International Journal of Informatics, Information System and Computer Engineering

INJIISCOM

Face Emotion Recognition Based on Machine Learning: A Review

*Adnan Mohsin Abdulazeez *, Zainab Salih Ageed ***

* Technical Informatics., College of Akre, Duhok Polytechnic University, Duhok, Iraq
** College of Science, Computer Science Dept., Nawroz University, Duhok, Iraq
*Corresponding Email: Adnan.mohsin@dpu.krd.edu

ABSTRACTS

Computers can now detect, understand, and evaluate emotions thanks to recent developments in machine learning and information fusion. Researchers across various sectors are increasingly intrigued by emotion identification, utilizing facial expressions, words, body language, and posture as means of discerning an individual's emotions. Nevertheless, the effectiveness of the first three methods may be limited, as individuals can consciously or unconsciously suppress their true feelings. This article explores various feature extraction techniques, encompassing the development of machine learning classifiers like k-nearest neighbour, naive Bayesian, support vector machine, and random forest, in accordance with the established standard for emotion recognition. The paper has three primary objectives: firstly, to offer a comprehensive overview of effective computing by outlining essential theoretical concepts; secondly, to describe in detail the state-of-the-art in emotion recognition at the moment; and thirdly, to highlight important findings and conclusions from the literature, with an emphasis on important obstacles and possible future paths, especially in the creation of state-of-the-art machine learning algorithms for the identification of emotions.

ARTICLE INFO

Article History:
Received 01 Dec 2023
Revised 27 Dec 2023
Accepted 05 Jan 2024
Available online 11 Mar 2024
Publication Date 01 Jun 2024

Keywords:
Technology,
Information System,
Computer Science

1. INTRODUCTION

Despite their best efforts, humans cannot fully suppress emotions, as some researchers argue that emotions are inherent abilities. Emotion detection, an automated technique for determining an individual's affective state, is becoming more and more important in the field of human-computer interaction (HCI) for a variety of applications, such as automobile safety (Hudlicika & Broekens, 2009). Unfortunately, most modern HCI systems lack emotional intelligence, rendering them incapable of processing or understanding emotional data and making decisions based on such information (Newell & M. Marabelli, 2015). Typically, emotions are assessed by analyzing patterns of facial expressions, head movements, eyelid movements, or a combination of these factors. While the visual sense of facial emotions is valuable for emotion identification, it is not always sufficient (Vankalayapati et al., 2011). In advanced intelligent systems, addressing the disconnect between humans and machines is crucial. A system that cannot recognize human affective states is prone to inadequate responses to those states. Therefore, it is crucial to train machines in interpreting and understanding human emotional states (Amanoul et al., 2021). As a result, the development of a reliable, accurate, flexible, and resilient emotion identification system becomes imperative for successful implementation in intelligent Human-Computer Interaction (HCI). With the overarching goal of instilling machines with emotions, an increasing number of researchers in artificial intelligence (AI) have explored affective computing, particularly emotion recognition, establishing it as an emerging and promising area of study (Kratzwald et al. 2018). Numerous studies on emotion recognition in audio-visual formats have been conducted over the years. The literature generally exhibits three primary methods: visual-based, audio-visual, and audio-based approaches. Early research primarily concentrated on independently addressing auditory and visual data modalities. The basis of audio-based emotion detection endeavors involves extracting and identifying emotional states from human voice signals (Tawari & Trivedi 2011; Kashevnik et al., 2021).

An increasing cohort of experts in the fields of ergonomics and intelligent systems is focused on improving the efficiency and flexibility of Human-Computer Interaction. This dedication arises from the rising prevalence of collaboration between humans and machines in diverse contexts. In intelligent HCI systems, computers must exhibit adaptability to accurately comprehend human communication styles and deliver appropriate responses (Zhihan et al., 2022). Human intentions are conveyed through both verbal and nonverbal means, encompassing a spectrum of emotional expressions. The understanding of human emotions and behavior plays a pivotal role in the adaptability of computers, giving rise to a burgeoning field known as affective computing [Shantanu et al., 2022; shadeeq et al., 2023]. The selection of distinctive features for discerning various emotions is a critical consideration in this context. Two categories of features, namely prosodic and spectral features, have been identified as valuable for the identification of emotions in speech (Chen et al., 2005; Wu et al., 2014).

2. BACKGROUND THEORY

A person's intricate emotional state is shaped by the interplay of behavior, thoughts, and feelings, manifested through psychophysiological reactions to internal or external stimuli. The quest to quantify beauty has prompted numerous studies across diverse fields, including psychology, philosophy, biology, and the arts, with a particular focus on facial analysis and aesthetics (Saeed et al., 2022). Affective computing serves various purposes, notably contributing to intelligent and user-friendly Human-Computer Interaction (HCI). Precise real-time detection of the human operator's emotional state can enhance HCI systems significantly (Dave, 2023). In military and aerospace domains, it is feasible to promptly identify the high-risk functional status of soldiers, pilots, and astronauts in real-time. Beyond this, emotion recognition technology finds application in public transit, enhancing driving safety by monitoring a driver's real-time emotional state and preventing risky driving during periods of extreme emotional stress (Zhang et al., 2020). In eLearning systems, the primary emphasis is on single-user face detection, where facial expressions are used to represent the user's emotions, enabling appropriate adjustments to instructional tactics (Ashwin et al., 2020; Abdullah & Abdulazeez 2021). This interdisciplinary field draws from cognitive science, psychology, and computer science. Emotions wield substantial influence over human behavior, impacting processes like perception, attention, learning, and decision-making (Zhao et al., 2016). Within the training process of machine-learning systems, loss functions are fundamental components. The optimal parameter values for the system are obtained by minimizing the mean loss value across a labeled training set, as outlined in reference (Saeed et al., 2023). Affective computing applications span diverse domains, including automatic driving assistance. Physiological signals are employed in alert systems to monitor the user's state. For instance, if a driver is too fatigued, unresponsive, or unwell to drive, the system can issue alerts and take appropriate actions, such as reducing speed or stopping the vehicle, to enhance driving safety and security (Kashevnik et al., 2021; Yang et al., 2018).

The rapid advancements in Artificial Intelligence emphasize the urgent requirement for intelligent HCI. Researchers are increasingly drawn to emotional computing as a significant area of study within AI (Anderson & Mc Owen, 2006). Emotion recognition research aims to improve human-computer interaction by making it more smooth, natural, and friendly. This entails a transition from machine-centric to human-centric machine design, evolving the computer from a solely logical computing unit into an intuitive perceptron. To achieve this transformation, a crucial prerequisite is the incorporation of affective computing capabilities into the machine or computer. Without emotional intelligence, the computer or device would be unable to attain a level of intelligence equivalent to that of humans (Zhang et l., 2020; Jenke et al., 2014).

2.1. The Concept of Emotion

The initial step toward recognizing emotion involves defining the concept itself. The interdisciplinary nature of the inquiry, spanning computer science, philosophy, and neuroscience, has led to multiple efforts to address this question and formulate a comprehensive definition of emotion. However, consensus is lacking, and discord persists, with no universally accepted definition. The significance of defining emotion is particularly pronounced in Machine Learning (ML), where a clear definition is essential for establishing success criteria.

To address this challenge, a common strategy involves categorizing emotions using two models: continuous and discrete (Chen & Zhang, 2017). The general process of emotional computing based on physiological inputs can be outlined in three phases (Halfon et al., 2011):

i) Feature extraction: This involves extracting features from diverse sources of heterogeneous physiological signals, including respiration, pulse rate, galvanic skin reaction, EEG, and ECG.

ii) Emotion recognition: The second step is the identification of the emotional state through the processed physiological data.

iii) Emotional regulation: The final step involves the control or modification of emotions using psychological techniques, completing the cycle of emotional computing.

2.1.1. Discrete Emotion Spaces

Throughout history, philosophers have delved into the realms of emotion and feelings, with traces of such contemplation dating back to (Gravier, 2017). During the Roman Empire, emotions were conceptualized and categorized into four fundamental types: fear, pain, lust, and pleasure. Meanwhile, posits that emotions have an evolutionary origin (Colzato, 2017). transcending cultural boundaries and supporting the notion of natural selection (Deng & Ren, 2021).

In order to capture the complexities and subtleties of emotions, numerous authors have embraced the concept of continuous multi-dimensional space models. In these models, emotions are measured along predetermined axes within a continuous multi-dimensional space, facilitating easier comparison and classification of emotions (Trnka et al., 2021). This approach offers a framework for overcoming the challenges associated with understanding and categorizing the intricate landscape of human emotions.

2.1.2. Continuous Dimensions

Two important factors need to be taken into account in a continuous portrayal of emotion: the ability to delineate correlations between various emotional states, such as sadness and admiration or trust, and the quantification of a specific condition, distinguishing, for example, between very sad, sad, and not sad. Arousal, ranging from calm to excited, and valence, spanning from negative to positive, are fundamental dimensions taken into account in this context. Figure 1 shows the mapping of several emotions inside the two-dimensional valence-arousal space (Gunes, et al., 2011; Sadeeq & Abdulazeez, 2022). This representation aids in capturing the nuanced variations in emotional experiences by considering both the intensity and the positive or negative nature of emotions.

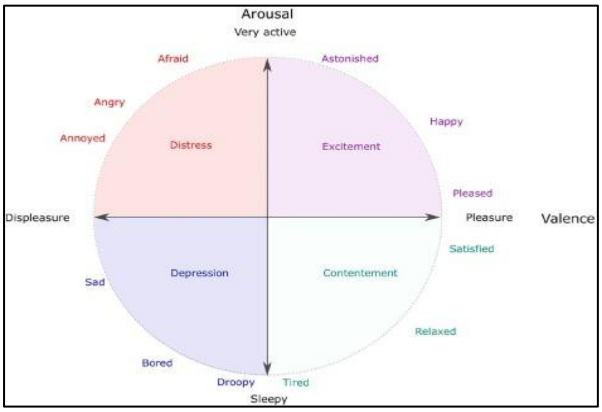

Fig. 1. Valence-arousal model of emotion.

2.2. Autonomic Nervous System

Emotions have been evolutionarily maintained as a crucial mechanism for efficiently mobilizing and coordinating rapid responses from diverse systems when environmental stimuli threaten existence. Despite general agreement with Levenson's theory, theorists diverge on the number of distinct emotional states connected to specific Autonomic Nervous System (ANS) patterns (Mohsin & Beltiukov, 2019; Mendl et al., 2022).

On one side, some theorists assert that the "on" and "off" states represent the only two ANS patterns. Conversely, other researchers propose the existence of numerous ANS activation patterns, each linked to a distinct emotion. The intricate functionality of the Autonomic Nervous System (ANS) introduces an additional challenge in establishing correlations between an individual's emotional state and their present physiological signals. Physiological signal alterations, such as an increase in heart rate or breathing, are more likely attributed to non-emotional ANS functions rather than emotional ones, further complicating the association between physiological responses and emotional states (Giannakakis, 2019).

2.3. Physiological Signals

As noted, there are observable physiological reactions of the Autonomic Nervous System (ANS) that are associated with emotional states. Several wearable physiological sensors can capture these reactions, including the Electrocardiogram (ECG), Electrodermal Activity (EDA), Electroencephalogram (EEG), Blood Volume Pulse (BVP), and others, briefly discussed below (Ping et al., 2013; Sadeeq & Abdulazeez).

i) Electrocardiography (ECG) ECG presents a numerical depiction of potential fluctuations originating from the heart's electrical activity. It records the heart muscle's contraction and relaxation in response to electrical stimuli reaching the skin's surface (Park, 2023).

ii) Electrodermal Activity (EDA) / Galvanic Skin Response (GSR) To measure skin resistance, EDA measures the voltage or current variations between two sensor leads. By introducing a small amount of current or voltage through the body, the skin's reaction resembles that of a variable resistor (Affanni & Chiorboli 2015).

iii) Photoplethysmography (PPG) or Blood Volume Pulse (BVP) PPG quantifies the light backscattered by a skin voxel. In a BVP signal, the amount of light traversing the finger and returning to a sensor correlates directly with the blood volume in the tissue (Blackford er al., 2018).

iv) Respiration (RESP) Administering a chest belt across the abdomen or thorax is necessary to monitor breathing patterns. The belt expands and contracts during air intake and exhalation, providing insights into the subject's breathing rate and depth (Al Alshaikh, 2018).

v) Skin Temperature (Temp) or (SKT) An infrared thermopile or a temperature-dependent resistor are used to measure skin temperature (SKT) at the skin's surface. It indicates changes in skin temperature and represents the impact of the Autonomic Nervous System (ANS) on variables such as physical activity, ambient conditions, and emotional responses (Patil & Pawar, 2022).

vi) Electromyography (EMG) Electromyography (EMG) gauges the electrical activity of skeletal muscles through the use of needle or surface electrodes. Muscle contractions cause an increase in the amplitude of the EMG signal (Patil & Pawar, 2022; Ibrahim et al., 2016).

vii) Electroencephalography (EEG) EEG measures the electrical field generated by currents during synaptic connections between neurons in the cerebral cortex (Brienza, & Mecarelli, 2019).

viii) Eye Gaze Methods including electrooculography (EOG), photoelectric, and infrared reflection detect the eye's resting potential as well as variations during vertical and horizontal eye movements. This provides insights into visual attention and emotional responses (Pazvantov & Petrova, 2022). It is essential to use pre-validated emotional stimuli to achieve a comprehensive portrayal of feelings at varying intensities because of the subjectivity and variation in emotion elicitation (Somarathna et al., 2022).

2.4. Machine Learning Algorithms

Machine learning algorithms have found applications in various sectors, ranging from medicine to economics. One specific field that focuses on analyzing data patterns in an educational setting is education data mining (Ahmed et al., 2021). Once the necessary dataset has been generated with relevant attributes, employing a robust classification method becomes a crucial next step. Encouragement for the multi-class classification of human expressions, vector machines (SVM) are commonly used, frequently in combination with different feature extraction techniques (Raut, 2018). Within the realms of computer vision and machine learning, forecasting face attractiveness stands as a challenging yet pivotal undertaking. The intricacies of human perception and the diversity of facial appearances pose challenges in developing reliable and efficient Face Beauty Prediction (FBP) models (Saeed, et al., 20223). The discussion of the benefits and drawbacks of applying deep learning models is essential for the development of glaucoma screening, diagnosis, and detection systems (Kako & Abdulazeez, 2022). Understanding both the advantages and limitations will contribute to the effective utilization of these models in enhancing glaucoma-related healthcare practices.

2.4.1. Support Vector Machine (SVM)

Among the most effective classification methods, Support Vector Machines (SVM) aim to identify the best hyperplane that accurately separates two classes. The concept of a margin is crucial, representing the maximum distance from both classes to prevent any overlap. To manage non-linear data, kernel functions like polynomial and radial basis function (RBF) are utilized. Rather than using a binary classification, a multi-class Support Vector Machine (SVM) is frequently used in the context of emotion detection to identify a variety of emotions, including fury, fear, disgust, contempt, happiness, sorrow, and surprise.

When comparing multiple machines learning methods and mitigating variations from the database, k-fold cross-validation is commonly applied (Rajesh & Naveenkumar, 2016). Figure 2(a) illustrates the SVM classifier, where the decision boundary is optimized by gamma, and C serves as the misclassification penalty function. Figure 2(b) displays the Optimal Hyperplane generated using the SVM algorithm. Fine-tuning the variables, gamma and C, has a notable impact on the accuracy of classifiers, allowing for optimization in both binary and multi-class classification scenarios (Loconsole, et al., 2014).

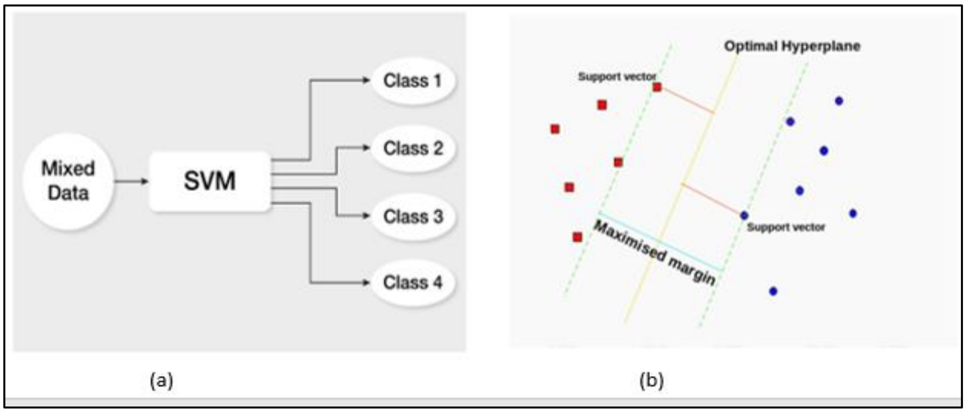

Fig. 2. (a): SVM Classifier, (b) Optimal Hyperplane using the SVM algorithm

2.4.2. Hidden Markov Models (HMM)

Hidden Markov Models (HMM) are statistically useful for revealing hidden patterns within data and are particularly popular for speech-based emotion recognition (Lee, et al., 2008). In this approach, a series of observable features serves as input. The advantage of using both Hidden Markov Models and k-Nearest Neighbors (k-NN) lies in the fact that HMM can perform sophisticated computations, while k-NN only needs to classify between the given samples [53] (Zhou, et al., 2004).

Figure 3 depicts several existing Hidden Markov Model (HMM) topologies: (A) a fully connected HMM; (B) a circular HMM; (C) a left-right HMM. To achieve optimal results for speech emotion recognition, Hidden Markov Models are often employed in serial multiple classifier systems (SVC +

HMM). In such systems, Hidden Markov Models (HMMs) are employed for sample training, and classification is managed by Support Vector Machines (SVM). SVM provides a direct classification rather than a score, enhancing its applicability in this context (Nijs, et al., 2016; Sadeeq & Abdulazeez, 2023).

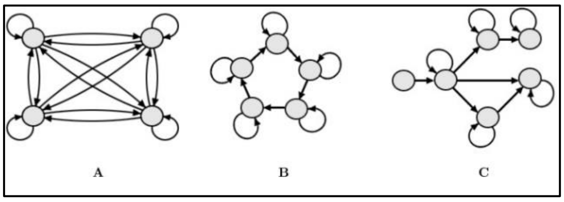

Fig. 3. HMM topologies: **(A)** a fully connected HMM; **(B)** a circular HMM; **(C)** a left-right **HMM**

2.4.3. Random Forest Classifiers

In certain instances, decision trees have demonstrated their superiority over Support Vector Machines (SVM). Built on decision trees, random forests improve performance by using several forests or classifiers rather than just one to identify the target variable's class. S1 through S5 represent the subset of emotions that are used for detection. Various techniques, including K-Nearest Neighbor, Neural Networks (ANN), and Linear Discriminant Analysis, are employed for emotion prediction and categorization (Chen, et al., 2017). Figure 4 illustrates the classifier for this type, showcasing the architecture and methodology used in these approaches.

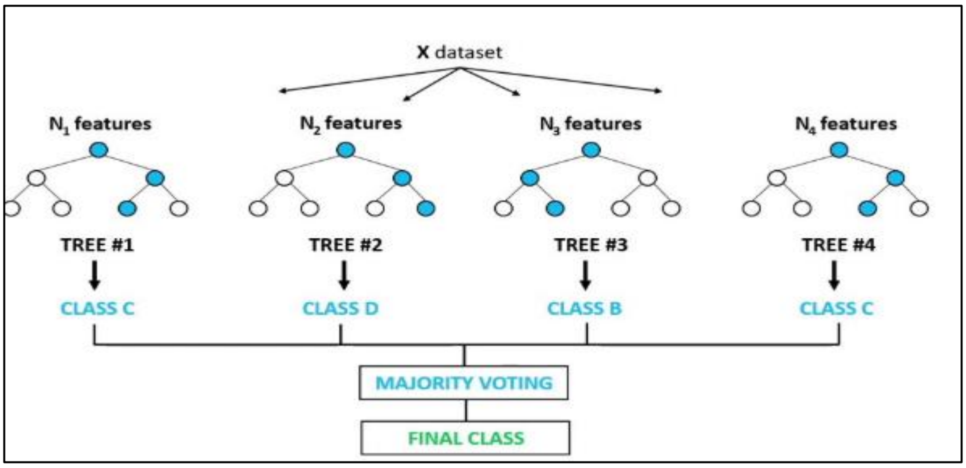

Fig. 4. Random forest classifier (Vaidya, et al., 2022)

3. METHODS

The diagram provided in the accompanying image outlines the key processes essential for building a machine learning system dedicated to emotion recognition, as depicted in Figure 5.

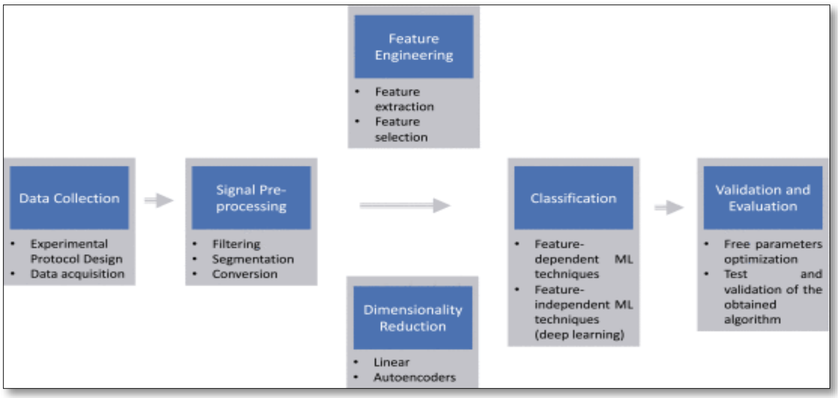

Fig. 5. Schematic illustration of an emotion identification machine learning technique (Zhang et al., 2020)

The data acquisition protocol in developing a machine learning (ML) system for emotion recognition is susceptible to various issues that introduce noise and external interference into the sensor signal. Factors such as subject movement, electrode disconnection, environmental changes in humidity and temperature, electrostatic artifacts, and unexpected user movements can lead to signal degradation. Consequently, the initial stage in ML system development typically involves applying signal pre-processing techniques to the raw signal. This involves filtering, noise reduction, and outlier removal, synchronizing signals from various sensors, and addressing null values and data loss through methods such as linear interpolation (Hosseini & Khalilzadeh, 2010).

Convolutional Neural Networks (CNNs) are recognized as trainable systems with the ability to reduce dimensionality and acquire discriminative features. Identifying emotional states can be accomplished through two primary methods: (1) feature-dependent machine learning techniques that rely on feature-class representation, and (2) feature-independent machine learning methods, including approaches within deep learning (DL) (Patil & Pawar, 2022; Haji, et al., 2021). To address challenges like overfitting and limited dataset size, researchers have recently turned to transfer learning and components of deep convolutional neural networks (DCNN). However, the implementation of large CNN systems remains challenging due to their computational intensity and the millions of parameters involved, particularly on small devices with limited hardware resources (Saeed, et al., 2022). In traditional machine learning system design, a feature engineering stage is often introduced after signal pre-processing to optimize the useful content of physiological signals. Figure 6 depicts a common diagram of the ML model building process.

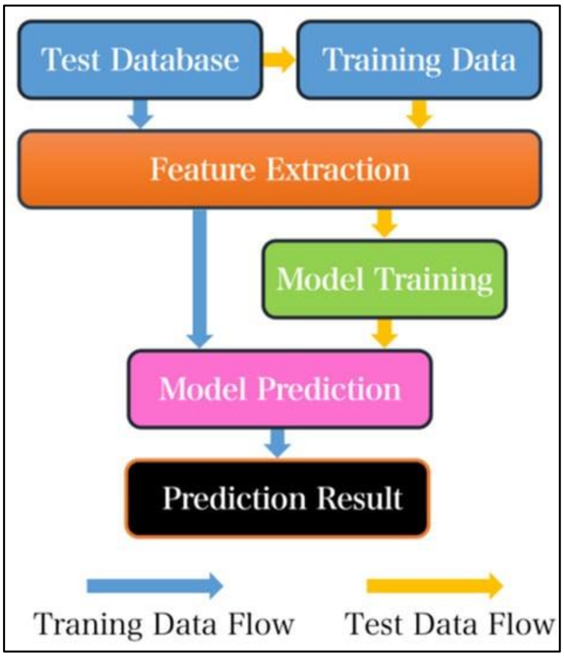

Fig. 6. Common ML model building process diagram.

Once feature engineering is completed on the incoming input, a classifier outputs the subject's emotion class label (Zhu, et al., 1988; Tripathi, et al., 2017). These metrics, commonly known as features, offer a succinct description of the signal, enabling comparisons across different signals in transformed dimensions and augmenting the informative characteristics of the signals. These features may be linear or non-linear, unimodal or multimodal, and can belong to the temporal, statistical, or spectral domains. Figure 7 illustrates the data variable types (Patil & Pawar, 2022).

Research findings indicate that Convolutional Neural Networks (CNNs), as a deep data-driven technique, exhibit effectiveness in extracting or predicting facial attractiveness from images (Abdulkareem & Abdulazeez. 2021). Improved performance is achieved by CNN models with deeper structures, larger input images, and smaller convolution kernels (Saeed et al., 2023; Alzubaidi, et al., 2021). Genetic algorithms (GA) and differential evolution (DE) are two widely recognized algorithms designed to simulate the genetic process of reproduction (Sadeeq & Abdulazeez, 2023) .

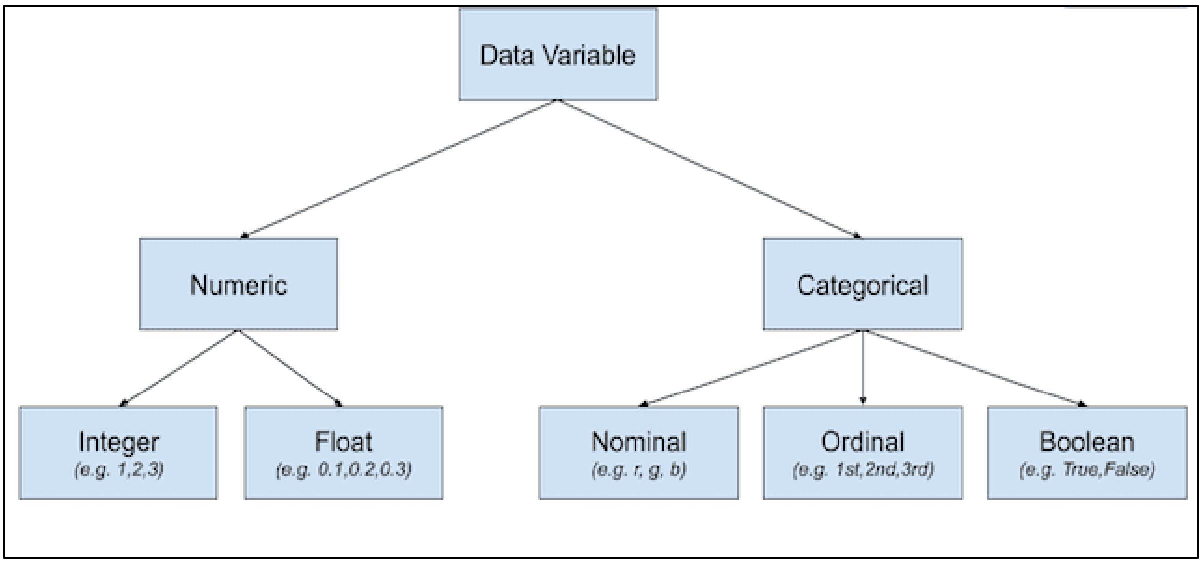

Fig. 7. Data variable types

4. LITERATURE REVIEW

K. P. Seng and colleagues proposed an approach aimed at emotion recognition in audio and video streams (Seng, et al., 2016). Their method integrates machine learning and rule-based strategies to enhance the effectiveness of emotion recognition. The visual route is established to achieve dimensionality discrimination and reduction, utilizing Bi-directional Principal Component Analysis (BDPCA) and Least-Square Linear Discriminant Analysis (LSLDA). The visual characteristics are subsequently analyzed by an Optimized Kernel-Laplacian Radial Basis Function (OKL-RBF) neural classifier. In the audio route, features such as Mel-scale frequency cepstral coefficients, spectral properties, log-energy, pitch, teager energy operator, and zero crossing rates, are employed. This comprehensive approach seeks to improve recognition efficacy by combining advanced techniques for both visual and auditory emotional cues.

Introduced a sophisticated linear model designed to distinguish facial movements in expressive face recordings with diverse linearly-representable characteristics (Xiang & Tran, 2017). In contrast to previous approaches that required a clear but somewhat unrealistic dissociation of identity and expression, their approach uses sparse representation just on the residual expression components and simultaneously captures the underlying neutral face. This is accomplished by implicitly subtracting the neutral face and leveraging the low-rank characteristic between frames. In experiments conducted on manually created expression components, their one-shot C-HiSLR, when applied to raw-face pixel intensities, demonstrated superior performance compared to traditional shape + SVM models with landmark detection and two-stepped Sparse Representation Classification on CK+.

Employed transfer learning by training Convolutional Neural Networks (CNNs) with millions of images, allowing the knowledge gained in training to be applied to a different task (Shaees, et al., 2020). AlexNet, the selected pre-trained CNN, uses a hybrid classifier that blends transfer learning with a Support Vector Machine (SVM)-like classification methodology. The evaluation of their approach involved testing it on the Cohn-Kanade+ (CK+) and Natural Visible and Infrared Expression (NVIE) databases, both widely used expression databases.

Presented a hybrid convolution-recurrent neural network method for recognizing facial emotions in photos (FER) (Jain, et al., 2018). Convolution layers and a recurrent neural network (RNN) are used in the suggested network design to make it easier to discover correlations from facial photos. During the classification process, temporal dependencies in the images are accommodated by the recurrent network. The researchers assessed their hybrid model using two publicly available datasets, demonstrating promising experimental results. In the data preparation stage, each facial outline is initially identified using a face and point of interest finder. Following normalization, the nose, lips, and nose are aligned, and mean subtraction and contrast normalization are applied when processing each facial image through the CCN.

Introduced a system incorporating a random forest classifier for facial expression detection (Arora, et al., 2018). The experiment assessed the system's performance in recognizing five common emotions—sadness, joy, anger, neutrality, and surprise—utilizing data from the Japanese Female Facial Emotion (JAFFE) database. The proposed framework shows promise for real-life applications, particularly in conjunction with electroencephalograms and brain-computer interfaces. Considering that facial emotions arise from facial muscle deformations, the system utilizes gradient features, well-known for their sensitivity to object deformations, to encode these facial components. Emotion classification is the next testing phase, when assessment parameters like false acceptance rate, false rejection rate, and recognition accuracy are measured.

devised a system focused on recognizing students' emotions based on facial cues, employing a three-phase approach: The process involves face detection using Haar Cascades, normalization, and emotion recognition using Convolutional Neural Networks (CNN) on the FER 2013 database, which includes seven distinct expression types. The findings suggest the feasibility of detecting facial emotions in educational settings, offering potential assistance to teachers in adjusting their presentations based on students' emotional states (Lasri, et al., 2019).

Presented a method for recognizing facial expressions using image edge detection and a Convolutional Neural Network (CNN) (Zhang, et al., 2019). The process involves normalizing the facial expression image, extracting edges using convolution, incorporating edge information onto feature images to preserve texture details, reducing dimensionality through maximum pooling, and ultimately employing a Softmax classifier for expression classification. A simulation experiment evaluates the method's robustness in facial expression identification against complex backgrounds, combining the Fer-2013 facial expression database is combined with the Labeled Faces in the Wild (LFW) dataset for scientific testing purposes.

In order to assess primary emotions in images, like happiness or sadness, Verma built a Convolutional Neural Network (CNN) with two components: the first one predicts the secondary emotion, while the second one analyzes the fundamental emotion (Verma & Verma, 2020). After being trained using the FER2013 and Japanese female facial expression (JAFFE) datasets, the model showed superior capacity to predict emotions from facial expressions than existing state-of-the-art approaches.

The authors' use of facial expression analysis has improved the ability to identify emotions. The phases of creating, honing, and testing an algorithm for emotion recognition based on logistic regression are described in their work. The study offers comprehensive information about the optimization process and outcomes in training and test sets (Barrionuevo et al., 2020).

A Convolutional Neural Network (FERC) was used to propose a novel approach to facial emotion recognition (Mehendale, 2020). The CNN in the FERC is divided into two segments: the first segment eliminates the backdrop of the image, and the second segment concentrates on extracting facial feature vectors. An expressional vector (EV) is used by the FERC model to distinguish between five different categories of typical facial expressions. Supervisory information was gathered from a 10,000-photo collection that included 154 different people. The last perceptron layer modifies weights and exponent values in each iteration of the two-level CNN, which functions in a sequential fashion. Notably, FERC differs from standard methods by employing a single-level CNN, which adds to improved accuracy. Moreover, a new backdrop removal method that is applied prior to expressional vector (EV) production tackles possible problems such changes in camera distance.

Introduced a modular system designed for the recognition of human facial emotions, consisting of two machine learning algorithms for offline training and subsequent real-time application, specifically for detection and classification (Alreshidi & Ullah, 2020). In the initial phase, AdaBoost cascade classifiers are utilized to detect faces in images. Subsequently, neighborhood difference features (NDF) are extracted based on localized appearance information, representing facial features. Unlike focusing solely on intensity data, the NDF predicts various patterns based on interactions between nearby zones, emphasizing significant facial expressions commonly encountered in daily life. To handle mis-/false detection, the researchers train a random forest classifier incorporating a hidden emotional state. Additionally, the proposed emotion identification technique is insensitive to gender or skin tone. Robust feature models are developed to eliminate unnecessary and redundant data from detected faces. These characteristics are orientation and illumination invariant, and the approach avoids both overfitting and underfitting due to the strengths of random forest trees in classification. Even with limited samples, random forests can deliver reliable results.

A genetic algorithm (GA) was integrated with support vector machine (SVM)-based classification to address a multi-attribute optimization problem involving feature and parameter selection. The research used two datasets: the Multimedia Understanding Group (MUG) dataset and the enlarged Cohn-Kanade dataset (CK+). They contrasted their method with convolutional neural networks (CNNs), a popular method for identifying emotions on faces (Liu, et al., 2020).

A directed graph neural network (DGNN), more precisely a graph convolutional neural network that uses landmark information for facial emotion recognition (FER), was introduced by Q. T. Ngo et al. in their proposal (Ngoc, et al., 2020). In this model, landmarks serve as nodes in the graph structure, and the Delaunay method constructs the edges in the directed graph. By leveraging the underlying geometrical and temporal information present on faces, emotional cues are extracted using graph neural networks, aiming to avoid the vanishing gradient issue. Additionally, their model includes a stable temporal block in the graph architecture.

The researchers introduced a modeling technique for Facial Emotion Recognition (FER) based on a Deep Convolutional Neural Network (DCNN) and Transfer Learning (TL) (Aknand, 2021). They leveraged a pre-trained DCNN model, adapting its dense top layer(s) for FER and further fine-tuning the model with facial expression data.

Their innovative pipeline involved initially training the dense layer(s) and sequentially adjusting each pre-trained DCNN block. This iterative process consistently improved FER accuracy to a higher level. The validation utilized well-known facial image datasets (JAFFE and KDEF) and eight different pre-trained DCNN models (VGG-16, VGG-19, ResNet-50, ResNet-34, ResNet-152, ResNet-18, DenseNet-161, and Inception-v3).

A. Poulose et al. successfully extracted FER features with their proposed Facial Emotion Recognition (FER) technique, which is based on foreground extraction (Poulse, et al., 2021). For model training, their approach makes use of a deep learning model that makes the best use of these features. Comparative findings show that, as compared to the conventional FER approach, the Facial Emotion Recognition (FER) approach yields more accurate classification results during model training. Nine users' emotions were gathered to validate the proposed FER method, which makes use of the deep learning model is the Xception architecture.

A unique method for facial expression identification using deep neural networks inside a decision tree framework was reported by M. A (Ruzainie, et al., 2021). Discrete Cosine Transform (DCT) coefficients with low frequencies are arranged to represent expression features. By applying 2-D DCT as an unsupervised feature extractor on the difference images between neutral and expression photos, these coefficients are discovered. They are not used because of the practical difficulties in obtaining geometric properties for practical uses. The first two decision tree nodes focus on three expressions: surprise, smiling, and melancholy. Three efficient one-hidden-layer multilayer projections (OHL-MLP), trained by the back-propagation approach, are used to implement them. The third node in the decision tree, which is a deep neural network implementation, handles and evaluates the three additional expressions: disgust, fury, and fear. An autoencoder is first applied to the directly concatenated DCT coefficients of several facial components, such as the mouth, nose, and eyes, in order to integrate and fine-tune the features. The next step is to train an OHL-MLP to classify the target expressions.

A system that divides emotions into happy, normal, and shocked categories was presented (Sujanaa, et al., 2021). The video frames in the dataset were retrieved at a rate of twenty frames per second and depict a range of moods. A Haar-based cascade classifier is used to segment the mouth area in the facial images. The system extracts edge and local information as well as gradient information from the emotion image using the local binary pattern (LBP) and the histogram of oriented gradients (HOG). Every mouth image is represented by a single histogram that is created by combining these elements.

In order to facilitate multimodal emotion detection, proposed a method to increase heterogeneity between different modalities and create a complementary relationship (Chen, et al., 2022). Experiments were conducted on the SAVEE and eNTERFACE'05 datasets to assess the accuracy of the suggested method. In order to combine multimodal data—which includes time and frequency domain data gathered from voice and facial expressions—Kernel canonical correlation analysis was utilized. To select features from several modalities and lower dimensionality, K-means clustering was used.

The goal was to determine which classifier could best recognize negative emotions including fear, rage, disgust, and melancholy. Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron Neural Network (MLPNN), Radial Basis Function (RBF), K-Nearest Neighbor (KNN), Decision tree (J48), and Neural Network (NN) were among the classifiers whose efficacy was evaluated (Tiwari & Veenadhari). The dimensionality was reduced using Principal Component Analysis (PCA), and feature extraction techniques like Gabor wavelet, Chi-Square, Local Binary Pattern (LBP), and Histogram of Gradient (HOG), were used.

A unique method for facial expression identification using convolutional neural networks (FERC) (Savakar, et al., 2023). Two segments make up the CNN network's FERC model: the first segment eliminates the background of the image, and the second segment gets rid of the face vector. The FERC model detects five distinct types of regular facial expressions using the expressional vector (EV). The exponent values and weights of the last perceptual layer change with each iteration of the continuous double-level CNN. FERC stands out from typical CNN single-level technology, enhancing accuracy. CNN involves facial recognition, image processing, object identification, and related technologies, utilizing multiple layers in a deep neural network to potentially extract significant features from the data.

The conducted experiments to extract facial traits with the aim of improving facial emotion recognition (Subudhiray, et al., 2023). They used feature extraction approaches such as Gabor, local binary pattern (LBP), and histogram of oriented gradient (HOG) in addition to conventional k-nearest neighbor (kNN) classification. Performance metrics, such as recall, kappa coefficient, computation time, precision, average and overall recognition accuracy, and recall, were compared. The experiment utilized two databases: Japanese female facial expression (JAFFE) and Cohn-Kanade (CK+).

5. Comparison Among Reviewed Works

Table 1 show A comparison table with different algorithms.

Table 1. A comparison table with different algorithms

#	Ref.	Year	Description	Results
1.	(Seng, et al., 2016)	2016	The Least-Square Linear Discriminant Analysis (LSLDA) and Bi-directional Principal Component Analysis (BDPCA) are used for discriminating and	The system's performance is evaluated using standard databases to assess both the visual path and the proposed auditory path. Comparative analyses and experimental results indicate

#	Ref.	Year	Description	Results
			dimensionality reduction. .	the effectiveness of the suggested system. To further improve recognition efficacy, the authors suggest an audio-visual emotion identification system that combines rule-based and machine learning techniques.
Dataset Type:			Standard databases are used to assess the performance of the finished system, the visual path, and the suggested aural path	
2.	(Xiang & Tran, 2017)	2017	SVM models with two-step SRC of the same type and landmark detection, but applied to manually created expression components. It is also comparable to temporal models like Bayes nets and CRF	The results provide encouragement for employing a piecewise linear model as an approximation for an expression. This approach can be compared to temporal models such as Bayes nets and CRF, as well as the piecewise linear model DCS. When applied to recognize action units (AUs) on MPI-VDB, the obtained results are noteworthy.
Dataset Type:			the piecewise linear model DCS used an action unit (AUs) on MPI-VDB, obtaining a respectable performance	
3.	(Shaees, et al., 2020)	2018	Used transfer learning and trained CNNs. Among the greatest CNNs with pre-training is AlexNet.Two widely used expression databases	Clearly demonstrates that CNNs trained through learning outperform manually crafted methods in terms of outcomes.
Dataset Type:			Tested have been Cohn-Kanade+ (CK+) and Natural Visible and Infrared Expression (NVIE).	

#	Ref.	Year	Description	Results
4.	(Jain, et al., 2018)	2018	A hybrid deep CNN and RNN model was proposed; it was assessed under various conditions and with the use of hyper parameters to ensure that it was tuned appropriately.	The efficacy of the proposed model was confirmed, as the combination of the two neural networks (CNN-RNN) significantly improved the overall detection results.
Dataset Type:			Two publicly available datasets are used to assess the suggested hybrid model: ISO and FA-KES.	
5.	(Arora, et al., 2018)	2018	An efficient framework proposed system hybridization of Gradient filter, For facial emotion recognition, an efficient framework of Gradient filter, PCA, and PSO hybridization has been presented.	The experimental results indicate that our proposed technique achieves a competitive classification accuracy. Specifically, on the JAFEE dataset, the average classification rate is 91.3%. According to the experimental findings, our approach demonstrates competitive classification accuracy. The utilization of a novel hybridization involving Gradient filter, PCA, and PSO for facial emotion identification, not explored before, outperforms existing algorithms.
Dataset Type:			JAFEE dataset had been used	
6.	(Lasri, et al., 2019)	2019	A CNN model that has been trained to evaluate photos and recognize face emotion.	The results obtained suggest the feasibility of facial emotion detection in educational settings. Consequently, it can aid teachers in adjusting their presentations based on the emotions of their students. This is accomplished by employing a Convolutional Neural Networks (CNN) model trained to analyze

#	Ref.	Year	Description	Results
				images and identify facial emotions.
Dataset Type:			FER 2013 database with seven types of expression	
7.	(Zhang, et al., 2019)	2019	A method for recognizing facial expressions based on a convolutional neural network (CNN) and image edge detection. By systematically combining the Fer-2013 facial expression database with the	The experimental results demonstrate that the proposed algorithm shows a training speed approximately 1.5 times faster than the comparative method on the training set. In addition, it requires less iterations to reach an average recognition rate of 88.56%.
Dataset Type:			LFW data set, a simulation experiment is created.	
8.	(Verma & Verma, 2020)	2020	A convolution neural network (CNN) is used in an enhanced deep learning method to analyze facial expressions in an image and predict emotions.	The findings imply that facial expressions, rather than the most advanced techniques available today, are a more accurate way to anticipate moods.
Dataset Type:			FERC was thoroughly evaluated using expanded Cohn-Kanade expression on over 750K photos.	
9.	(Barrionuevo et al., 2020)	2020	Using Supervised Learning with Logistic Regression	Capable of utilizing facial expression analysis to differentiate emotions.
Dataset Type:			The Extended Cohn-Kanade (**CK+**) dataset contains 593 video sequences from a total of 123 different subjects	
10.	(Mehendale, 2020)	2020	utilizing convolutional neural networks to recognize facial emotions (FERC).	Emphasizing the emotion achieved a 96% accuracy rate when employing an EV of length 24 values. The two-level CNN operates

#	Ref.	Year	Description	Results
			Based on a two-part convolutional neural network (CNN), the FERC Caltech faces, CMU, NIST	sequentially, each iteration, the final layer of the perceptron modifies the weights and exponent values. FERC employs a single-level CNN in a distinct manner compared to most techniques, leading to enhanced accuracy.
Dataset Type:			Extended Cohn-Kanade expression datasets were used to test FERC extensively on over 750K photos, and the 10,000 photos (154 people) in the database that was saved provided supervisory data.	
11.	(Alreshidi & Ullah, 2020)	2020	Utilizing a random forest classifier trained to identify emotions during testing, detect faces and extract neighborhood difference features (NDF) based on the relationships between neighboring regions.	The proposed method exhibits superior performance on the static facial expressions in the wild (SFEW) and real-world emotional faces (RAF) datasets, surpassing the reference methods by 13% and 24%, respectively.
Dataset Type:			The method is evaluated on several benchmark datasets and contrasted with the five-reference method (SFEW) 2.0 dataset and the real-world affective faces (RAF) dataset.	
12.	(Liu, et al., 2020)	2020	The method combines a genetic algorithm (GA) with support vector machine (SVM) based classification to solve a multi-attribute optimization issue involving feature and parameter selection.	The test accuracy for the proposed model achieved 95.85%, 97.59%, and 96.56%, respectively, with corresponding validation accuracy values of 93.57%, 95.58%, and 96.29%. The F1-score, recall, and overall precision were approximately 0.96, 0.95, and 0.97, demonstrating strong performance across these metrics.
Dataset Type:			The Multimedia Understanding Group (MUG) dataset and the enlarged Cohn-Kanade dataset (CK+) were	

#	Ref.	Year	Description	Results
			used for experimental evaluations. For CK+ classes 8, 7, and MUG classes 7.	
13.	(Ngoc, et al., 2020)	2020	A directed graph neural network (DGNN) is designed that makes use of landmark characteristics for FER.	The testing results show that the suggested approach works well on datasets like AFEW (32.64%), MMI (69.4%), and CK+ (96.02%). Moreover, a fusion network utilizing both landmarks and image data is introduced and examined, showcasing high performance with 98.47% on the CK+ dataset and 50.65% on the AFEW dataset.
Dataset Type:			The CK+, MMI, and AFEW databases were been used	
14.	(Aknand, 2021)	2021	An efficient DCNN using TL with pipeline tuning strategy has been proposed for emotion recognition from facial images. The experiment's findings indicate that eight distinct pre-trained	The evaluation's conclusions demonstrate that the recommended FER system outperforms the current one in terms of emotion recognition accuracy. Furthermore, promising results are observed on the KDEF dataset, especially with profile views, suggesting the requisite expertise for practical uses. The recommended approach, which made use of pre-trained models, showed excellent accuracy on both datasets. Using a 10-fold cross-validation technique, the best FER accuracies achieved with DenseNet-161 on the test sets of KDEF and JAFFE are 96.51% and 99.52%, respectively.

#	Ref.	Year	Description	Results
Dataset Type:			DCNN models using the popular KDEF and JAFFE emotion datasets with various profile views	
15.	(Poulse, et al., 2021)	2021	The system's deep learning model efficiently uses the FER features that are successfully extracted by the suggested foreground extraction-based FER approach for model training.	The FER experiment and result analysis show that the suggested foreground extraction-based strategy greatly minimizes the classification error seen in the conventional FER approach. In comparison to the normal FER approach, the suggested method's FER results indicate a 3.33% improvement in model correctness.
Dataset Type:			Extended Cohn-Kanade (CK+), AffectNet, JAFFE	
16.	(Ruzainie, et al., 2021)	2021	Decision tree framework and deep neural networks Using deep neural networks and decision trees with 123 subjects and 70 photos for every expression being gathered and used (50 for the training phase and 20 for the testing phase),	The initial node achieves a recognition rate of 99.17% for both surprised and smiling faces. The second node successfully detects sadness in every instance (100%). The last node exhibits accuracy rates of 100.00%, 70.00%, and 55.00% for the facial expressions of fear, rage, and disgust, respectively.
Dataset Type:			Applying the recommended recognition algorithm to the enlarged Cohn-Kanade (CK+) database	
17.	(Sujanaa, et al., 2021)	2021	Support vector machines (SVMs) and one-dimensional convolutional neural networks (1D-CNNs) are used to extract	According to the experimental findings, 1D-CNN and SVM attain accuracies of 97.44% and 98.51%, respectively. These texture features are

#	Ref.	Year	Description	Results
			gradient information from emotion images, extract the edge and local information from the histogram of oriented gradients (HOG) and local binary pattern (LBP), and finally combine them into a single histogram to form the features, where each histogram represents a mouth image.	employed to train both the support vector machine (SVM) and the one-dimensional convolutional neural network (1D-CNN), enabling the detection of emotions in test video frames using the trained algorithms.
Dataset Type:			Scale-invariant feature transform (SIFT) and accelerated robust features (SURF) techniques are used to extract unique points with changing numbers of key-points as features from emotion images.	
18.	(Chen, et al., 2022)	2022	Clustering based on K-means in human-robot interaction (HRI), a kernel canonical correlation analysis approach is suggested for multimodal emotion recognition. .	The results indicate that the proposed method achieves favorable recognition rates, surpassing those obtained by methods without K-means clustering. Specifically, the proposed method demonstrates recognition rates that are 4.7% higher in eNTERFACE'05 and 2.77% higher in SAVEE compared to methods that do not incorporate K-means clustering.
Dataset Type:			The correctness of the suggested strategy is assessed through experiments on two datasets: SAVEE and eNTERFACE'05	
19.	(Tiwari & Veenadhari)	2022	Evaluation is done on the classifiers for Support Vector Machine (SVM), Radial basis function (RBF),	The K-nearest neighbors (KNN) classifier exhibits a commendable 93.46% classification accuracy in

#	Ref.	Year	Description	Results
			Random Forest (RF), Multilayer Perceptron Neural Network (MLPNN), Decision tree (J48), K-Nearest Neighbor (KNN), and Neural Network (NN).	identifying emotions associated with stress.
Dataset Type:			The Extended Cohn Kanade (CK+), JAFFE, and MMI facial expression datasets were used in the evaluation procedure.	
20.	(Savakar, et al., 2023)	2023	Convolutional Neural Network (CNN) technique A deep neural network (DNN) is made up of multiple neural network layers. This could potentially extract significant features from the information.	The primary challenge arose from insufficiently measured information, hindering the development of a comprehensive framework. However, for the sake of maintaining logical coherence, this proposition needs to be rejected. A model utilizing remote data was substantiated through initial research in the startup phase, confirming the hypothesis.
Dataset Type:			Using the Keras library	
21.	(Subudhiray, et al., 2023)	2023	A simple k-nearest neighbor (kNN) classifier: Three retrieved face traits were tested using a basic k-nearest neighbor (kNN) classifier to see if they could improve the identification of facial emotions.	Although it demands the most computing time, it achieves the highest average accuracy. In contrast, HOG exhibited the lowest average accuracy at 55.2% with the shortest calculation time, while LBP demonstrated an average accuracy of 88.2%, surpassing Gabor's accuracy with a shorter computational time.

#	Ref.	Year	Description	Results
Dataset Type:			Two databases have been utilized: Japanese female facial expression (JAFFE) and Cohn-Kanade (CK+).	

6. DISCUSSION

This paper provides a comprehensive analysis and comparison of facial emotion recognition (FER) algorithms. By utilizing conventional Machine Learning (ML) classification techniques like random forest, AdaBoost, KNN, and SVM, FER relies less on face physics-based models [65]. An audio feature level fusion module employs pre-established rules to identify the most likely emotion in the audio stream after extracting the audio characteristics. The outputs from the visual and acoustic routes are combined via a fusion module, and the efficacy of the system is assessed by employing benchmark datasets. In another work [66], the study explores the use of a piecewise linear model to approximate facial expressions, considering expressions as mixtures of Action Units (AUs). Their one-shot C-HiSLR on raw-face pixel-intensities outperforms two-step SRC and traditional shape+SVM models on CK+, demonstrating comparable results to temporal models like Bayes nets, CRF, and the piecewise linear model DCS when applied to recognizing AUs on MPI-VDB.

In a study by researchers [67], it was observed that training Convolutional Neural Networks (CNNs) from scratch requires substantial data, leading to the adoption of transfer learning—a technique involving the training of CNNs on vast image datasets. This method, known as transfer learning, allows the knowledge gained during training to be applied to a different task. Notably, comparisons were made between pre-trained CNNs like AlexNet that used hybrid classifiers and transfer learning with classification techniques like Support Vector Machine (SVM). The evaluation was conducted on two widely used expression databases, Natural Visible and Infrared Expression (NVIE), and Cohn-Kanade+ (CK+), with results unequivocally indicating the superior performance of pre-trained CNNs over handcrafted techniques. In another study in the same year], a methodology for recognizing emotions on faces was proposed, involving a combination of deep CNN and Recurrent Neural Network (RNN) models (Jain, et al., 2018). After testing the proposed model in many scenarios and fine-tuning its hyperparameters, it was discovered that combining the two kinds of neural networks (CNN-RNN) greatly enhanced the overall detection results. Furthermore, in a different work, the researchers employed gradients to encode facial traits as components of the face, training a random forest classifier to identify emotions (Arora, et al., 2018). The suggested system was accurate in identifying common emotions, such as happy, sad, angry, neutral, and astonished, when tested on the Japanese Female Facial Emotion (JAFFE) database. Furthermore, a study demonstrated how CNN models can be trained to recognize facial emotions, suggesting that facial expression detection is a feasible application in educational contexts (Lasri, et al., 2019). This capability could enable teachers to adapt their lessons based on the emotional states of students (Zhang, et al., 2019).

The proposed approach aims to mitigate the limitations imposed by artificial design elements and facilitate automatic learning of pattern features. This method utilizes image data from training samples, directly inputting pixel values from each image. Through unconscious, autonomous learning, the model can capture more abstract features of the images. The appropriate initialization of weights during the training phase significantly influences weight updates, contributing to the effectiveness of the suggested strategy. Furthermore, the suggested method demonstrates its capability to enhance the identification of facial expressions, particularly in complex background scenarios. Compared to FRR-CNN and R-CNN models, the suggested model converges substantially faster in complicated backdrop environments. Additionally, the suggested strategy achieves a higher recognition rate, showcasing its effectiveness in handling complex background environments ((Verma & Verma, 2020)).

The authors employ a two-CNN approach, where the first CNN determines the primary emotion (happiness or sadness), and the second CNN identifies the secondary emotion. The results suggest that this method outperforms current state-of-the-art techniques in accurately detecting emotions from facial expressions. In another study, the development of a logistic regression-based emotion detection system is detailed, outlining the phases of algorithm development, training, and testing (Barrionuevo et al., 2020). The logistic regression algorithm enhances the ability to discern emotions through facial expression analysis. The FERC (Facial Expression Recognition Convolutional) model proposed by consists of a two-part CNN (Mehendale, 2020). The initial section concentrates on obtaining vectors of the face's features, whilst the subsequent section removes the image's background. The FERC model's expressional vector (EV) successfully distinguishes five different types of regular facial expressions. With an EV of length 24 values, the model exhibits great accuracy, indicating emotions with 96% accuracy. Neighborhood difference features (NDF) are used by the modular approach presented to identify faces and categorize facial emotions into seven distinct states (Alreshidi & Ullah, 2020). A random forest classifier is trained to categorize facial expressions into seven groups during testing. In the work of Support vector machine (SVM) based classification is combined with a genetic algorithm (GA) to solve a multi-attribute optimization issue that includes feature and parameter selection. For the purpose of recognizing facial emotions, convolutional neural networks (CNNs) and this approach are contrasted in the studies (Liu, et al., 2020).

A convolutional neural network (CNN), a well-liked approach for identifying emotions on faces, and the suggested method were compared in the study by (Liu, et al., 2020). In terms of test accuracy, CNN was slightly surpassed by the suggested technique, which combines a genetic algorithm (GA) with support vector machine (SVM) based classification, for both the 8-class CK+ (95.85% (SVM) vs. 95.43% (CNN)) and 7-

class CK+ (97.59 vs. 97.34) tests. On the 7-class MUG dataset, CNN fared somewhat better (96.56 vs. 99.62). The suggested approach shows promise for real-time machine vision applications in automated systems since it makes use of simpler models than CNN-based algorithms.

A directed graph neural network (DGNN) for landmark-based facial expression recognition (FER) was shown in (Ngoc, et al., 2020). The DGNN outperformed cutting-edge image- or video-based algorithms in terms of performance. For both the CK+ and AFEW datasets, state-of-the-art performance of 98.47% and 50.65% was obtained when the recommended strategy was paired with a conventional video-based method. In the future, the FER system might be expanded to incorporate more modalities, like facial features, auditory signals, and physical movements. The recommended strategy has remarkably high recognition accuracy Aknand, 2021). The overall test setup and a few challenging face photos demonstrated the technology's potential for real-world commercial applications, such as hospital patient monitoring and surveillance security. The concept of face emotion detection may be extended to body language or speech recognition for new industrial applications. The kind of FER dataset utilized for model training was discovered to have an impact on the FER system's performance in [79]. Training using publicly available datasets including FER 2013, Affect Net, JAFFE, and extended Cohn-Kanade (CK+), the proposed method showed a 3.33% increase in model validity over the traditional FER approach. A decision tree structure and deep neural networks were used in the method by (Ruzainie, et al., 2021). The 2-D DCT was used on difference images between neutral and expressive shots for unsupervised feature extraction. High accuracy rates for identifying a range of facial emotions were shown by the final decision tree nodes (Sujanaa, et al., 2021).

A framework consisting of the happy, normal, and surprise emotion categories was proposed (Chen, et al., 2022). The video frames, acquired at a rate of twenty frames per second, were segmented using a Haar-based cascade classifier. The classifier was trained to concentrate on the facial image's mouth area. The histogram of oriented gradients (HOG) and the localized binary pattern (LBP) were employed to extract the gradient information from the emotion image. These features were combined into a single histogram, each of which represented a mouth image, to form the characteristics.

A strategy for enhancing heterogeneity between various modalities and creating a complementary relationship between them for multimodal emotion detection was suggested (Chen, et al., 2022). The suggested approach utilized Kernel canonical correlation analysis to merge multimodal information from speech and facial expression. In order to choose features from many modalities and reduce dimensionality, K-means clustering was used. SAVEE and eNTERFACE'05 were the two datasets used in the studies.

Finding the classifier that could most accurately recognize negative emotions including fear, fury, contempt, and sorrow was the aim of research (Tiwari & Veenadhari). The classifiers evaluated were Support Vector Machine (SVM), Radial Basis Function (RBF), Decision tree (J48), Random Forest (RF), Multilayer Perceptron Neural Network (MLPNN), Neural Network (NN), and K-Nearest Neighbor (KNN), using Principal Component Analysis (PCA) as the dimensionality reduction method. Histogram of Gradient (HOG), Local Binary Pattern (LBP), and Chi-Square, Gabor wavelet was some of the feature extraction techniques used.

In a brand-new technique called convolutional neural networks (FERC) for face expression recognition was unveiled (Savakar, et al., 2023). The backdrop of the image in the first segment and the face vector in the second are eliminated by the two-segment FERC model. The expressional vector (EV) is used by the FERC model to identify the five distinct categories of typical facial expressions. The study emphasizes the value of deep learning, and more especially the CNN technique and the Keras framework.

In a study, the effectiveness of three derived features—the Histogram of Gradient, Research was done on the recognition of facial expressions using Gabor and the Local Binary Pattern (Subudhiray, et al., 2023). Gabor outperformed the others, achieving a recognition accuracy of 94.8%. Every feature extraction method's performance metrics, such as recall, kappa coefficient, computation time, precision, average and total recognition accuracy, and calculation time, were investigated.

7. CONCLUSION

Recent developments in emotion recognition from visual data—particularly facial expressions—are thoroughly examined in this paper. It includes a range of study topics from the last ten years, such as gadgets, emotion models, and classification strategies. SVM was applied and with GA (Aknand, 2021; Sujanaa, et al., 2021; Tiwari & Veenadhari). Yielding accuracy of 98.51% and 93.46% for respectively, out of the twenty-one research that met the criteria. The validation accuracy ranged from 93.57% to 96.29%, while the test accuracy varied from 95.85% to 96.56% (Sujanaa, et al., 2021; Tiwari & Veenadhari). The approximate values for the F1-score, recall, and total precision were 0.96, 0.95, and 0.97. Compared to non-users, produced a greater recognition rate by using k-means (Chen, et al., 2022). CNNs were chosen, and who used 1D-CNN to achieve 97.44% (Lasri, et al., 2019; Zhang, et al., 2019; Verma & Verma, 2020; Mehendale, 2020; Aknand, 2021; Sujanaa, et al., 2021). Compared to existing techniques, application of CNNs in enhanced deep learning showed greater accuracy in emotion prediction (Mehendale, 2020). Introduced FERC using a two-part CNN, emphasizing emotions with 96% accuracy using an EV of length 24 values. FERC's single-level CNN approach differs from typical techniques, enhancing accuracy. In 1D-CNN achieved an accuracy of 97.44% (Sujanaa, et al., 2021). Showcased an algorithm with a training speed approximately 1.5 times faster than the contrast method, utilizing fewer iterations to attain an average recognition rate of 88.56% (Zhang, et al., 2019). The collective findings highlight diverse approaches and promising outcomes in facial emotion recognition.

REFERENCES

Abdulkareem, N. M., & Abdulazeez, A. M. (2021). Science and Business. *International Journal*, *5*(2), 128-142.

Abdullah, S. M. S., & Abdulazeez, A. M. (2021). Facial expression recognition based on deep learning convolution neural network: A review. *Journal of Soft Computing and Data Mining*, *2*(1), 53-65.

Affanni, A., & Chiorboli, G. (2015). Design and characterization of a real-time, wearable, endosomatic electrodermal system. *Measurement*, *75*, 111-121.

Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., & Ahmed, F. Y. (2021, June). Predicting university's students performance based on machine learning techniques. In *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)* (pp. 276-281). IEEE.

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics*, *10*(9), 1036.

Al Alshaikh, F. (2018). *The reliability and responsiveness of components of breathing pattern* (Doctoral dissertation, University of Southampton).

Alreshidi, A., & Ullah, M. (2020, February). Facial emotion recognition using hybrid features. In *Informatics* (Vol. 7, No. 1, p. 6). MDPI.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.

Amanoul, S. V., Abdulazeez, A. M., Zeebare, D. Q., & Ahmed, F. Y. (2021, June). Intrusion detection systems based on machine learning algorithms. In *2021 IEEE international conference on automatic control & intelligent systems (I2CACIS)* (pp. 282-287). IEEE.

Anderson, K., & McOwan, P. W. (2006). A real-time automated system for the recognition of human facial expressions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *36*(1), 96-105.

Arora, M., Kumar, M., & Garg, N. K. (2018). Facial emotion recognition system based on PCA and gradient features. *National Academy science letters*, *41*, 365-368.

Ashwin, T. S., Jose, J., Raghu, G., & Reddy, G. R. M. (2015, December). An e-learning system with multifacial emotion recognition using supervised machine learning. In *2015 IEEE seventh international conference on technology for education (T4E)* (pp. 23-26). IEEE.

Barrionuevo, C., Ierache, J., & Sattolo, I. (2020, October). Emotion Recognition Through Facial Expressions Using Supervised Learning with Logistic Regression. In *Argentine Congress of Computer Science* (pp. 233-246). Cham: Springer International Publishing.

Blackford, E. B., Estepp, J. R., & McDuff, D. J. (2018, February). Remote spectral measurements of the blood volume pulse with applications for imaging photoplethysmography. In *Optical diagnostics and sensing XVIII: toward point-of-care diagnostics* (Vol. 10501, pp. 192-199). SPIE.

Brienza, M., & Mecarelli, O. (2019). Neurophysiological basis of EEG. *Clinical electroencephalography*, 9-21.

Chen, C. Y., Huang, Y. K., & Cook, P. (2005, July). Visual/acoustic emotion recognition. In *2005 IEEE International Conference on Multimedia and Expo* (pp. 1468-1471). IEEE.

Chen, L., Wang, K., Li, M., Wu, M., Pedrycz, W., & Hirota, K. (2022). K-means clustering-based kernel canonical correlation analysis for multimodal emotion recognition in human–robot interaction. *IEEE Transactions on Industrial Electronics*, *70*(1), 1016-1024.

Chen, P., & Zhang, J. (2017). Performance comparison of machine learning algorithms for EEG-signal-based emotion recognition. In *Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part I 26* (pp. 208-216). Springer International Publishing.

Chen, X., Yang, X., Wang, M., & Zou, J. (2017, May). Convolution neural network for automatic facial expression recognition. In *2017 International conference on applied system innovation (ICASI)* (pp. 814-817). IEEE.

Colzato, L. S., Sellaro, R., & Beste, C. (2017). Darwin revisited: The vagus nerve is a causal element in controlling recognition of other's emotions. *Cortex*, *92*, 95-102.

Dave, J. (2023). Enhancing user experience in e-learning: Real-time emotional analysis and assessment. *International Journal Software Engineering and Computer Science (IJSECS)*, *3*(2), 57-64.

Deng, J., & Ren, F. (2021). A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing*.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440-460.

Graver, M. (2017). The Performance of Grief: Cicero, Stoicism, and the Public Eye. *Emotions in the Classical World: Methods, Approaches, and Directions*, 195-206.

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011, March). Emotion representation, analysis and synthesis in continuous space: A survey. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 827-834). IEEE.

Haji, S. H., Abdulazeez, A. M., Zeebaree, D. Q., Ahmed, F. Y., & Zebari, D. A. (2021, July). The impact of different data mining classification techniques in different datasets. In *2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA)* (pp. 1-6). IEEE.

Halfon, S., Doyran, M., Türkmen, B., Oktay, E. A., & Salah, A. A. (2021). Multimodal affect analysis of psychodynamic play therapy. *Psychotherapy Research*, *31*(3), 313-328.

HD, V., KR, A., & K, K. (2011). Emotion recognition from decision level fusion of visual and acoustic features using Hausdorff classifier. In *Computer Networks and Intelligent Computing: 5th International Conference on Information Processing, ICIP 2011, Bangalore, India, August 5-7, 2011. Proceedings* (pp. 601-610). Springer Berlin Heidelberg.

Hosseini, S. A., & Khalilzadeh, M. A. (2010, April). Emotional stress recognition system using EEG and psychophysiological signals: Using new labelling process of EEG signals in emotional stress state. In *2010 international conference on biomedical engineering and computer science* (pp. 1-6). IEEE.

Hudlicka, E., & Broekens, J. (2009, September). Foundations for modelling emotions in game characters: Modelling emotion effects on cognition. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (pp. 1-6). IEEE.

Ibrahim, A. F. T., Gannapathy, V. R., Chong, L. W., & Isa, I. S. M. (2016). Analysis of electromyography (EMG) signal for human arm muscle: A review. *Advanced Computer and Communication Engineering Technology: Proceedings of ICOCOE 2015*, 567-575

Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. *Pattern Recognition Letters*, *115*, 101-106.

Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective computing*, *5*(3), 327-339.

Kako, N. A., & Abdulazeez, A. M. (2022). Peripapillary Atrophy Segmentation and Classification Methodologies for Glaucoma Image Detection: A Review. *Current Medical Imaging*, *18*(11), 1140-1159.

Kashevnik, A., Lashkov, I., Axyonov, A., Ivanko, D., Ryumin, D., Kolchin, A., & Karpov, A. (2021). Multimodal corpus design for audio-visual speech recognition in vehicle cabin. *IEEE Access*, *9*, 34986-35003.

Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., & Prendinger, H. (2018). Deep learning for affective computing: Text-based emotion recognition in decision support. *Decision support systems*, *115*, 24-35.

Lasri, I., Solh, A. R., & El Belkacemi, M. (2019, October). Facial emotion recognition of students using convolutional neural network. In *2019 third international conference on intelligent computing in data sciences (ICDS)* (pp. 1-6). IEEE.

Lee, J. J., Uddin, M. Z., & Kim, T. S. (2008, August). Spatiotemporal human facial expression recognition using fisher independent component analysis and hidden markov model. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2546-2549). IEEE.

Liu, X., Cheng, X., & Lee, K. (2020). GA-SVM-based facial emotion recognition using facial geometric features. *IEEE Sensors Journal*, *21*(10), 11532-11542.

Loconsole, C., Miranda, C. R., Augusto, G., Frisoli, A., & Orvalho, V. (2014, January). Real-time emotion recognition novel method for geometrical facial features extraction. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)* (Vol. 1, pp. 378-385). IEEE.

Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), 446.

Mendl, M., Neville, V., & Paul, E. S. (2022). Bridging the gap: Human emotions and animal emotions. *Affective Science*, *3*(4), 703-712.

Mohsin, M. A., & Beltiukov, A. (2019, May). Summarizing emotions from text using Plutchik's wheel of emotions. In *7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)* (pp. 291-294). Atlantis Press.

Newell, S., & Marabelli, M. (2020). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. In *Strategic Information Management* (pp. 430-449). Routledge.

Ngoc, Q. T., Lee, S., & Song, B. C. (2020). Facial landmark-based emotion recognition via directed graph neural network. *Electronics*, *9*(5), 764.

Nijs, Y., PDEng, C. S., & Swerts, M. G. J. (2016). Children's lying behavior towards personified robots: an experimental study.

Pal, S., Mukhopadhyay, S., & Suryadevara, N. (2021). Development and progress in sensors and technologies for human emotion recognition. *Sensors*, *21*(16), 5554.

Park, K. S. (2023). *Humans and Electricity: Understanding Body Electricity and Applications*. Springer Nature.

Patil, V. K., & Pawar, V. R. (2022, August). How can emotions be classified with ecg sensors, ai techniques and iot setup?. In *2022 International Conference on Signal and Information Processing (IConSIP)* (pp. 1-6). IEEE.

Patil, V. K., & Pawar, V. R. (2022, August). How can emotions be classified with ecg sensors, ai techniques and iot setup?. In *2022 International Conference on Signal and Information Processing (IConSIP)* (pp. 1-6). IEEE.

Pazvantov, S., & Petrova, G. (2022, September). Development kit for recording and processing of EOG signals for eye tracking. In *AIP Conference Proceedings* (Vol. 2449, No. 1). AIP Publishing.

Ping, H. Y., Abdullah, L. N., Halin, A. A., & Sulaiman, P. S. (2013). A study of physiological signals-based emotion recognition systems. *Int. J. Comput. Technol*, *11*(1), 2189-2196.

Poulose, A., Reddy, C. S., Kim, J. H., & Han, D. S. (2021, August). Foreground Extraction Based Facial Emotion Recognition Using Deep Learning Xception Model. In *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)* (pp. 356-360). IEEE.

Rajesh, K. M., & Naveenkumar, M. (2016, December). A robust method for face recognition and face emotion detection system using support vector machines. In *2016 international conference on electrical, electronics, communication, computer and optimization techniques (ICEECCOT)* (pp. 1-5). IEEE.

Raut, N. (2018). Facial emotion recognition using machine learning.

Ruzainie, M. A., Xiao, Y., Uno, T., Ma, L., & Khorasani, K. (2021, October). A new facial expression recognition system using decision tree and deep neural networks. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)* (pp. 49-54). IEEE.

Sadeeq, H. T., & Abdulazeez, A. M. (2022). Giant trevally optimizer (GTO): A novel metaheuristic algorithm for global optimization and challenging engineering problems. *IEEE Access*, *10*, 121615-121640.

Sadeeq, H. T., & Abdulazeez, A. M. (2022, September). Improved Northern Goshawk Optimization Algorithm for Global Optimization. In *2022 4th International Conference on Advanced Science and Engineering (ICOASE)* (pp. 89-94). IEEE.

Sadeeq, H. T., & Abdulazeez, A. M. (2023). Metaheuristics: A Review of Algorithms. *International Journal of Online & Biomedical Engineering*, *19*(9).

Sadeeq, H. T., & Abdulazeez, A. M. (2023, September). Car side impact design optimization problem using giant trevally optimizer. In *Structures* (Vol. 55, pp. 39-45). Elsevier.

Saeed, J. N., Abdulazeez, A. M., & Ibrahim, D. A. (2022, September). 2D Facial Images Attractiveness Assessment Based on Transfer Learning of Deep Convolutional Neural Networks. In *2022 4th International Conference on Advanced Science and Engineering (ICOASE)* (pp. 13-18). IEEE.

Saeed, J. N., Abdulazeez, A. M., & Ibrahim, D. A. (2022, September). FIAC-Net: Facial image attractiveness classification based on light deep convolutional neural network. In *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1-6). IEEE.

Saeed, J. N., Abdulazeez, A. M., & Ibrahim, D. A. (2023). An Ensemble DCNNs-Based Regression Model for Automatic Facial Beauty Prediction and Analyzation. *Traitement du Signal*, *40*(1), 55.

Saeed, J. N., Abdulazeez, A. M., & Ibrahim, D. A. (2023). Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles. *Applied Sciences*, *13*(17), 9728.

Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J. S., Manjunath, R., & Jaiswal, S. (2023). Facial emotion recognition using convolutional neural networks. *Materials Today: Proceedings*, *80*, 3560-3564.

Seng, K. P., Ang, L. M., & Ooi, C. S. (2016). A combined rule-based & machine learning audio-visual emotion recognition approach. *IEEE Transactions on Affective Computing*, *9*(1), 3-13.

Shaees, S., Naeem, H., Arslan, M., Naeem, M. R., Ali, S. H., & Aldabbas, H. (2020, September). Facial emotion recognition using transfer learning. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)* (pp. 1-5). IEEE.

Somarathna, R., Bednarz, T., & Mohammadi, G. (2022). Virtual reality for emotion elicitation–a review. *IEEE Transactions on Affective Computing*.

Subudhiray, S., Palo, H. K., & Das, N. (2023). K-nearest neighbor based facial emotion recognition using effective features. *IAES International Journal of Artificial Intelligence*, *12*(1), 57.

Sujanaa, J., Palanivel, S., & Balasubramanian, M. (2021). Emotion recognition using support vector machine and one-dimensional convolutional neural network. *Multimedia Tools and Applications*, *80*, 27171-27185.

Tawari, A., & Trivedi, M. M. (2011, July). Audio visual cues in driver affect characterization: Issues and challenges in developing robust approaches. In *The 2011 International Joint Conference on Neural Networks* (pp. 2997-3002). IEEE.

Tiwari, P., & Veenadhari, S. (2022, November). An Efficient Classification Technique For Automatic Identification of Emotions Leading To Stress. In *2022 IEEE 6th Conference on Information and Communication Technology (CICT)* (pp. 1-5). IEEE.

Tripathi, S., Acharya, S., Sharma, R., Mittal, S., & Bhattacharya, S. (2017, February). Using deep and convolutional neural networks for accurate emotion classification on DEAP data. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 2, pp. 4746-4752).

Trnka, M., Darjaa, S., Ritomský, M., Sabo, R., Rusko, M., Schaper, M., & Stelkens-Kobsch, T. H. (2021). Mapping discrete emotions in the dimensional space: An acoustic approach. *Electronics*, *10*(23), 2950.

Vaidya, R., Nalavade, D., & Kale, K. V. (2022). Hyperspectral Imagery for Crop yield estimation in Precision Agriculture using Machine Learning Approaches: A review. *Int. J. Creat. Res. Thoughts*, *9*, a777-a789.

Verma, G., & Verma, H. (2020). Hybrid-deep learning model for emotion recognition using facial expressions. *The Review of Socionetwork Strategies*, *14*, 171-180.

Wu, C. H., Lin, J. C., & Wei, W. L. (2014). Survey on audiovisual emotion recognition: databases, features, and data fusion strategies. *APSIPA transactions on signal and information processing*, *3*, e12.

Xiang, X., & Tran, T. D. (2017). Linear disentangled representation learning for facial actions. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(12), 3539-3544.

Yang, Y., Wu, Q., Qiu, M., Wang, Y., & Chen, X. (2018, July). Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1-7). IEEE.

Z. Lv, F. Poiesi, Q. Dong, J. Lloret, and H. Song, "Deep learning for intelligent human–computer interaction," *Applied Sciences,* vol. 12, p. 11457, 2022.

Zhang, H., Jolfaei, A., & Alazab, M. (2019). A face emotion recognition method using convolutional neural network and image edge computing. *IEEE Access*, *7*, 159081-159089.

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, *59*, 103-126.

Zhao, S., Yao, H., Gao, Y., Ding, G., & Chua, T. S. (2016). Predicting personalized image emotion perceptions in social networks. *IEEE transactions on affective computing*, *9*(4), 526-540.

Zhou, X., Huang, X., Xu, B., & Wang, Y. (2004, December). Real-time facial expression recognition based on boosted embedded hidden Markov model. In *Third International Conference on Image and Graphics (ICIG'04)* (pp. 290-293). IEEE.

Zhu, Q. (1988, March). Pattern classification in dynamical environments: Tagged feature-class space and univariate sequential classifier. In *International Conference on Pattern Recognition* (pp. 517-526). Berlin, Heidelberg: Springer Berlin Heidelberg.