**Development and Assessment of a CNN-Based Framework for Facial Emotion Recognition Utilizing the VGG16 Architecture**

**1Lucky Ehizojie**

Department of Data Science Management, Rome Business School, Nigeria

lehizojie@acm.org

**2Abiodun Sojobi**

Centre for Information and TechnologyDepartment, University of Lagos, Nigeria

asojobi@unilag.edu.ng

\*Corresponding author: 1Lucky Ehizojie

**Abstract**

This study examines the role of AI in HCI, noting that accurate emotion discernment has become a vital area of concentration. Traditional methods strive with the complexity and multiformity of facial cues, often effecting mercurial performance across diverse demographics, cultural settings and one’s mood. The proposed system exploits the VGG16 CNN model, a deep learning architecture known for its efficiency in image scrutiny, to rectify these complications by identifying the facial muscles and their position. Hence, leveraging robust preprocessing methodologies, attribute extraction, and a systematic training framework employing the FER2013 and self-captured datasets, the model classifies seven basic emotions with enhanced accuracy which ordinarily, the ML algorithm deployed by diverse researchers in the past years had difficulty of getting optimally accurate results when extracting the facial emotion. The experimental results add to innovating facial emotion detection by proffering a credible and scalable model, catalyzing more compassionate and interactive human-machine interfaces.

Keywords: Convolutional Neural Network (CNN), facial emotion detection, VGG16 architecture, human-computer interaction, preprocessing methodologies, attribute extraction, FER2013, self-captured datasets.

**Significance Study**

This study concentrates on forming a framework that utilizes a special type of artificial intelligence, called a Convolutional Neural Network (CNN), to detect human emotions from facial signals. By leveraging the VGG16 model, a potent CNN architecture for image analysis, the investigation seeks to enhance how well computers can decode facial emotions, enabling machines to better discern and respond to human feelings. The research tackles some major issues in this field, such as the intricacy and diversity of human emotional responses and the limitation of traditional methods of recognizing emotions accurately and efficiently across diverse populations and cultural backgrounds. It supplies a solution by developing a CNN-based architecture trained on a carefully prepared dataset, facilitating it to classify emotions like happiness, sadness, anger, and surprise with greater accuracy. This study is significant because it provides a more robust base for emotion-aware systems that could be adopted in various industries such as healthcare, entertainment and education, where detecting emotions can improve interactions between humans and technology.

**1.0 Introduction and Background**

**1.1 Introduction**

In recent years, the domain of artificial intelligence (AI) has seen remarkable progress, particularly in computer vision and machine learning. Among the intriguing applications of AI is the recognition of human emotions through facial expressions, which holds significant potential across various domains such as human-computer interaction, healthcare, and marketing. The advancements in human-machine interaction are increasingly concentrated on facial analysis, as the face depicts the most expressive and communicative part of human behavior, essential for establishing effective dialogue between the two entities [1], [2].

In their scholarly endeavor, [3]describe emotions as delineations of basic aspects of human life, which comprise a complicated interaction of mental, physical, and behavioral reactions to external stimuli. They [3] further documented in their empirical research that faces serve as a potent medium of communication, visually conveying emotions and the motives individuals wish to express. Consequently, emotions play an essential role in human interaction. Facial expressions communicate a range of ideas and emotions, and Facial Emotion Recognition (FER) systems are designed to interpret these expressions accurately.

Facial emotion recognition systems aim to decode the intricate nuances of human emotions depicted through facial expressions, thereby enabling machines to perceive and respond to human emotions in a more intuitive and empathetic manner [4]. This capability has profound implications for enhancing user experience in interactive systems and fostering deeper connections between humans and machines [5], [6].

Despite advancements in AI technology, accurately recognizing facial emotions remains challenging due to the complexity and variability of human emotions, as well as the diversity of facial expressions across different individuals and cultures [7], [8]. Traditional approaches to facial emotion recognition often rely on handcrafted features and shallow learning algorithms, which may struggle to capture the intricate patterns present in facial expressions [9],[10]. However, with the emergence of deep learning techniques, particularly convolutional neural networks (CNNs), there has been a paradigm shift in the field, enabling more sophisticated and data-driven approaches to facial emotion recognition.

CNNs have transformed computer vision by leveraging hierarchical feature learning and automatic feature extraction from raw pixel data [11]. These networks can learn rich representations of facial features and expressions directly from image data, bypassing the need for manual feature engineering [12]. Architectures like VGG16 have demonstrated exceptional performance in image classification tasks, making them ideal for facial emotion recognition [13]. By harnessing the power of deep learning and CNNs, researchers can develop more robust and accurate facial emotion recognition systems that surpass the limitations of traditional methods [14].

The significance of facial emotion recognition extends beyond technological innovation; it has profound implications for various sectors, including healthcare, education, and entertainment [15]. In healthcare, emotion-aware systems can aid in diagnosing and treating mental health disorders by analyzing patients' facial expressions and detecting signs of distress or discomfort [16]. Similarly, in education, emotion-aware tutoring systems can adapt their teaching strategies based on students' emotional states, thereby enhancing learning outcomes and engagement [17]. As a matter of fact, [18] opine that Facial Emotion Recognition (FER) can be strategically applied in marketing to evaluate consumer responses to products and advertisements, hence, facilitating more informed, data-driven decisions. Furthermore, in the entertainment industry, emotion-aware interfaces can personalize user experiences in gaming and virtual reality environments, leading to more immersive and emotionally resonant interactions [19].

In light of these considerations, this study aims to contribute to the ongoing research efforts in facial emotion recognition by developing and evaluating a CNN-based face emotion recognition system using the VGG16 architecture. By leveraging deep learning techniques and state-of-the-art neural network architectures, the study seeks to advance the accuracy and efficiency of facial emotion recognition systems, thereby paving the way for more empathetic and intelligent human-machine interactions in diverse applications.

**1.2 Statement of the Problem**

Despite the advancements in facial emotion recognition technology, several challenges persist in achieving accurate and robust emotion detection from facial expressions. One of the primary challenges is the variability and complexity of human emotions, which can manifest differently across individuals and cultures [11]. This variability introduces ambiguity and uncertainty into the recognition process, making it difficult for algorithms to generalize effectively across diverse populations. Additionally, the presence of occlusions, such as facial hair, accessories, or partial facial expressions, further complicates the recognition task, leading to decreased accuracy and reliability in real-world applications [7].

Another significant challenge in facial emotion recognition is the limited availability of annotated datasets, particularly for specific demographics or cultural groups. Without sufficient and diverse training data, machine learning models may exhibit biases or limitations in their ability to recognize emotions accurately across different populations [4]. Moreover, the lack of standardized evaluation protocols and benchmarks makes it challenging to compare the performance of different facial emotion recognition systems objectively. As a result, researchers face difficulties in assessing the generalizability and effectiveness of their algorithms in real-world settings, hindering the adoption of facial emotion recognition technology in practical applications [9],[20]. Addressing these challenges is crucial for advancing the field of facial emotion recognition and unlocking its full potential in various domains.

**1.3 Study Objectives**

The objectives of the study are:

- Develop and evaluate a convolutional neural network (CNN)-based facial emotion recognition system using the VGG16 architecture.

- Enhance the accuracy and efficiency of facial emotion recognition, particularly in diverse populations and cultural backgrounds.

- Contribute to the advancement of facial emotion recognition technology by exploring innovative approaches to overcome existing challenges.

- Improve the overall performance of emotion detection from facial expressions [11].

**1.4 Research Hypothesis**

The research hypothesis formulated for this study is:

- H0: There is no significant difference in the accuracy and efficiency of facial emotion recognition between the CNN-based system developed in this study and existing methods.

- H1: The CNN-based facial emotion recognition system developed in this study demonstrates superior accuracy and efficiency compared to existing methods, leading to improved emotion detection and classification.

**1.5 Scope and Delimitations of the Study**

The scope of this study encompasses the development and assessment of a CNN-based facial emotion recognition system using the VGG16 architecture. The research focuses on recognizing seven basic emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral) from facial expressions captured in images. However, the study is limited to the analysis of grayscale images and does not consider other modalities such as video or depth data. Additionally, the evaluation of the system's performance is limited to controlled experimental settings and may not fully capture its real-world effectiveness [7].

**1.6 Definition of Key Terms**

**Facial Emotion Recognition**: The process of identifying human emotions based on facial expressions, and classifying them into different emotional states such as happiness, sadness, anger, fear, disgust, surprise, and neutrality. That is, with the aid of machine learning algorithms.

**Convolutional Neural Network (CNN)**: A deep learning architecture designed for processing and analyzing visual data, particularly images.

**Accuracy**: The measure of how often the facial emotion recognition system correctly identifies and classifies emotions from facial expressions.

**Efficiency**: The ability of the system to achieve accurate emotion recognition within a reasonable time frame.

**Human-Machine Interaction**: The interaction between humans and machines, facilitated by AI technologies, such as facial emotion recognition systems.

**Convolutional Neural Network (CNN)**: A class of deep neural networks commonly used for image recognition and classification tasks, characterized by their ability to automatically learn spatial hierarchies of features.

**VGG16 Architecture**: A specific CNN architecture consisting of 16 layers, developed by the Visual Geometry Group (VGG) at the University of Oxford, known for its simplicity and effectiveness in image classification tasks.

**Preprocessing**: The initial step in data preparation, involving various techniques such as image transformation, scaling, cropping, and filtering to enhance the quality of input data before feeding it into machine learning algorithms.

**Convolutional Neural Network (CNN)**: A deep learning architecture designed for pattern recognition in visual data, particularly suited for tasks such as image classification and object detection.

**Precision**: The ratio of correctly predicted positive instances to the total predicted positive instances, indicating the model's ability to avoid false positives.

**Recall**: The ratio of correctly predicted positive instances to the total actual positive instances, indicating the model's ability to capture all positive instances.

**F1 Score**: The harmonic mean of precision and recall, providing a balance between the two metrics.

**ROC AUC**: Receiver Operating Characteristic Area Under the Curve, a measure of the model's ability to distinguish between classes, particularly relevant in binary classification tasks.

**2.0 Literature Review**

This section examines the extant literature on the research topic, its variables, and the reviews of related empirical studies, explaining the main concepts of the research work and the theory that forms the basis for the study.

**2.1 Conceptual Framework**

Facial Emotion Recognition (FER) is a burgeoning field that has garnered significant attention due to its wide-ranging applications in various domains such as healthcare, human-computer interaction, and security systems. Fundamentally, Facial Emotion Recognition (FER) involves the automated identification and assessment of human emotions by interpreting facial expressions [20], [18]. In their study, [7] provide a comprehensive overview of FER, defining it as the process of automatically identifying and categorizing human emotions from facial expressions captured in images or videos. The authors highlight the importance of FER in enabling machines to understand human emotions, thereby facilitating more natural and intuitive human-machine interactions

Facial expression recognition (FER) is a vital form of visual data that aids in interpreting an individual’s emotional state. It provides a wealth of social cues to observers, including insights into a person’s focus, emotions, motivations, and intentions, making it a powerful tool for non-verbal communication. AI-based FER systems can be strategically deployed in high-traffic areas such as bus terminals, railway stations, airports, and stadiums, where they assist security personnel in detecting potential threats efficiently [21].

Khajuria, O., Kumar, R., & Gupta M. (2023) [2]assert thatFacial expressions offer a reliable means for identifying human emotions across diverse contexts. However, manual facial expression recognition (FER) presents challenges, as interpretations can be influenced by an individual's state of mind [14]. Theoretical Foundations of Facial Emotion Recognition delve into psychological theories that underpin FER systems. Notably, Paul Ekman's Facial Action Coding System (FACS) serves as a cornerstone in understanding facial expressions. According to Oguntuase et al. (2021) [11], FACS provides a systematic framework for describing facial muscle movements associated with different emotional states, forming the basis for many FER algorithms. Additionally, the basic emotions theory proposed by Ekman suggests that there are universal facial expressions corresponding to basic emotions such as happiness, sadness, anger, fear, disgust, and surprise, which further informs the design and development of FER systems. Components of Facial Emotion Recognition (FER) Systems encompass various stages, including image preprocessing, feature extraction, and classification. Adewole (2018) [4] explains that image preprocessing involves techniques such as normalization, resizing, and noise reduction to enhance the quality of facial images and improve the performance of FER algorithms. Feature extraction techniques, such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), aim to capture discriminative facial features that are indicative of different emotional states [11].Finally, classification algorithms, often implemented using machine learning or deep learning approaches, are trained to classify facial expressions into predefined emotion categories [23].

Challenges and Limitations in Facial Emotion Recognition highlight the complexities inherent in accurately recognizing and interpreting facial expressions. In their study, Adebayo et al. (2017) [7] identify challenges such as variability in facial expressions across individuals and cultures, occlusions, and insufficient annotated datasets. These challenges pose significant obstacles to the development of robust and reliable FER systems, underscoring the need for innovative solutions and advancements in the field.

**2.3 Theoretical Framework**

Deep Learning and Convolutional Neural Networks (CNNs) have emerged as powerful tools in FER research, enabling automated feature learning and classification directly from raw pixel data. Jhadi et al., (2024) [24] emphasize the effectiveness of CNNs in capturing spatial dependencies in facial images and extracting hierarchical features, leading to improved performance in emotion recognition tasks. CNN architectures, such as VGG16 and ResNet, have been widely adopted in FER systems, demonstrating superior performance compared to traditional machine learning algorithms.

Psychological Models of Emotion provide insights into the cognitive processes involved in facial emotion recognition. According to Sun et al. (2023) [25], the dimensional model of emotion posits that emotions can be represented along dimensions such as valence and arousal, influencing the perception and interpretation of facial expressions. Moreover, cultural differences in emotion expression and perception highlight the need for culturally sensitive FER systems that account for diverse cultural norms and expressions (Sun et al., 2023) [25].

Cognitive Processing of Facial Expressions explores the cognitive mechanisms underlying the perception and interpretation of facial expressions. Sun et al. (2023) [25] discusses how cognitive processes such as attention, memory, and interpretation bias influence individuals' ability to recognize and attribute emotions to facial expressions. Understanding these cognitive processes is crucial for designing FER systems that mimic human-like emotion recognition capabilities.

Social and Cultural Factors in Facial Emotion Recognition examine the influence of social and cultural contexts on emotion expression and perception. Adebayo et al. (2017) [7] highlight the importance of considering cultural differences in facial expression norms and the impact of social factors such as gender and age on emotion recognition accuracy. Cross-cultural studies have revealed variations in emotion perception across different cultural groups, emphasizing the need for culturally informed FER algorithms.

**2.4 Empirical Review**

State-of-the-Art Facial Emotion Recognition Systems showcase recent advancements and innovations in FER technology. A study by Olabiyisi et al. (2018) [17] reviews state-of-the-art FER systems based on deep learning approaches, highlighting their improved performance and robustness compared to traditional methods. These systems leverage large-scale datasets, such as the Facial Expression Recognition 2013 (FER2013) dataset, and employ sophisticated CNN architectures to achieve high accuracy in emotion recognition tasks.

Evaluation Metrics and Benchmark Datasets are essential for assessing the performance of FER systems. Jhadi et al (2024) [24] discuss common evaluation metrics such as accuracy, precision, recall, and F1-score, which are used to measure the effectiveness of FER algorithms. Benchmark datasets, including FER2013, CK+, and MMI, provide standardized datasets for training and evaluating FER systems, enabling fair comparisons between different approaches.

Applications of Facial Emotion Recognition span various domains, including healthcare, education, and human-computer interaction. Adewole (2018) [4] explores the potential applications of FER in healthcare, such as mental health diagnosis and treatment, where emotion-aware systems can aid clinicians in assessing patients' emotional states and monitoring their progress. Similarly, in education, FER technology can be used to develop emotion-aware tutoring systems that adapt learning content based on students' emotional responses, leading to improved learning outcomes and engagement.

Comparative Studies and Performance Analysis evaluate the effectiveness of different FER algorithms and techniques. Jhadi et al (2024) [24] conduct a comparative study of CNN architectures for FER, comparing the performance of VGG16, ResNet, and Inception networks on benchmark datasets. The results demonstrate variations in performance across different architectures, highlighting the importance of selecting appropriate models and parameters for FER tasks.

**3.0 System Analysis, Methodology and Design**

This section begins by giving an overview of a face emotion classification system. It then looks at some of the methods used in face emotions classification system. It then goes on to explain the methodology used in the implementation.

**3.1 Typical Face Emotion Classification System**

Face Emotion Recognition systems typically use image preprocessing and feature extraction followed by training on selected training architectures [26]. The end result of training is the generation of a model capable of assigning emotion categories to newly provided image examples. The steps that are required to perform face emotion recognition are shown in figure 3.1.

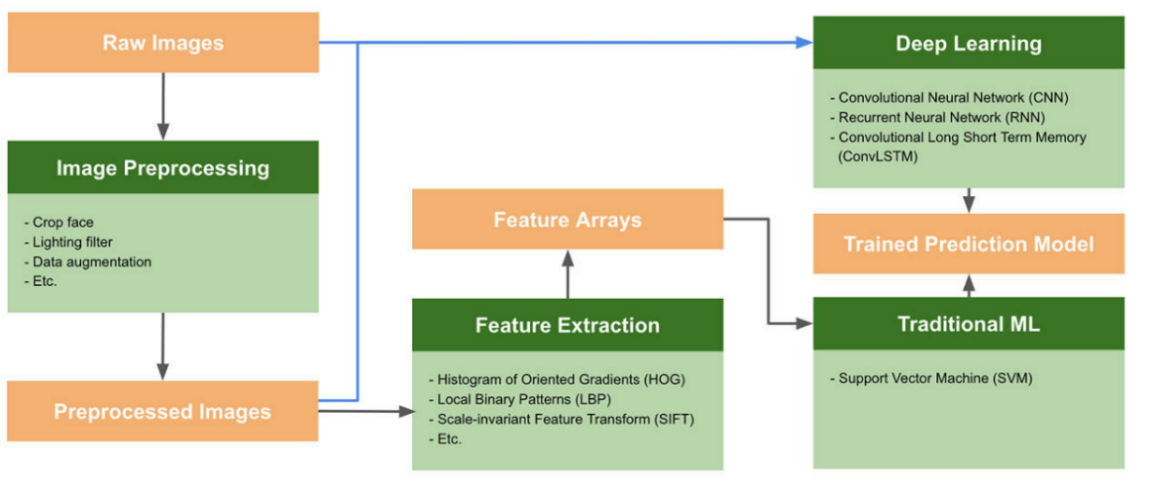


Figure 3.1: Block diagram of a typical face emotion recognition system.

The preprocessing stage of the image may involve image transformations, such as scaling, cropping, or filtering. This is also used to accentuate contextual details about the image, such as cropping an image to remove a background. It can also be used to improve a dataset, such as creating several versions from an original image with varying cropping or transformations.

The extraction stage of the feature goes even further in finding the more descriptive parts of an image. This also involves seeking details that might be most representative of a certain class, such as outlines, textures, or colors.

The training phase takes place according to the specified training architecture, which specifies the combinations of layers in the neural network that are fed into one another. Architectures must be planned for training, taking into account the composition of the extraction function and the preprocessing phases of the image.

The training architecture and model is very important for the general training cycle because it is the deciding factor for prediction accuracy. Something very important to remember is the optimizer value, the optimizer value is also a key factor as it follows the idea that the higher the value the faster the training process but also the lower the accuracy and at the same time the lower the optimizer, the slower the training process but then the higher the probability for the predictive model to be accurate. Finding a perfect value to strike a balance between accuracy and time is therefore also a key mathematical ability needed to provide a near-perfect system. Following the general method for forecasts, no matter the activation function chosen for any specific system, one will find a particular order.

**3.2. Face Database**

For credibility and for research to be easily accepted a standardized database is used. In this study we will be using the standard FER2013 and self-captured databases.

**3.2.1. FER2013 Face Database**

The FER2013 database consists of 48x48 pixel grayscale images of faces. The faces have been saved and stored so that the face is more or less centered and covers about the same amount of space in each image consisting facial details and expressions of both male and female [27].

**3.2.2. Self-Captured Face Database**

The Self Captured database consists of images of black faces. The images are gotten from snapshots of different people both male and female. The faces will be saved and stored so that the face consisting facial details and expressions will make up the dataset. The dataset contains 3841 images, with 514 ‘anger’ images, 533 ‘disgust’ images, 349 ‘fear’ images, 793‘happiness’ images, 614 ‘sadness’ images, 455 ‘surprise’ images and 583 ‘neutral’ images. The dataset contains seven folders, with each folder containing a category of emotion. The 3841 images were divided into two sub folders, the training folder and the testing folder. The training folder contains 3072 images with each image categorized under the proper emotion.

**3.3. Preprocessing**

Preprocessing applies to all the transformations on the raw data before they are fed into the machine learning or deep learning algorithm. Training a CNN on raw images, would possibly lead to bad results in classification [28]. In essence, it helps in planning the data for the processing phase to improve the results.

**3.4. Processing**

Current face analysis systems, which are able to determine the emotional state of an individual from the facial expressions analysis, operate in three basic phases, as defined by [26]:

1. Face detection phase,
2. Feature extraction,
3. Classification of emotions according to the selected model.

**3.4.1 Feature Extraction**

Feature Extraction helps to reduce the number of features in a dataset by generating new features (and then discarding the original features) from existing ones. This new reduced feature set will then be able to summarize much of the details in the original feature set. In this way, the original features can be summarized from a combination of the original set [29].

Working with datasets with hundreds (or even thousands) with features is becoming nowadays very popular. When the number of features is close (or even greater!) to the number of observations contained in a dataset then this will most likely lead to an overfitting Machine Learning model. To avoid this type of problem, either the regularization or the reduction of dimensionality (Feature Extraction) techniques is required. In Machine Learning, a dataset's dimensionality is equal to the number of variables used to describe it.

**3.4.3 Classifier**

A classifier is an algorithm that maps the input data to a specific category [30], [31] described the effective image classification as follows:

* Acquire our input, which is a training dataset that consists of *N* images, each labeled with one of the different classes.
* Then use this training set to train a classifier to learn what every one of the classes looks like.
* We evaluate the quality of the classifier by asking it to predict new set of images that it has never seen before. We will then check the true labels of these images and compare to the ones predicted by the classifier. This process is called testing.

**3.5 Model Development Tools**

These are the tools that were used to develop our model, test the model and validate the model.

**3.5.1 Python**

Python is a programming language based on high-level, interpreted Object Oriented. It is an effective, highly useful language that focuses on the Rapid Development of applications (RAD) and Don’t Repeat Yourself (DRY). It also works beautifully as a linking language; it links different components together. It is now becoming one of the fastest developing languages because of Python's ease of learning, scalability, and adaptability [32]. Python 3.9 is used in this project.

**3.5.2 Jupyter Notebook Environment**

Jupyter Notebook is an extremely powerful tool for the interactive development and presentation of data science projects. A notebook incorporates the code and its output into one single document that blends narrative text, visualizations, mathematical equations, and other rich media. The intuitive plan encourages iterative, rapid growth, making notebooks a gradually popular choice at the heart of data scientists. Jupyter provides enough test and development environment. But a web service would be needed to utilize and access the model even from an external app. The brainwave project works on TensorFlow, it is one of the most widely used systems for AI operations. Microsoft Cognitive Toolkit (CNTK), a Microsoft product that is another common deep learning platform, is being worked on for Brainwave support (Chollet 2018),

**3.5.3 Libraries**

These are the libraries that were used during the development of the proposed model.

**3.5.3.1 Keras**

Keras is used in a deep learning framework for Python which provides a suitable way of de-fining and training any form of deep learning models. Keras was first developed for researchers and scientists to provide fast research.

Keras has the following key features:

1. It allows the implementation of both CPU and GPU with the same code.
2. It is a user-friendly API which allows scientist to make Deep learning models prototype easily.
3. Recurrent networks (sequence processing) and convolutional networks (for computer vision) are supported by Keras build-in libraries.
4. It supports random network architectures: multi-output models or multi-input, model sharing, layer sharing, and so on. This indicates that Keras is suitable for building any type of deep learning model, from a generative adversarial net-work to a neural Turing machine.
5. Keras has been distributed under the permission of the MIT licence, which can be freely used in commercial projects. It’s well-suited with any Python version from 2.7 to 3.6 (as of mid-2017) [32] and more as time goes on.

**3.5.3.2 Sklearn**

Scikit-learn offers a range of supervised, unsupervised learning and reinforcement algorithms via a reliable interface in Python. It is licensed under the permission of a simplified BSD license and is distributed under encouraging academic, Linux distributions, and commercial use.

The library is built upon the SciPy (Scientific Python) which must be installed before you can use sci-kit-learn. This stack includes:

1. NumPy: Base n-dimensional array package.
2. SciPy: Fundamental library for scientific computing.
3. Matplotlib: Comprehensive 2D/3D plotting.
4. IPython: Enhanced interactive console.
5. Sympy: Symbolic mathematics.
6. Pandas: Data structures and analysis.

Extensions or modules for SciPy care are conventionally named SciKits. As such, the module provides learning algorithms and is named sci-kit-learn [32].

**3.5.3.3 NumPy**

NumPy is the fundamental library used for scientific computing with Python. Among its functionalities are:

1. A powerful N-dimensional array object.
2. Sophisticated (broadcasting) functions.
3. Tools for integrating C/C++ and Fortran code.
4. Useful linear algebra, Fourier transform, and random number capabilities.

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be de-fined. This allows NumPy to integrate with a wide variety of databases seamlessly and speedily [33].

**3.5.3.4 Pandas**

Pandas Is among python packages that provide flexible, fast, and expressive data structures intended to make working with labelled or relational data both intuitive and easy. It aims to provide a high-level fundamental building block for performing real-world, practical data analysis in Python

Here are just a few of the things that pandas do well:

1. Easy handling of missing data in floating point and as non-floating-point data.
2. Size mutability: columns can be inserted and deleted from Data Frame and higher dimensional objects.
3. Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, Data-Frame, etc. automatically align the data for you in computations.
4. Powerful, flexible group by functionality to perform split-apply-combine operations on data sets for aggregating and transforming data.
5. Make it easy to convert ragged, differently indexed data in other Python and NumPy data structures into Data Frame objects.
6. Intelligent label-based slicing, fancy indexing, and subsetting of large data sets.
7. Intuitive merging and joining data sets.
8. Flexible reshaping and pivoting of data sets.
9. Hierarchical labeling of axes (possible to have multiple labels per tick).
10. Robust IO tools for loading data from flat files (CSV and delimited), Excel files, databases, and saving/loading data from the ultrafast HDF5 format.
11. Time series-specific functionality: date range generation and frequency conversion, moving window statistics, date shifting, and lagging [34].

**3.5.3.5 Matplotlib**

Matplotlib is a library in anaconda used for making 2D plots of arrays in Python, even though it was initially emulated from MATLAB graphics commands, MATLAB is not needed and can be used in an object-oriented and pythonic way. Although matplotlib is coded primarily in Python, it uses NumPy and other extension codes to offer good performance with a small or large array. Matplotlib is designed with the idea that a user is to create a simple plot with only a few commands, if the user wants to see a histogram of his data should not need to call methods, instantiating objects, set properties, and so; it should just work [35].

**3.5.3.6 Seaborn**

Seaborn is a matplotlib-based Python data visualization library. It provides a high-level GUI for drawing attractive statistical graphics and giving information. Seaborn is an open source, BSD-licensed Python library that provides high-level APIs to access data using Python's programming language [36].

**3.5.4 Requirement Analysis**

Specifications apply to the various features required to be incorporated in the project to get the project's desired result. The project specifications are classified into three groups: main requirements recommended requirements and optional specifications.

**3.5.4.1 Essential Requirements:**

These are the minimal and primary features required to be implemented in the project. They are:

1. Obtain and prepare a dataset to be used as input to the model.
2. Perform visualization and statistical analysis of the obtained data.
3. Adopt a Machine Learning model to carry out the prediction task from scratch.
4. Visualize and track the performance of our trained model to avoid overfitting and underfitting.
5. Evaluate the performance of the model on unseen test data.

**3.5.4.2 Recommended Requirements:**

These are major features that were implemented in the project. They are:

1. Obtain a minimum accuracy of 80 percent on classification.
2. Handle imbalanced data if any.

**3.6 Methodology**

This face emotion recognition system will be implemented based on CNN AND DNN principles. It first starts from creation of database of all the face images, which it can be termed as face gallery. The images in this research work are gotten from two standardized face gallery called Self Captured face database and FER2013.

Figure 3.2 and figure 3.3 shows the seven (7) different category of face labels we wish to derive from our data base, (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).



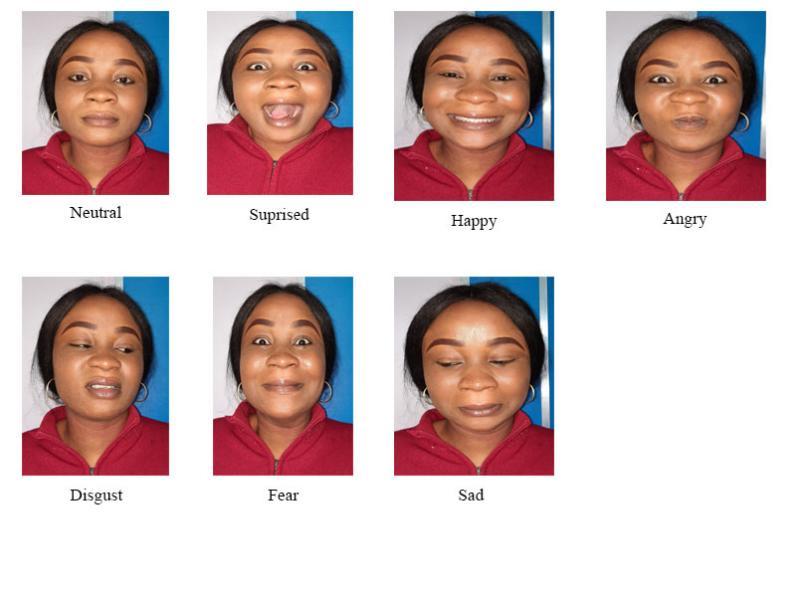
Figure 3.2: Categories of Images (FER2013 DATASET)

Figure3.3: Categories of images

For training and research purposes the face database is split into two sub-sets. The dataset will be fed into a data matrix and then partitioned into a test set and training set. Along with the labels the train and test images will be loaded and stored in variables x train, y train, x test and y test. The training set is used to learn how to discriminate between the different faces and is used to test the algorithm using the test set. The data will be divided into two parts, 80% designed for training and the remaining 20% for validation. The Jupyter notebook environment will be used in this research work.

The network will be trained for 20 epochs with a batch size of 64 and all required models will be imported to model the data and layers. Then the model will be compiled using Adam optimizer to minimize our loss function, the loss type used is categorical cross entropy because it is a job of multi-class classification and the metrics will be defined as accuracy, confusion matrix and other available metric values obtained from the analysis. Visualization of the layers will be done using the summary function.

Neural Networks are the basis of deep learning. These Networks will be connected in the fashion of feed forward. The input will be multiplied by the weight and added, the bias will also add, and then the activation function applied will ReLU as the activation function used in this research.

In summary, we will create a convolutional network using the deep learning libraries specifically Keras and Tensor flow, we split the data into training and test data and we then load the data to a VGG 16 network. After training, we get the metrics results of the model.

**4.0 System Implementation and Analysis of Results**

This paper highlights the results and discussions of the models used. The dataset was validated using the split sample validation technique of train/test split with 80% train set and 20% test set. Performance Evaluation using Accuracy, Precision, Recall, AUC, was also presented in this chapter.

**4.1 Data Splitting**

To train a model, the dataset must usually be divided into at least two sections: training and testing data. The model is based on a collection of validations and a set of training and testing. The model's output on the validation package is used to tune the model. Splitting the dataset into three parts: preparation, validation, and testing is a good idea. The dataset was divided into 80 percent (3072) and 20 percent (768) for training and testing, respectively; the 768 samples were used to determine the model's performance 50 percent (384) and 50 percent (384) for validation and testing, respectively. [36] stated that the fewer training data you have, the more likely your work will be overfitted. The data consists of 48x48 pixel grayscale images of faces. The task was to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

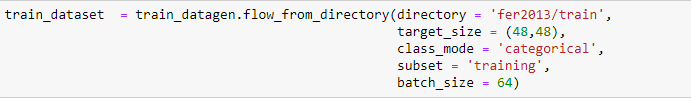


Figure 4.1: Code showing training data being loaded.

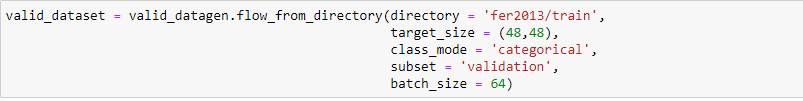


Figure 4.2: Code showing validation data being loaded.

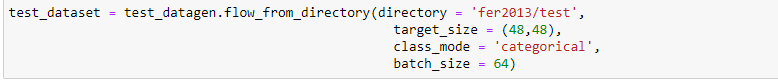


Figure 4.3: Code showing testing data being loaded.

**4.2 VGG 16 Model Architecture, Training and Testing Performance**

VGG 16 was used for the training and testing, the model's precise number of input layers, which includes two hidden levels and one output layer. In this architecture, the sequential DNN model is employed, which is in the form of a linear stack of layers. Compiling a method to specify loss functions, optimisers, and metrics is also standard.

**4.2.1 VGG 16 Architecture**

The VGG 16 model has a sequence of layers which includes the first layer which is the input lay-er having 10 neurons, an input shape of 10 since our input variables are 10 and ReLU as the activation function, the model had 2 hidden layers with 11 neurons with ReLU activation function and the final layers which had only 1 neuron that predicts either 0 or 1. We create a function call model. Compile which accumulates optimizer, loss function, and metrics, our model used adam as its optimizer, binary cross entropy for loss function. The model also had a function fit that fits our parameters for the training (x-train and Y-train), specification mini batch as 64, the period it takes for the model to train the dataset to 100 times and validation parameters (x-val and Y-val).

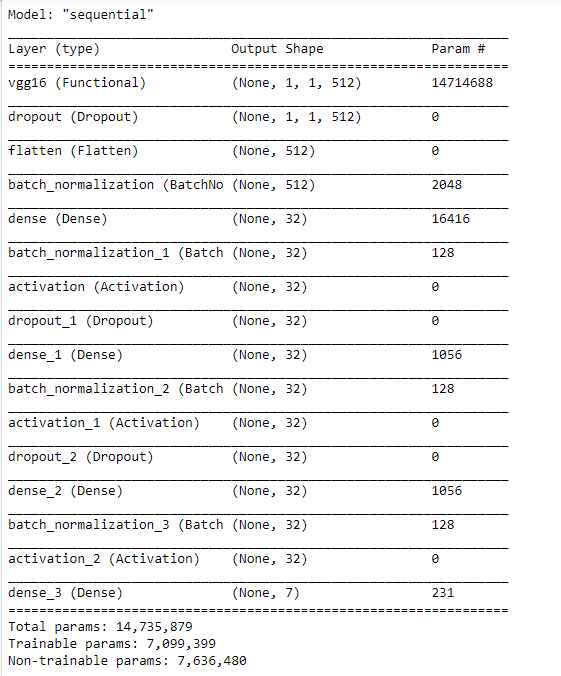


Figure 4.4: Model Architecture

Figure 4.4 is a summary of the model. These are convolutional stacks, there are three fully connected layers. With the last layer serving as an output layer with SoftMax activation. A flatten layer was added before the first fully connected layer. Dropout layers were also used following dense layers for regularization. The size of the output layer can be adjusted to match the no. of classes in any recognition problem. I have kept the output layer to match 7 class categories following the proposed classes of emotion recognition.

**4.4.2 Training**

For the training of the model, Adam optimizer was used as an optimizer and categorical cross entropy was since the model was to deal with multiclass problems. The training was set to run for 20 epochs with verbose set to 1. After the training, the model was saved into “model.h5” file.

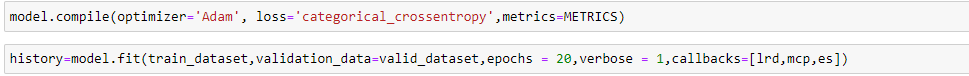


Figure 4.5: Adam Optimizer and Categorical cross entropy

The figure below shows the result of the first 11 epochs, with accuracy, precision, recall, AUC, F1 score been record throughout the training.

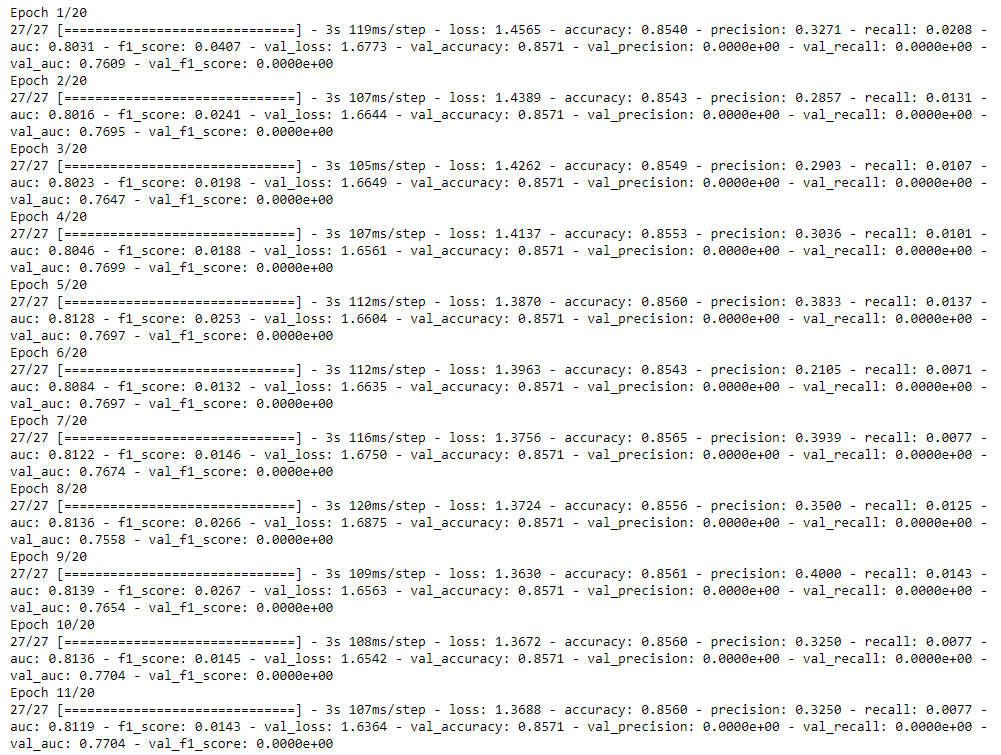


Figure 4.6: First 11 epochs

The figure below shows the result of the last 9 epochs, with accuracy, precision, recall, AUC, F1 score been record throughout the training.

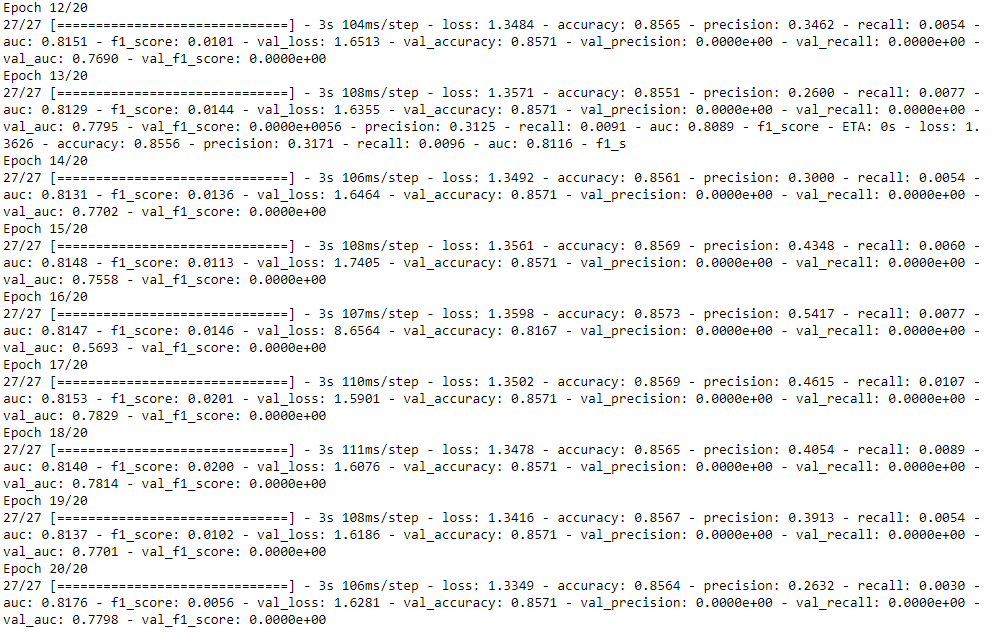


Figure 4.7: Last 9 epochs

During the training process, two variables are frequently reported, as shown in Figures 4.6 and 4.7: the network's loss over the training data and its accuracy over the same training data. The figures illustrate that the validation loss diminishes as the loss increases. The validation accuracy improves as backpropagation and validation progress. Our model was able to learn from the pool of datasets provided, as shown in both graphs.

## 4.3 Performance Evaluation

Table 4.7: Performance Evaluation

|  |  |
| --- | --- |
| Measures | Results |
| Accuracy | 85% |
| Precision | 50% |
| F1 Score | 50% |
| ROC AUC | 85% |

Table 4.7 displayed the model performance, revealing that the model was successful in classifying facial emotions into the seven categories with an overall prediction accuracy of 85 percent when evaluated on unseen data. The model was shown to be hyper-intelligent to anticipate emotions with an accuracy of 85%. Precision: 50%, F1 Score: 50%, and ROC Score: 85% are other measures used to evaluate model performance.

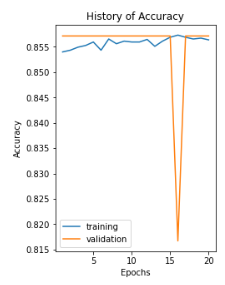


Figure 4.8: *Graph of Accuracy for Training and Validation Set*

Figure 4.8 depicts the level of accuracy of our model as it was iterated 20 times and based on the results in both the training and validation sets, the accuracy increases in both sets and with nearly the same level of advancement, with the validation set even outperforming the training set, indicating the effectiveness of our model and no overfitting.

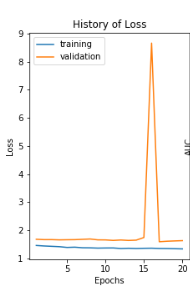


Figure 4.9: *Graph of Function Loss for Training and Validation Set*

Figure 4.9 depicts the accuracy loss of our model on both the validation and training sets as it is iterated 20 times and based on the results of the figure. The graph reveals no overfitting because there is an improvement because the loss accuracy reduces in both the training and testing sets at about the same rate.

**Conclusion**

This study successfully created and assessed a Convolutional Neural Network (CNN) framework utilizing the VGG16 architecture for accurate facial emotion identification. Regardless of conventional constraints, such as insufficient data and the intricate nature of facial features, the proposed system capitalizes on a combination of the Facial Expression Recognition 2013 dataset (FER2013) and self-captured datasets, coupled with fortified preprocessing and feature extraction methodologies, to build a dependable emotion classifier.

The technique used optimized image preprocessing and a tuned VGG16 architecture, enhancing the algorithm’s ability to extract meaningful features from facial images and categorize them effectively into appropriate emotional classes while reducing overfitting. These experimental results reveal the potency of deep learning in innovating facial emotion identification, adding to more compassionate and accurate AI-powered systems in healthcare, education and human-computer interaction. This study lays the groundwork for subsequent research to nuance and amplify CNN-based emotion discernment, enabling enriched user experiences in emotionally responsive systems.

**Recommendations for Future Research**

This scholarly endeavor like other similar studies has its strength and limitations. Hence, the research thereby offers the following suggestions for further investigations to improve its core values in these areas utilizing CNNs and similar architectures:

1 Tackle the challenge of dataset diversity; obtaining and integrating larger, more real-world datasets across varied population and cultural backdrops to enhance model performance.

2 Leverage multimodal datasets that synthesize video, depth, infrared imaging to improve the model’s ability to recognize emotional expression and control visual obstructions such as facial hair or accessories, that may block expressions.

3 Optime CNN architectures and experiment with simplified models to improve latency and mitigate computational demands by allowing deployment to be feasible on mobile and smart devices.

4 Develop real-time emotion recognition systems that ensure user privacy and maintain high accuracy.

5 Study how to reduce model bias and improve interpretability in order to increase user trust and clarify the way CNN-based architectures make emotion predictions across varied populations.

**Originality/value**

To the best of the knowledge of the authors, this study is novel, underscoring the efficacy of deep learning in propagating facial emotion recognition. It’s an innovative contribution to more empathetic and accurate AI-application systems that explores new approaches to overcome existing challenges and improves the overall performance of emotion recognition from facial features.

**Conflicts of Interest**

The authors declare no conflicts of interest, Including technical assistance, donations or organizational aids from individuals or organizations. The authors also declare that there is no financial, commercial, or other affiliations that may be perceived as potential conflicts of interest by the academic community.

**References**

[1] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In British Machine Vision Conference.

[2] Khajuria, O., Kumar, R., & Gupta, M. (2023). Facial Emotion Recognition using CNN and VGG-16. International Conference on Inventive Computation Technologies (ICICT). Pp. 472-477, DOI: 10.1109/ICICT57646.2023.10133972.

[3] Desai, K., Patel, H. N., Dashlaniya, D., Desai, P., Shah, M., & Goel, P. (2024). A Transfer Learning Framework for Facial Emotion Recognition: Leveraging Pre-Trained Convolutional Neural Networks. *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*. Pp. 663-667, DOI: 10.1109/InCACCT61598.2024.10551190.

[4] Adewole, K. S. (2018). Advances in facial expression recognition: A survey. Brain Informatics, 5(1), 1-13.

[5] Oladimeji, T.O., Charles, Ayo., & Adewumi, S. E. (2019). Review on Insider Threat Detection Techniques. Journal of Physics Conference Series. DOI: [10.1088/1742- 6596/1299/1/012046](http://dx.doi.org/10.1088/1742-6596/1299/1/012046).

[6]Pagar, M. C., Savale, T.K., Patil, P.U., Mutha, N., Wanjare, H. F., & Mogare, S. B. (2024).

Human-Computer Interaction in Smart Living Environments: Challenges and Opportunities. Journal of Informatics Education and Research. Vol 4 Issue.

[7] Adebayo, A., Afolabi, O., & Salau, T. (2017). A review on facial expression recognition techniques. International Journal of Computer Science and Information Security, 15(10), 103-116.

[8] Salawu, S. O., Adeniyi, D. A., & Olabiyisi, S. O. (2020). A Comprehensive Survey on Facial Emotion Recognition Techniques. Journal of Computer Science and Its Application, 27(2), 12-25.

[9] Oyebode, O., & Akinade, O. O. (2020). A review on facial expression recognition using deep learning techniques. In 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 1-7). IEEE.

[10] Li, L., Mu, X., Li, S., & Peng, H. (2020). A review of face recognition technology. IEEE Access, 8, 139110–139120. <https://doi.org/10.1109/ACCESS.2020.3011028>.

[11] Oguntuase, J. A., Oluwatobi, A. A., & Olugbara, O. O. (2021). Performance Evaluation of Convolutional Neural Networks in Facial Expression Recognition. Journal of Engineering and Applied Sciences, 16(4), 3532-3541.

[12] Adegoke, M. E., Adewale, A. S., & Misra, S. (2018). Facial expression recognition using deep learning: A survey. In 2018 IEEE 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-5). IEEE.

[13] Olaniyi, A., Okediran, O., & Sogbesan, T. (2022). Application of deep learning algorithms for facial expression recognition: a systematic review. In Proceedings of the 2022 5th International Conference on Soft Computing & Machine Intelligence (pp. 1-6).

[14] Onaolapo, J. O., Awelewa, A. A., & Shittu, K. K. (2016). Facial Emotion Recognition Using Convolutional Neural Networks. International Journal of Computer Applications, 134(5), 29-34.

[15] Igbe, A., Akinbulire, T. O., & Idowu, T. B. (2016). A survey of facial expression recognition techniques. International Journal of Computer Science.

[16] Fayemi, P., Wanieck, K., Zollfrank, C., Maranzana, N., & Aoussat, A. (2017). Biomimetics: process, tools and practice. *Bioinspiration & Biomimetics, 12*. DOI:[10.1088/1748- 3190/12/1/011002](https://doi.org/10.1088/1748-3190%2F12%2F1%2F011002).

[17]Olabiyisi, S. O., Adelaja, A. O., & Aremu, O. S. (2018). Face Recognition System: A Deep Learning Approach. Journal of Computer Science and Its Application, 25(1), 1-12.

[18] Punuri, S. B., Kuanar, S. K., & Mishra, T.K. (2023). Facial Emotion Recognition in Unconstrained Environments through Rank-Based Ensemble of Deep Learning Models using 1-Cycle Policy. International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM). Pp. 1-8. DOI: 10.1109/IC-RVITM60032.2023.10435159.

[19] Xi, N., Buruk, O. O., Chen, B., J, S., & Hamari, J. (2024). Wearable gaming technology: A study on the relationships between wearable features and gameful experiences. International Journal of Human-Computer Studies. Volume 181. <https://doi.org/10.1016/j.ijhcs.2023.103157>.

[20] Shi, Y., Zhang, Z., Huang, K., Ma, W., & Tu, S. (2020). Human-computer interaction based on face feature localization. Journal of Visual Communication and Image Representation, 70, 102740. <https://doi.org/10.1016/j.jvcir.2019.102740>

[21] Kanwar, N., & Verma, B. (2023). From methods to datasets: a detailed study on facial emotion recognition. Springer Nature. DOI: [10.1007/s10489-023-05052-y](http://dx.doi.org/10.1007/s10489-023-05052-y).

[22] Sun, Y., Wang, X., & Tang, X. (2014). Deep Learning Face Representation by Joint Identification-Verification. In Advances in Neural Information Processing Systems.

[23] Kaliszyk, C., Chollet, F., & Szegedy, C., (2017). Holstep: A Machine Learning Dataset for Higher Order Logic Theorem Proving. arXiv preprint arXiv:1703.00426. https://doi.org/10.48550/arXiv.1703.00426.

[24] Jhadi, K., Tiwari, N., & Chawla, M. (2024). Review of Machine and Deep Learning Techniques for Expression based Facial Emotion Recognition. *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp.1-6.

[25] Sun, J., Dong, T., & Liu, P., (2023). Holistic processing and visual characteristics of regulated and spontaneous expressions. Journal of Vision. Vol.23, 6. DOI:https://doi.org/10.1167/jov.23.3.6.

[26] Madgin. M., Benko, L., & Koprda, S. (2019). A Case Study of Facial Emotion Classification Using Affdex. Sensor Applications on Face Analysis. [https://www.mdpi.com/1424- 8220/19/9/2140](https://www.mdpi.com/1424-%098220/19/9/2140). DOI: [10.3390/s19092140](http://dx.doi.org/10.3390/s19092140).

[27] Tang, J., Zhou, X., & Zheng J. (2019). Design of Intelligent classroom facial recognition based on Deep Learning. Journal of Physics Conference Series. DOI: [10.1088/1742- 6596/1168/2/022043](http://dx.doi.org/10.1088/1742-6596/1168/2/022043).

[28] Pal, K. K., & Sudeep, K. S. (2016). Preprocessing for image classification by convolutional neural networks. IEEE International Conference on Recent Trends in Electronics, Information \& Communication Technology (RTEICT)}. DOI:10.1109/RTEICT.2016.7808140.

[29] Baulin, F. B. & Buryi, E. V. (2019). Performance Estimate of Range Profile Feature Extraction for the Case of Defined Viewing Aspect by Means of Fisher Score. *PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring.* Pp. 2922-2926, DOI: 10.1109/PIERS-Spring46901.2019.9017824.

[30] Murugappan, M., Ramachandran, N., & Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. J. Biomedical Science and Engineering. doi:10.4236/jbise.2010.34054.

[31] Setiawan, D., Gunawan, G., & Arif, Z. (2024). Application of nearest neighbor interpolation method and naïve bayes classifier for identification of bespectacled faces. Journal of Intelligent Decision Support System (IDSS).

[32] Bisong, E. (2019). Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Springer Nature. <https://link.springer.com/book/10.1007/978-1-4842-4470-8>.

[33] Numpy Community Survey 2020 Results. (2020). <https://numpy.org/surveys/NumPy_usersurvey_2020_report.pdf>.

[34] Wes McKinney & PyData Development Team. (2014). Pandas: Powerful Python Data Analysis Toolkit Release 0.14.0. <https://pandas.pydata.org/pandas-> docs/version/0.14.0/pandas.pdf.

[35] Hunter, J., Dale, D., Firing, E., & Droettboom, M. (2016). Matplotlib development team, [En línea]. https://matplotlib. org/.[Último acceso: 28 11 2021].

[36] Chollet, F. (2018) Deep Learning with Python. Manning Publications Co.