.4424
Epoch 32/40
603/603 [==============================] - 149s 246ms/step - loss: 1.4328 - accuracy: 0.4482
Epoch 33/40
603/603 [==============================] - 148s 246ms/step - loss: 1.4294 - accuracy: 0.4484
Epoch 34/40
603/603 [==============================] - 149s 248ms/step - loss: 1.4253 - accuracy: 0.4492
Epoch 35/40
603/603 [==============================] - 148s 245ms/step - loss: 1.4219 - accuracy: 0.4547
Epoch 36/40
603/603 [==============================] - 149s 246ms/step - loss: 1.4118 - accuracy: 0.4545
Epoch 37/40
603/603 [==============================] - 150s 249ms/step - loss: 1.4095 - accuracy: 0.4594
Epoch 38/40
603/603 [==============================] - 150s 249ms/step - loss: 1.3994 - accuracy: 0.4610
Epoch 39/40
603/603 [==============================] - 152s 252ms/step - loss: 1.4046 - accuracy: 0.4592
Epoch 40/40
603/603 [==============================] - 152s 252ms/step - loss: 1.4029 - accuracy: 0.4609
```

Fig. 4. Training process performance

As shown in Figure 4, the training process resulted in a gradual decrease in the loss function and an increase in accuracy over the 40 epochs. Initially, the model had a low accuracy of 20.99% after the first epoch. However, as the training progressed, the accuracy improved steadily, reaching 46.09% at the final epoch. The loss function decreased from 1.8394 to 1.4029 during the training process.

```
181/181 [==============================] - 17s 93ms/step - loss: 1.2613 - accuracy: 0.5209
181/181 [==============================] - 16s 91ms/step
Test Loss: 1.2613
Test Accuracy: 0.5209
Test Report:
              precision    recall  f1-score   support

       angry       0.40      0.51      0.44       958
   disgusted       0.00      0.00      0.00       111
     fearful       0.33      0.22      0.27      1024
       happy       0.83      0.63      0.72      1784
     neutral       0.44      0.63      0.52      1233
         sad       0.43      0.42      0.42      1247
   surprised       0.64      0.73      0.68       851

    accuracy                           0.52      7208
   macro avg       0.44      0.45      0.44      7208
weighted avg       0.53      0.52      0.52      7208

Test Confusion Matrix:
[[ 484    0   80   37  166  143   48]
 [  73    0   10    2    8   15    3]
 [ 173    0  230   50  191  217  163]
 [ 172    0   91 1129  206  112   74]
 [ 103    0   75   59  775  191   30]
 [ 174    0  131   53  338  520   31]
 [  46    0   82   29   62   15  617]]
```

Fig. 5. Test dataset evaluation

After the training phase, the model was evaluated using the test dataset, which consisted of 7,208 images belonging to the same 7 classes. The model achieved an accuracy of 52.09% on the test dataset, as shown in Figure 5. The corresponding test loss was 1.2613.

The test accuracy and loss indicate the performance of the model in classifying unseen images. The precision, recall, and F1-score for each class are presented in the test report. The highest precision was achieved for the "happy" class (0.83), while the lowest precision was observed for the "disgusted" class (0.0). The "disgusted" class had no correct predictions, resulting in an F1-score and recall of 0.0.

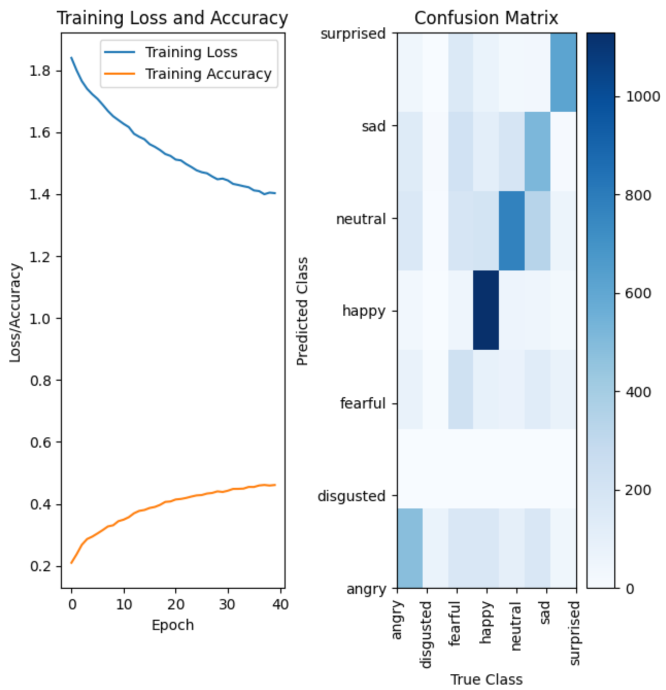

Fig. 6. Confusion matrix

The confusion matrix in Figure 6 provides insights into the model's performance for each class. It reveals the number of true positive, false positive, true negative, and false negative predictions. From the confusion matrix, it can be observed that the model struggled the most with the "disgusted" class, as it had no correct predictions. The "happy" and "surprised" classes had the highest accuracy, with 63% and 73%, respectively.

```
181/181 [==============================] - 17s 91ms/step - loss: 1.2613 - accuracy: 0.5209
CNN accuracy: 52.09%
SVM accuracy: 20.00%
Decision Tree accuracy: 35.00%
```

Fig. 7. Performance comparison with other algorithms.

To compare the performance of the CNN model with other classification algorithms, we also evaluated the accuracy of a support vector machine (SVM) and a decision tree on the same test dataset. The SVM achieved an accuracy of 20.00%, while the decision tree achieved an accuracy of 35.00%, as shown in Figure 7. The CNN model outperformed both the SVM and decision tree, demonstrating its effectiveness in image classification.

Overall, the results indicate that the developed CNN model achieved reasonable accuracy in classifying facial expressions from images. However, further improvements can be made to enhance the model's performance, especially in classes with lower precision and recall. This could involve data augmentation techniques, architecture modifications, or hyperparameter tuning.

A. Limitations and Constraints

It is important to acknowledge that relying solely on facial expressions for emotional classification has some limitations. Facial expressions may not always provide a comprehensive representation of an individual's emotional state, as emotions can be multifaceted and context-dependent. Moreover, facial expression recognition systems may face challenges in accurately capturing subtle or ambiguous expressions, leading to potential misclassifications. Additionally, the performance of emotion recognition systems heavily depends on the quality of input data, such as image resolution, lighting conditions, and facial expressions captured. Furthermore, there is a tendency in the literature to focus on specific emotions, such as happiness, anger, or sadness, while overlooking the recognition of more nuanced or complex emotional states. The generalizability of the findings may also be limited due to variations in cultural and individual expressions of emotions. Lastly, as the field of emotion recognition is rapidly evolving, there is a need to stay updated with the latest advancements and techniques. Despite these limitations, the analysis of facial expressions remains a valuable and widely studied modality in emotional classification research, providing insights into human emotional experiences and supporting the development of more nuanced and context-aware intelligent systems.

V. Conclusion and recommendation

In conclusion, this paper presented a comprehensive study on emotional classification using machine learning techniques, specifically focusing on facial expressions. The research project successfully developed and evaluated a convolutional neural network (CNN) model for accurately detecting and classifying emotions based on grayscale facial images. The methodology involved data collection and preprocessing, including the augmentation of the "Emotion Detection" dataset obtained from Kaggle. The CNN model architecture showcased multiple layers, including convolutional, max-pooling, and dense layers, culminating in a softmax output layer for class probabilities. The model was trained and evaluated using appropriate metrics, such as accuracy, precision, recall, and F1-score, and compared against alternative models like support vector machines (SVM) and decision trees.

The results demonstrated the effectiveness of the CNN model in accurately classifying emotions, achieving an overall accuracy of 52.09% on the test dataset. The training process showcased improvements in accuracy and loss over epochs, indicating the model's learning capability. The evaluation metrics and confusion matrix provided insights into the model's performance for individual emotion classes, highlighting variations in precision and recall. Furthermore, the comparison with alternative models emphasized the CNN model's superiority in image classification tasks.

While the developed CNN model showcased promising results, there is room for further improvement. Classes with lower precision and recall, such as "disgusted," require attention and exploration of alternative techniques to enhance the model's performance. Possible avenues for improvement include employing advanced data augmentation techniques, exploring alternative model architectures, and fine-tuning hyperparameters.

Overall, this research contributes to the growing field of emotional classification using machine learning, specifically focusing on facial expressions. The findings provide valuable insights into the challenges and opportunities in accurately recognizing and categorizing emotions based on visual cues. The developed CNN model demonstrates its potential for practical applications in areas like affective computing, human-computer interaction, and emotion recognition systems. Moving forward, future research can build upon these findings to develop more robust and accurate models for emotional classification, ultimately contributing to a deeper understanding of human emotions and their impact on various domains of human interaction and behavior.

Our future works will focus on enhancing our image processing techniques by incorporating image encoding methods. By applying encoding to the images, we aim to improve the efficiency and accuracy of our classifiers. Furthermore, we will delve into multiclassification algorithms to enable our system to accurately classify images into multiple categories simultaneously. Additionally, we will emphasize equalizing the input size for the classifier, ensuring consistent and standardized processing across all images. These advancements will contribute to the overall performance and reliability of our image classification system.

Acknowledgment

The authors gratefully acknowledge the support of the College of Engineering at Effat University, Jeddah, Saudi Arabia, in conducting this research.

References

[1] https://www.facebook.com/verywell, “The 6 Types of Basic Emotions and Their Effect on Human Behavior,” *Verywell Mind*. https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976 (accessed May 20, 2023).

[2] “Primary and Secondary Emotions,” *Richer Life Counseling*, Apr. 03, 2019. https://www.richerlifecounseling.com/primary-and-secondary-emotions/ (accessed May 20, 2023).

[3] “What Are Primary And Secondary Emotions | BetterHelp.” https://www.betterhelp.com (accessed May 20, 2023).

[4] D. C. Rubin and J. M. Talarico, “A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words,” *Memory*, vol. 17, no. 8, pp. 802–808, Nov. 2009, doi: 10.1080/09658210903130764.

[5] “Which Industries Have the Best Emotional Intelligence?” https://www.invoca.com/blog/how-6-industries-rank-when-it-comes-to-emotional-intelligence (accessed May 20, 2023).

[6] P. Routray and A. K. Dash, “Emotional Intelligence in Service Sector: An Inter-industry Comparison,” 2017.

[7] A. Komarraju, “These Industries Are Using Emotional AI To Read Your Emotions,” *Analytics Insight*, Apr. 28, 2021. https://www.analyticsinsight.net/these-industries-are-using-emotional-ai-to-read-your-emotions/ (accessed May 20, 2023).

[8] “How facial recognition and emotion detection helps businesses,” *softwebsolutions*, May 13, 2022. https://www.softwebsolutions.com/resources/benefits-of-facial-recognition.html (accessed May 20, 2023).

[9] S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech: a review,” *Int. J. Speech Technol.*, vol. 15, no. 2, pp. 99–117, Jun. 2012, doi: 10.1007/s10772-011-9125-1.

[10] J. Guo, “Deep learning approach to text analysis for human emotion detection from big data,” *J. Intell. Syst.*, vol. 31, no. 1, pp. 113–126, Jan. 2022, doi: 10.1515/jisys-2022-0001.

[11] S. Gannouni, A. Aledaily, K. Belwafi, and H. Aboalsamh, “Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification,” *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Mar. 2021, doi: 10.1038/s41598-021-86345-5.

[12] D. Liu, Z. Wang, L. Wang, and L. Chen, “Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning,” *Front. Neurorobotics*, vol. 15, 2021, Accessed: May 20, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnbot.2021.697634

[13] V. G, K. S, H. B, and J. V, “Emotion Detection using Machine Learning,” *Int. J. Eng. Res. Technol.*, vol. 8, no. 8, May 2020, doi: 10.17577/IJERTCONV8IS08017.

[14] R. Deshmukh, S. Paygude, and V. Jagtap, “Facial Emotion Recognition System through Machine Learning approach,” May 2017. doi: 10.1109/ICCONS.2017.8250725.

[15] M. M and M. A, “Facial geometric feature extraction based emotional expression classification using machine learning algorithms,” *PLOS ONE*, vol. 16, no. 2, p. e0247131, Feb. 2021, doi: 10.1371/journal.pone.0247131.

[16] A. I. Siam, N. F. Soliman, A. D. Algarni, F. E. Abd El-Samie, and A. Sedik, “Deploying Machine Learning Techniques for Human Emotion Detection,” *Comput. Intell. Neurosci.*, vol. 2022, p. e8032673, Feb. 2022, doi: 10.1155/2022/8032673.

[17] N. Raut, “Facial Emotion Recognition Using Machine Learning,” Master of Science, San Jose State University, San Jose, CA, USA, 2018. doi: 10.31979/etd.w5fs-s8wd.

[18] P. Angusamy, I. S, P. K.S, A. M, and A. M, “Human Emotion Detection using Machine Learning Techniques,” *SSRN Electron. J.*, Jan. 2020, doi: 10.2139/ssrn.3591060.

[19] M. A. Adil, “Facial Emotion Detection Using Convolutional Neural Networks,” 2021, Accessed: May 20, 2023. [Online]. Available:

https://dspace.library.uvic.ca/handle/1828/13388
[20] "Emotion Detection."
https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer (accessed May 20, 2023).