

Deep Learning Vs. Traditional Machine Learning For Facial Emotion Recognition In Psychiatric Applications

Rohini Narayan Pawar¹, Dr. Vaishali Chavan²

¹Assistant Professor (Research Scholar) Dr. G. Y. Pathrikar College of Computer Science & IT MGM University, Chatrapati Sambhaji Nagar, India.

²Assistant Professor (Ph.D. Guide) Dr. G. Y. Pathrikar College of Computer Science & IT MGM University, Chatrapati Sambhaji Nagar, India.

Abstract

The early diagnosis of mental illnesses including stress, worry, and depression may be greatly aided by facial expression identification. In order to classify six fundamental emotions (anger, disgust, fear, happiness, sadness, and surprise), this study assesses the performance of five machine learning models: K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and a general Neural Network (NN). Using the PyTorch and scikit-learn frameworks, experiments were carried out on a system with an Intel i7-12700F CPU, 32 GB RAM, and an RTX 3070 GPU. Accuracy, precision, recall, and F1-score were used to compare models. With an overall accuracy of 95% and high F1-scores across all emotion classes, the results show that the MLP model performs better than all other approaches, proving its capacity to capture intricate, non-linear emotional patterns. While a traditional NN performed worse (81%), KNN, RF, and SVM obtained modest accuracies (62–84%). While simpler models like KNN are still competitive for small datasets, deep neural networks perform better for larger or more homogeneous data, according to analysis across several datasets. The results encourage the inclusion of AI-driven technologies into real-time mental health monitoring and intervention systems by highlighting the efficacy of deep learning for accurate emotion recognition.

1. Introduction

This work uses a number of machine learning and deep learning classification techniques to predict face emotions. A method called facial emotion recognition (FER) is used to examine emotions in a variety of media, including pictures and movies. 55% of emotion detection is attributed to facial expressions, a type of nonverbal communication that offers clues about human feelings [1]. In human-computer interaction and psychology, the interpretation of emotional emotions has grown in importance. The study of emotion recognition's technological applications is being intensively researched because of its significant importance [2]. Machine learning is a discipline of Artificial Intelligence (AI) that focuses on constructing systems capable of learning from data without the need for explicit programming by humans [3]. Conversely, Deep Learning (DL) is a branch of Machine Learning. learning that uses a set of nonlinear transformation functions organised in a deep and layered framework to create algorithms to represent data abstractions at a higher level [4]. Several Machine Learning (ML) classification algorithms were used in the modelling experiment, including: (1) Support Vector Machine (SVM), which uses a hypothesis space in the form of linear functions on high-dimensional features, trained using optimization-based algorithms to produce classification or regression models [5]; (2) K-Nearest Neighbour (KNN), which can perform classification without requiring modelling of data distribution, making it suitable for datasets of varying sizes [6]; and (3) Random Forest (RF), an ensemble learning algorithm that builds predictive models using a bootstrap (bagging) approach [7]. Convolutional Neural Networks (CNN), a Deep Learning technique, are used to recognise expressions with great accuracy by learning complex features [4]. Because Python supports multiple

programming paradigms, such as object-oriented, functional, and rule-based programming, it is frequently used for data analysis [8]. The aim of this study is to conduct a FER modelling experiment and compare the performance of ML models such as SVM, KNN, RF, and the deep learning model CNN in detecting emotions based on facial expressions. It is anticipated that the resultant FER model will function as a substitute instrument to support psychological therapy and sustain dynamic mental health [9]. The Indonesian Psychiatric Association (PDSKJI) reported in March 2022 that 82.5% of self-assessment service users had psychiatric issues, up from 80.4% in 2020 and 70.7% in 2019. Anxiety, depression, and psychological trauma were among these psychological problems. According to this research, mental health issues in Indonesia are growing more concerning and need for greater care and intervention. One of the most important factors influencing both individual and society well-being is mental health. Nonetheless, evidence indicates that mental illnesses are becoming more common in Indonesia every year. About 10% of Indonesians suffer from psychiatric problems, according to the Ministry of Health of the Republic of Indonesia (Kemenkes). In the adult population, the prevalence of emotional mental problems increased significantly from 6.0% in 2013 to 9.8% in 2018 [10]. According to the Basic Health Research (Riskesdas) conducted in 2018, about 19 million Indonesians over the age of 15 have emotional psychological illnesses. Furthermore, according to a survey conducted by Khamelia and Terry [11], the COVID-19 pandemic also played a role in the increase in psychiatric issues.

2. Literature Review

Facial Emotion Recognition (FER) is described in the research journal by Yousif Khaireddin and Zhuofa Chen as an essential component of human-computer interaction, especially in clinical practice and behavioural assessment, which aids in the detection of mental health conditions like depression, anxiety, and mood disorders. However, the variety of human faces and image alterations, including variations in facial positions and lighting conditions, continue to pose challenges to the accuracy and robustness of FER models. Convolutional Neural Networks (CNNs) and other deep learning-based models have demonstrated significant promise because of their computational efficiency and capacity for automatic feature extraction. Without using any extra training data, a recent study that used the VGGNet architecture obtained the maximum accuracy of 73.28% on the FER2013 dataset. Hyperparameter tuning and the investigation of optimisation techniques were credited with the model's effectiveness. Because of their high accuracy, CNNs have a lot of potential for incorporation into mental health detection systems, offering automated insights into an individual's emotional state that can facilitate early diagnosis and more effective mental health treatment [12]. Simcock et al.'s study investigates the connection between early adolescents' mental health and their capacity to recognise facial emotions. Forty twelve-year-old children participated in the study. They completed a face expression detection test and a mental health questionnaire to evaluate symptoms of anxiety, depression, and somatization. The findings demonstrated that teenagers with more psychiatric problems tended to process emotions biasedly, especially when it came to negative expressions like fear and rage. Although more research is still needed to understand the neural mechanisms causing this bias, it may point to underlying mental health difficulties from an early age. These results demonstrate how facial expression recognition can be used as a tool to identify early indicators of mental illnesses, which is crucial for quicker teen diagnosis and treatment [13]. Facial Emotion Recognition (FER) employing artificial intelligence (AI) technology has demonstrated tremendous potential in supporting medical professionals, notably in the realm of mental and emotional health. Previous research has shown that FER systems that employ deep learning and machine learning methods, including Convolutional

Neural Networks (CNNs), can reliably identify facial expressions. This information can then be utilised as additional data for the diagnosis of mental illnesses or emotional stress. Healthcare practitioners can obtain real-time insights into patients' emotional states by incorporating FER technology into medical systems, which is essential for early intervention and clinical decision-making. For instance, with properly adjusted parameters, a CNN-based model trained using the FER2013 dataset was able to achieve high accuracy in emotion categorisation, demonstrating how this technology might be further modified to assist medical personnel in comprehending and responding to patients' circumstances. These developments present excellent prospects for the creation of AI-based mental health systems that are more responsive and inclusive [14].

3. Methods of Research

This study uses machine learning and deep learning approaches to investigate Face Emotion Recognition (FER) in a methodical manner. The procedure comprises: Figure 1 illustrates these sections: 1) Introduction; 2) Literature Review; 3) Methodology; 4) Experiments and Results; 5) Discussion; and 6) Conclusion.

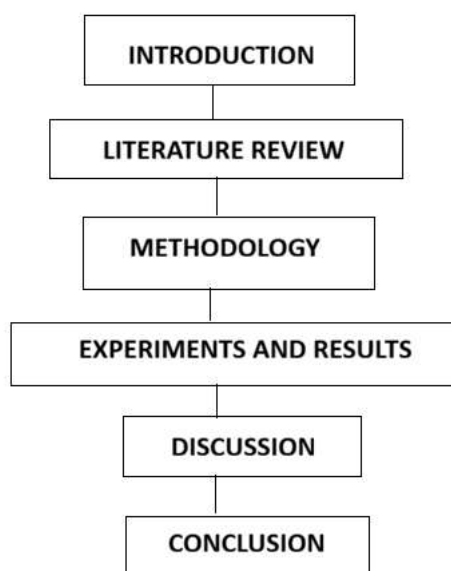


Figure 1: Research Process for Recognising Face Emotions

The performance of multiple machines is compared in this study. In the context of facial expression recognition (FER), learning methods such as Support Vector Machine (SVM), KNearest Neighbours (KNN), and Random Forest are combined with a Deep Learning model (Neural Network or NN). Preprocessing and data collection are the first steps in the process. Python's pickle module, which manages the serialisation (pickling) and deserialisation (unpickling) of Python objects for effective storage and reuse, is used to first convert facial expression photos to grayscale, resize them to 48 by 48 pixels, and save them [15]. Pixel value normalisation is used to improve model performance and lessen the impact of unimportant features [16]. Machine learning models can be trained more easily by using label encoding, which transforms categorical emotion labels into numerical values [17]. Principal Component Analysis (PCA) is applied for dimensionality reduction to reduce complexity, speed up training, and minimise the danger of overfitting [18]. In order to guarantee consistent feature scaling, which is crucial for numerous

learning methods, feature standardisation is also carried out using StandardScaler [19]. The dataset is divided into training and testing sets after preprocessing. Grid Search is used to optimise machine learning models in order to find the optimal combinations of hyperparameters. Metrics like accuracy and classification reports are used to evaluate the model. For example, the performance of the KNN model is evaluated to determine how well it classifies face expressions. Finally, the best-performing model is utilised to predict emotions on the test dataset. The results' visualisation aids in illustrating each model's ability to identify facial expressions. The purpose of this study is to provide light on how well Deep Learning models and conventional Machine Learning techniques perform in FER challenges. Figure 2 depicts the framework for facial emotion recognition.

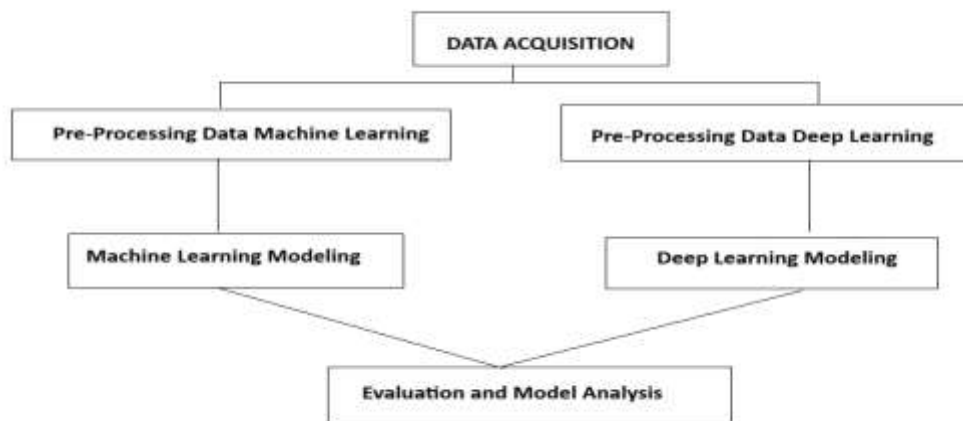


Figure 2: Framework for Recognising Face Emotions

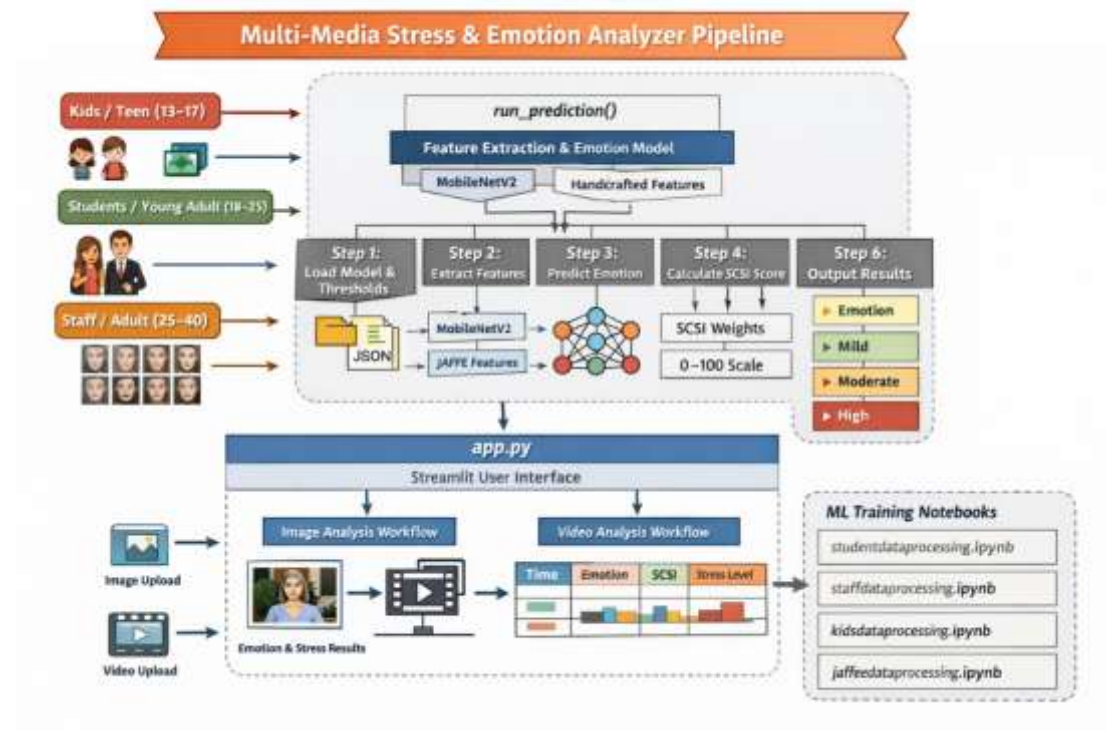


Figure 3. Architecture Diagram

The illustration shows a Multi-Media Stress and Emotion Analyser system that uses pictures and videos to determine a person's emotional state and stress level. It is made to be used by people of all ages and uses an interactive application to deliver numerical and visual results.

The system first takes into account several user categories, including adults, students or young people, and children and teenagers. The system employs age-specific models and thresholds to ensure more accurate stress interpretation because each group experiences and expresses stress differently. Users upload an image or a video to provide input. While movies are processed frame-by-frame to record changes in tension and emotion over time, images are analysed as a single frame. The system extracts features from the media after it is uploaded. It employs manually created facial characteristics based on datasets of facial expressions in addition to using a deep learning model (MobileNetV2) to extract high-level facial features. These characteristics work together to better correctly depict face emotions. An emotion prediction model receives the extracted features and uses them to categorise a person's facial expressions into emotional groups like happy, sad, furious, or neutral. This emotion prediction serves as the basis for stress analysis. The technology determines a Stress Score (SCSI) upon emotion recognition. The severity of stress is represented by a numerical scale that ranges from 0 to 100 after the projected emotions have been weighted and mapped. The system classifies the stress level as light, moderate, or high based on this score. Professionals and users can better comprehend the data thanks to these categories. All processing is incorporated into a Streamlit-based user interface, where users may upload media and receive results. For videos, the interface displays graphs showing emotion and stress changes over time. Lastly, machine learning training notebooks that are used to preprocess data and train models for various datasets and age groups complement the system. This ensures the system remains accurate, adaptable, and scalable.

A. The dataset

The Face Emotion Recognition Dataset, which Jonathan Oheix made available on the Kaggle platform, is used in this study. The dataset is made up of photos of people's faces that have been categorised into seven different emotions: Anger, joy, sadness, surprise, neutrality, fear, and disgust. The intricacy of the emotion recognition process is increased by the fact that each category contains photos with different lighting, positions, and people.

B. Preparing Information for Machine Learning

1. Data Loading

To make it easier to handle and process, the dataset is arranged into folders labelled by emotion categories (such as angry, joyful, and sad). Python modules like os and Pillow, which enable effective processing of image files, are used to load the data. The `convert('L')` method is used to convert all photos to grayscale in order to maximise processing and concentrate on key facial traits. This method removes extraneous color information while maintaining key texture characteristics crucial for emotion perception. Each image is reduced to 48 by 48 pixels to guarantee consistency and standardise the deep learning model's input dimensions. By improving consistency and streamlining the preprocessing pipeline, this step improves the dataset's suitability for training. Additionally, automatic labelling is carried out by extracting class names from the folder names, guaranteeing

accurate image categorisation and enabling smooth integration into the machine learning workflow.

2. Transformation and Cleaning of Data

The data must be processed once the photos are loaded in order for the machine learning model to be trained. Pixel normalisation, label encoding, dimensionality reduction, and data standardisation are all part of this process.

3. Data Splitting for Training and Assessing Machine Learning

The training data and the testing data are the two primary components of the processed dataset at this point. The split is carried out using `train_test_split`, which enables users to split the dataset into training and testing sets at random. Because it guarantees that training and assessment are conducted on different sets of data, avoiding overfitting and producing accurate performance measurements, this function is highly helpful in machine learning applications. This function contains numerous optional options, such as `test_size` to determine the fraction of the test data. For example, `test_size=0.3` signifies that 30% of the data is used for testing and the remaining data is utilised for training. `Random_state`, another crucial argument, ensures reproducibility by producing the same data split each time the function is executed with a particular seed value. Debugging, comparing models, and repeating experiments all depend on this consistency [20].

- **Data Split:** The dataset is divided into 20% testing data and 80% training data. The `train_test_split` function from the scikit-learn library is used to randomly split this data, making sure that the training and testing sets do not overlap.
- **Stratification:** Because the dataset is stratified, both the training and testing data maintain the same percentage of each emotion type. This avoids class imbalance, which might lower the performance and accuracy of the model.
- **Model Data Preparation:** A variety of models, including machine learning techniques (SVM, KNN, RF), are compatible with the training and testing data. The ensuing stages of model training and evaluation subsequently make use of the split dataset.

C. Modelling for Machine Learning

The pre-processed data is now used to train the machine learning model. There are three primary algorithms used:

KNearest Neighbours (KNN), Random Forest (RF), and Support Vector Machine (SVM). `GridSearchCV`, a technique for determining the optimal parameter combination through cross-validation, is used to optimise each algorithm. The user creates a dictionary with the list of values to be tested and the hyperparameter names as keys. `GridSearchCV` will test every possible combination of these parameters on various training data subsets, evaluate the outcomes, and refine the model. Based on the dataset utilised, this optimisation seeks to create a model that performs the best in categorising facial emotions [21].

D. Assessment of Machine Learning

Visualisation is utilised to evaluate the performance of face expression categorisation in the final evaluation stage. Five grayscale test images are randomly displayed with true and predicted labels for visual comparison. A confusion matrix highlights common misclassifications by summarising accurate and inaccurate predictions for each emotion. For every emotion class, a ROC curve is also presented using binary labels; the model's capacity to discriminate between emotions is indicated by the AUC values. A baseline for

random predictions is supplied. These visual aids offer information about the model's precision, dependability, and potential improvement areas.

E. Deep Learning-Based Face Emotion Recognition

a. Data Loading

The dataset is loaded using ImageFolder, which assigns labels based on folder names, then arranged into folders according to face expression class. The model is better able to distinguish between distinct emotions because to its structure. A Data Loader is used to load data with a batch size of 64, and shuffling is turned on to enhance generalisation. Memory pinning improves CPU-to-GPU data flow, while four parallel workers speed up loading. Together, these improvements increase training effectiveness and model performance on unknown data.

b. Transformation of Data

Before being fed into the Neural Network, the input images undergo a variety of preprocessing steps to increase feature extraction and assure consistency across the entire dataset. Converting photos into a single-channel format in order to exclude colour information and enable the model to concentrate on facial texture and structure is the first stage. This keeps important information required for emotion identification while streamlining calculation.

c. Data Augmentation

The dataset goes through several data augmentation stages to improve model resilience and generalisation. Images are converted to grayscale and scaled to 48×48 pixels to guarantee consistent input proportions. To replicate natural changes in facial positions, geometric augmentations include random affine transformations with rotations of ± 10 degrees, translations up to 10%, and scaling between 0.9 and 1.1. While brightness and contrast modifications within a range of 0.8 to 1.2 assist the model adapt to various lighting situations and lessen sensitivity to changes in illumination, random horizontal flipping with a 50% chance provides variation in orientation and symmetry.

4. Model Assessment

Following training, the model's performance and generalisation are assessed using the validation dataset. For consistent predictions, the model is in evaluation mode, which disables dropout and stabilises batch normalisation. To save memory and expedite inference, gradient computations are disabled; mixed precision inference further improves performance.

By comparing predicted labels—obtained using argmax on output probabilities—with ground truth labels, validation accuracy is calculated. While per-class accuracy shows performance across many emotion categories and aids in identifying class imbalances or frequent misclassifications, overall accuracy shows the percentage of correctly categorised samples.

A high validation accuracy suggests robust generalization, but a big difference between training and validation results may suggest overfitting, demanding changes like hyperparameter tuning, data augmentation, or regularisation.

5. Model Training Methodology

With the model in training mode (`model.train()`) to enable dropout and update batch normalisation statistics for improved generalisation, the training procedure lasted 100 epochs.

To produce predictions, each batch passed through a forward pass, and the Cross-Entropy loss was computed in relation to the true labels. While retaining stability, mixed precision training (`torch.amp.autocast ("cuda")`) increased computational efficiency. During the backwards pass, gradients were computed and updated using the AdamW optimizer, with gradient clipping applied to prevent inflating gradients. Gradient scaling (`torch.amp.GradScaler`) improved mixed precision convergence. After each batch, the OneCycleLR scheduler dynamically modified the learning rate to encourage balanced training and prevent early convergence. Per-class accuracy was tracked to identify misclassification or class imbalance. Accuracy was regularly assessed. When validation accuracy plateaued, early stopping with a patience of 15 epochs saved needless training. After enhanced validation performance, the optimal model checkpoint was retained. Every five epochs, cache clearing took place to control GPU memory. Following training, the model was returned to its optimal condition, ready for deployment or additional refinement, and loss curves were produced to show learning progress.

6. Final Training and Assessment Outcomes

The `train_part` function, which manages the training loop and includes mixed precision, gradient clipping, learning rate adjustment, and early halting to encourage robust learning, is where model training begins. The `evaluate_model` function is used to assess the model on the validation set following training. In order to stabilise batch normalisation and disable dropout, the model enters evaluation mode (`model.eval()`) during evaluation. Inference operates under `torch.no_grad()` to save memory usage and boost productivity by omitting gradient computations. While per-class accuracy offers insights into performance across various facial expressions and highlights any class-specific strengths or shortcomings, validation accuracy is determined as the proportion of right predictions overall.

4. Results and Discussion

Using PyTorch and scikit-learn packages, experiments were carried out on a system with an Intel i7-12700F CPU, 32 GB RAM, and an RTX 3070 GPU. K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN), a deep learning model, are compared in this study for facial emotion recognition. The objective is to evaluate accuracy, precision, recall, and F1-score in order to determine the best model for early identification of psychiatric illnesses such as stress, anxiety, and depression.

NNs are excellent candidates for this task since they automatically learn complicated patterns, in contrast to classical models that rely on manual feature extraction. Early identification is crucial because mental health illnesses are on the rise. Real-time monitoring and intervention can be enhanced by incorporating emotion detection into healthcare via wearables, telemedicine, and digital apps. In order to support psychological evaluation and treatment, this study provides insights into AI-driven mental health technologies.

4.1 KNN (K-Nearest Neighbors) Emotion

The precision, recall, F1-score, and support of a KNN-based emotion classification model across six emotion classes are summarised in the table. Overall, the model's classification performance is moderate, with an accuracy of 0.62 on 37 samples. Surprise has the highest recall (1.00) and F1-score (0.92) of all the emotions, indicating that the model accurately detects almost all instances of surprise. Fear likewise performs well; its recall (0.57) indicates that it may miss some fear samples, but its perfect precision (1.00) indicates that the model is always correct when it predicts fear. With F1-scores between

0.55 and 0.67, Disgust and Anger exhibit balanced but moderate performance. On the other hand, the model finds it more difficult to classify happiness and sadness in particular, as evidenced by their poorer precision, recall, and F1-scores. Although there is obviously need for improvement, especially for emotions with subtler or overlapping features like sadness and happiness, the macro and weighted averages both produce an F1-score of 0.62, indicating that the model's performance is generally constant across classes.

Table 1. KNN (K-Nearest Neighbors) Emotion

	precision	recall	f1-score	support
Anger	0.60	0.50	0.55	6
Disgust	0.67	0.67	0.67	6
Fear	1.00	0.57	0.73	7
Happiness	0.40	0.67	0.50	6
Sadness	0.40	0.33	0.36	6
Surprise	0.86	1.00	0.92	6
accuracy			0.62	37
macro avg	0.65	0.62	0.62	37
weighted avg	0.66	0.62	0.62	37

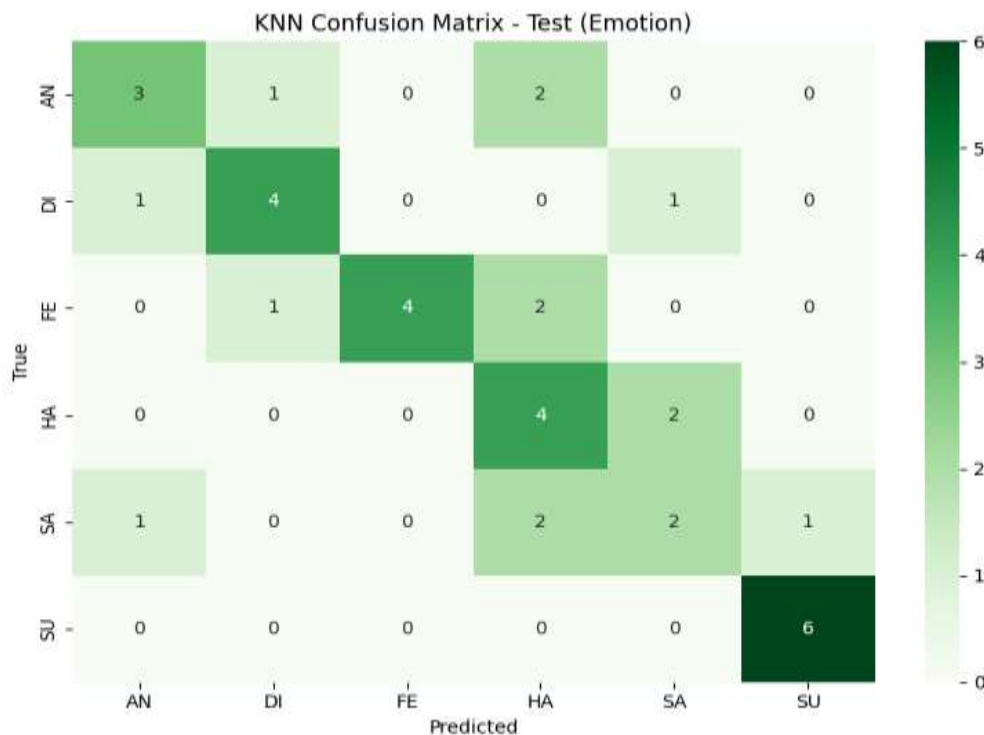


Figure 1. KNN (K-Nearest Neighbors) Emotion (KNN Confusion_matrix_Test)

4.2 MLP (Multilayer Perceptron) emotion

The Multilayer Perceptron (MLP) emotion classification model's performance is displayed in the table, which displays excellent and reliable outcomes for each of the six emotion classes. On 37 samples, the model obtains an overall accuracy of 0.95, demonstrating exceptional classification abilities. With 100% precision, recall, and F1-score, emotions like anger, fear, and surprise are all fully classified, meaning that every incidence of these emotions is accurately detected with no false positives or false negatives. With F1-scores of 0.92, which indicate only slight misclassifications, disgust and happiness also do remarkably well. Sadness shows significantly lower but still high performance, with balanced precision and recall of 0.83, showing some misunderstanding with other emotions but generally reliable detection. Despite some differences in class support, the model performs consistently well across all classes, as seen by the macro and weighted averages of 0.95. Overall, this table shows that the MLP model performs far better than more straightforward methods, catching intricate patterns in emotional characteristics and offering reliable and precise emotion recognition.

Table 2. MLP (Multilayer Perceptron) emotion

	Precision	recall	f1-score	support
Anger	1.00	1.00	1.00	7

Disgust	1.00	0.86	0.92	7
Fear	1.00	1.00	1.00	6
Happiness	0.86	1.00	0.92	6
Sadness	0.83	0.83	0.83	6
Surprise	1.00	1.00	1.00	5
Accuracy			0.95	37
macro avg	0.95	0.95	0.95	37
weighted avg	0.95	0.95	0.95	37

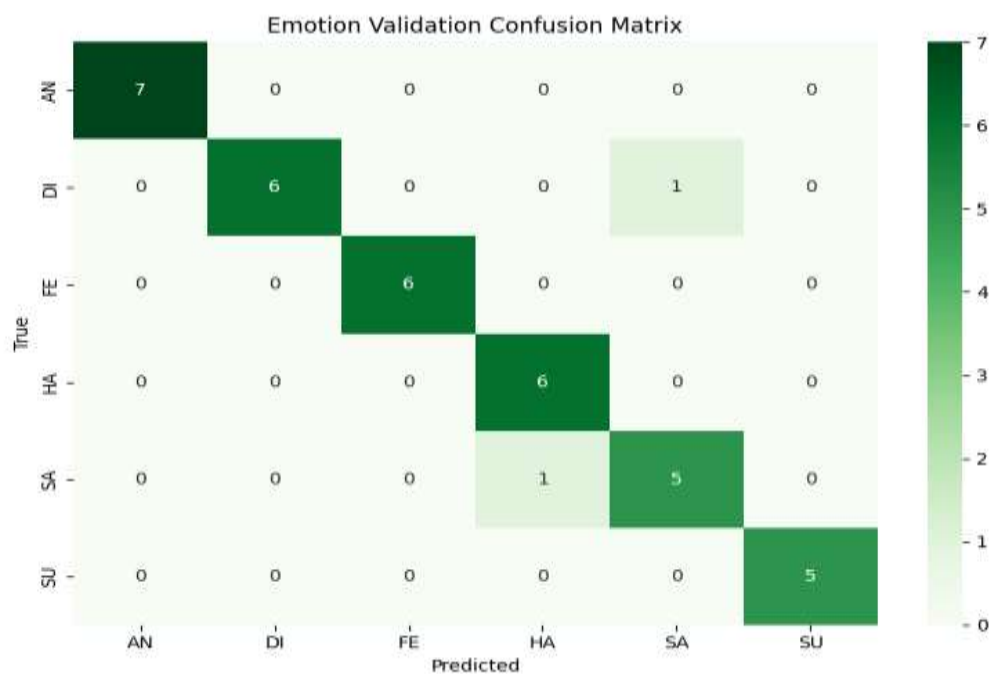


Figure 2. MLP (Multilayer Perceptron) Emotion (MLP Confusion_matrix_Test)

4.3. NN (Neural Network) emotion

Because of its capacity to identify intricate, non-linear patterns in data, a Neural Network (NN) model was used for the classification of emotions. The network automatically selects significant information from the input and translates them to emotional categories through numerous interconnected layers. The model learns to differentiate between emotions including anger, happiness, sadness, fear, disgust, and surprise by varying its weights throughout training. With an overall accuracy of 81%, the performance results show that the neural network successfully detects emotional inputs. Strong sensitivity is indicated by high recall for emotions like surprise and rage, whereas accurate predictions are demonstrated by high precision for fear and happiness. Lower recall for melancholy, however, indicates that the model would need more training data or adjustments to effectively capture nuanced emotional expressions. Overall, the findings support neural networks' applicability to tasks involving the recognition of emotions.

Table 3. NN (Neural Network) emotion

	precision	recall	f1-score	support
Anger	0.60	1.00	0.75	6
Disgust	0.71	0.83	0.77	6
Fear	1.00	0.71	0.83	7
Happiness	1.00	0.83	0.91	6
Sadness	1.00	0.50	0.67	6
Surprise	0.86	1.00	0.92	6
accuracy			0.81	37
macro avg	0.86	0.81	0.81	37
weighted avg	0.87	0.81	0.81	37

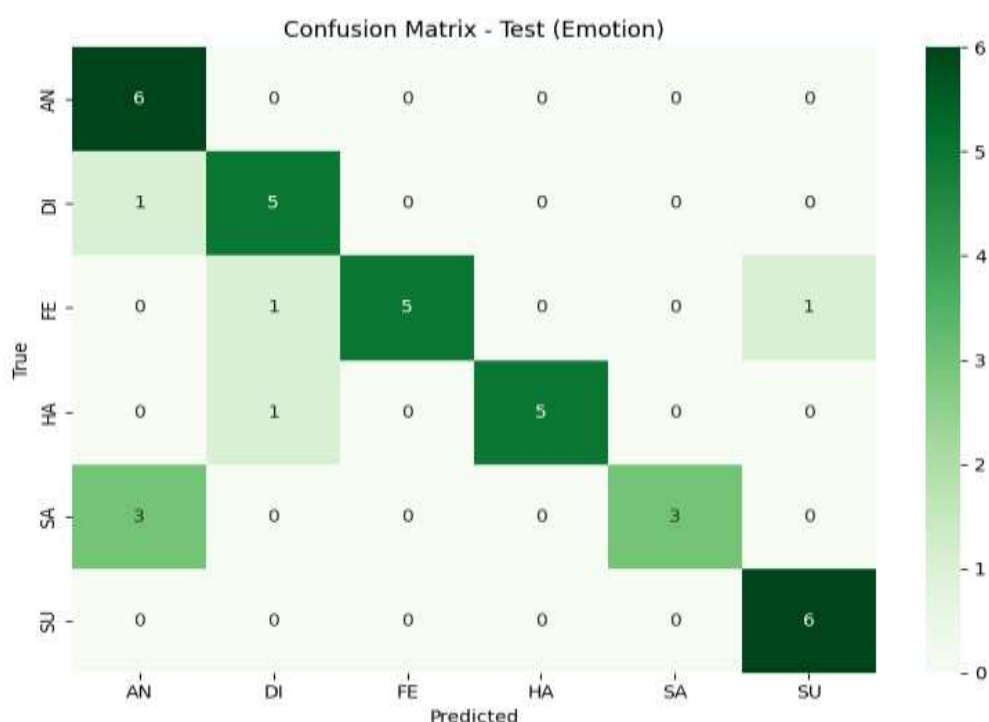


Figure 3. NN (Neural Network) Emotion (NN Confusion_matrix_Test)

4.4. RF (Random Forest) emotion

Data preparation is essential when utilising Random Forest for emotions. This could involve converting to embeddings, eliminating stopwords, and tokenizing text. Acoustic features such as energy or MFCCs are frequently extracted for speech. To prevent underfitting or overfitting, hyperparameters such as the number of trees, tree depth, and feature subset size must be adjusted. As your research demonstrates with an overall accuracy of 84%, RF can achieve excellent accuracy even on tiny datasets with the right tuning.

Table 4. RF (Random Forest) emotion

	precision	recall	f1-score	support
Anger	0.86	0.86	0.86	7
Disgust	1.00	0.71	0.83	7
Fear	0.71	0.83	0.77	6
Happiness	1.00	1.00	1.00	6

Sadness	0.57	0.67	0.62	6
Surprise	1.00	1.00	1.00	5
accuracy			0.84	37
macro avg	0.86	0.85	0.85	37
weighted avg	0.86	0.84	0.84	37

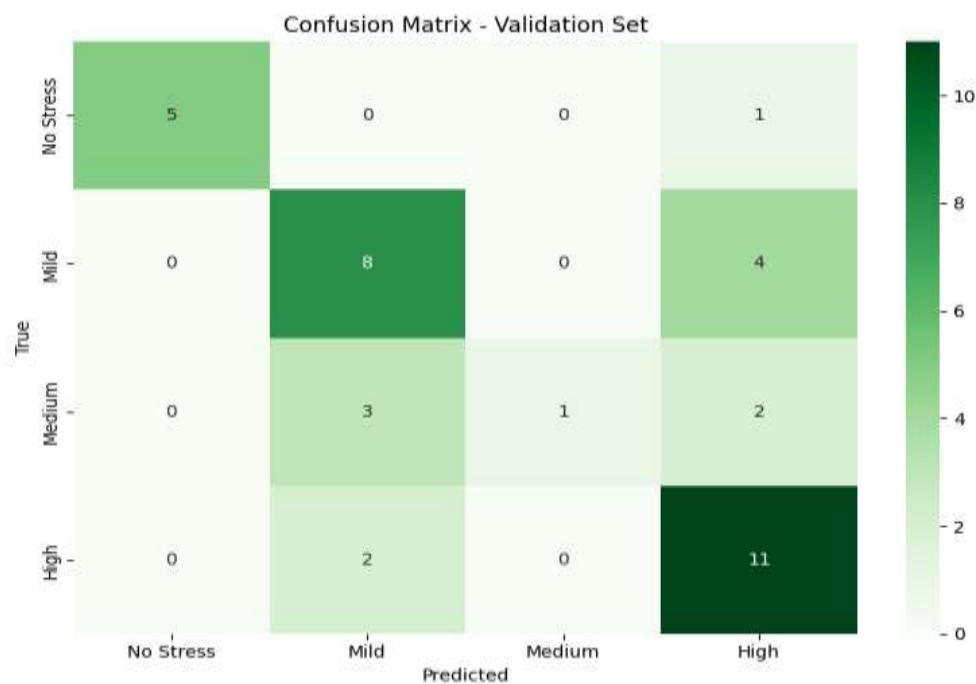


Figure 4. RF Stress (RF Confusion_matrix_Test)

4.5. SVM (Support Vector Machine)

The table displays how well an SVM model performs in categorising emotions in paragraphs. For each of the six emotions—anger, disgust, fear, happiness, sadness, and surprise—it provides precision, recall, F1-score, and support. Precision evaluates how many paragraphs predicted as a given emotion were actually correct, while recall measures how many paragraphs of that emotion were properly identified. By balancing recall and precision, the F1-score provides a single accuracy metric for every class. The number of paragraphs for each emotion in the dataset is shown by Support. The model performs exceptionally well for Happiness and Surprise, earning perfect scores in all criteria, indicating it accurately detected all paragraphs containing these emotions

without misclassifying any others. Anger exhibits strong recall but low precision, meaning that while the model identified all paragraphs containing anger, it incorrectly labelled some other paragraphs as such. Some disgust paragraphs were overlooked because disgust has great precision but poor memory. Sadness performs the worst, indicating that the model occasionally mistakes it with other emotions. Fear and Sadness both perform moderately. The model performs well across all classes, with an overall accuracy of 84% and macro and weighted averages of precision, recall, and F1-score between 0.84 and 0.86. This shows that the SVM is typically good at categorising emotions at the paragraph level, although it might do better at identifying sadness and striking a balance between the precision-recall trade-offs for anger and disgust.

Table 5. SVM (Support Vector Machine)

	precision	recall	f1-score	support
Anger	0.75	1.00	0.86	6
Disgust	1.00	0.67	0.80	6
Fear	0.83	0.71	0.77	7
Happiness	1.00	1.00	1.00	6
Sadness	0.57	0.67	0.62	6
Surprise	1.00	1.00	1.00	6
accuracy			0.84	37
macro avg	0.86	0.84	0.84	37
weighted avg	0.86	0.84	0.84	37

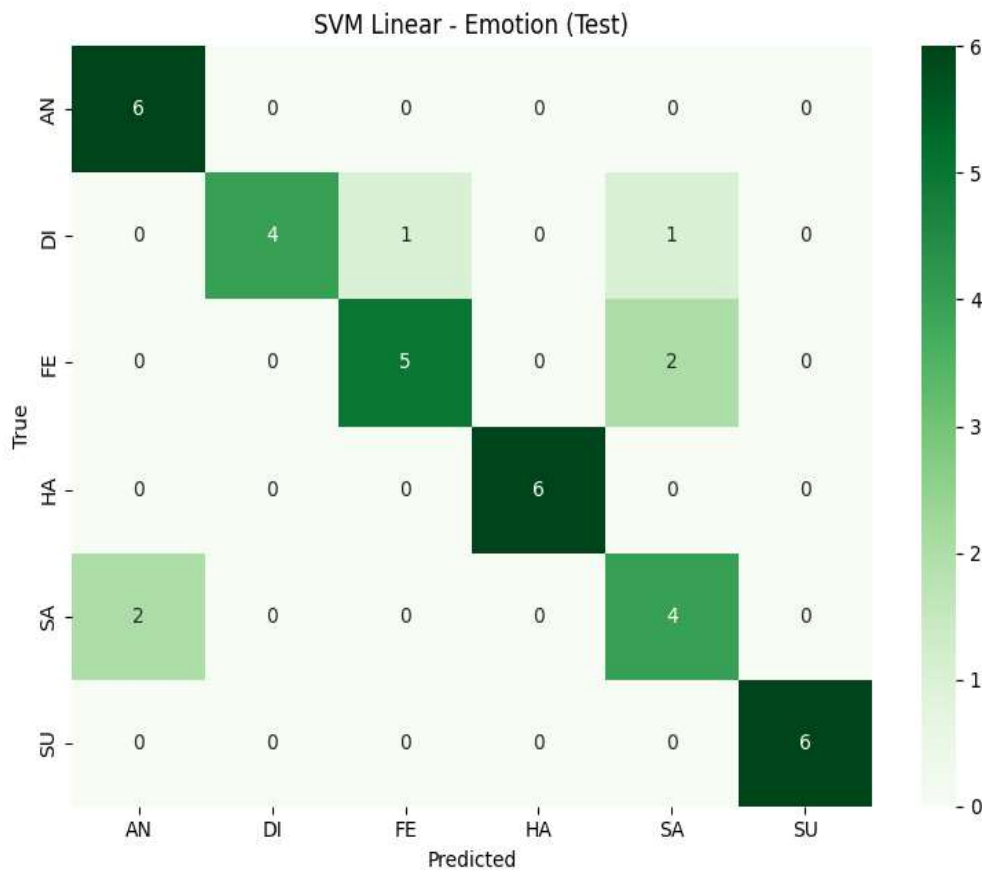


Figure 5. SVM (Support Vector Machine) Linear Emotion (SVM Confusion_matrix_Test)

4.6. Analysis of Primary Datasets

Five machine learning techniques—KNN, Keras Deep Neural Network (DeepNN), scikit-learn Neural Network (NN_Sklearn), Random Forest (RF), and SVM—are compared in the table using three datasets: Adult, Young Adult, and Teenager. The minor difference between training and testing accuracies shows that KNN somewhat overfits, but overall, it consistently obtains the highest testing accuracy for all datasets. In the Young Adult and Teenager datasets, when the training and testing accuracies are close, Keras DeepNN performs better, demonstrating good generalisation. With low training and testing accuracies, NN_Sklearn performs poorly on all datasets, indicating underfitting. Training accuracies are higher than testing, and Random Forest and SVM offer intermediate accuracy but typically overfit significantly. The Adult dataset is the most difficult for all models, while the Teenager dataset gets the highest overall accuracies. In conclusion, KNN is the best approach overall, NN_Sklearn has trouble capturing patterns, and DeepNN has promise for larger or simpler datasets.

Table 6. Analysis of Primary Datasets

Dataset	Method	Testing Accuracy	Training Accuracy
Adult	KNN	0.74	0.82
	Keras_DeepNN	0.65	0.73
	NN_Sklearn	0.35	0.42
	RF	0.47	0.6
	SVM	0.53	0.64
Young Adult	KNN	0.76	0.84
	Keras_DeepNN	0.73	0.76
	NN_Sklearn	0.45	0.52
	RF	0.52	0.66
	SVM	0.57	0.67

Teenager	KNN	0.83	0.91
	Keras_DeepNN	0.79	0.8
	NN_Sklearn	0.53	0.59
	RF	0.68	0.76
	SVM	0.69	0.72

4.7 Discussion

- With significant ramifications for mental health applications, the study's findings highlight crucial distinctions between deep learning techniques and traditional machine learning models for facial emotion recognition. The Multilayer Perceptron (MLP), which achieved 95% accuracy and consistently strong F1-scores across all six emotion classes, outperformed the other models that were assessed. This suggests that complex, non-linear patterns in facial expressions can be automatically extracted by deep learning algorithms, making it possible to accurately identify subtle emotional indicators like happiness and sadness, which frequently overlap or are less noticeable.
- K-Nearest Neighbours (KNN), on the other hand, performed somewhat (accuracy 62%), doing well on emotions with clear characteristics like surprise and terror but having trouble identifying more complex emotions. This illustrates a drawback of classical models: they are less able to capture small emotional fluctuations due to their reliance on manual feature selection. In a similar vein, Random Forest (RF) and Support Vector Machine (SVM) demonstrated diversity in precision and recall across emotions, especially for grief and rage, but reached respectable accuracy (84%). With an accuracy of 81%, the traditional Neural Network (NN) outperformed KNN but fell short of MLP, underscoring the significance of network depth and adequate training data for accurate pattern recognition.
- Analysis of various age-based datasets highlights the significant impact of dataset features on model performance. While KNN remained competitive on smaller datasets because of its simplicity, MLP-based deep neural networks generalised well on larger or more homogeneous datasets. This implies that while classical models might still be appropriate for applications with limited resources, deep learning techniques are best suited for scalable, real-time emotion identification systems.
- Overall, the work supports applications like real-time monitoring, telemedicine, and individualised mental health interventions by demonstrating that AI-driven

emotion identification can be a potent tool for early detection of psychiatric illnesses. In order to increase robustness and application in actual healthcare situations, future research should concentrate on increasing dataset size and diversity, incorporating multimodal characteristics (such as speech and physiological signals), and refining neural network topologies.

5. Conclusion

The effectiveness of five machine learning models—K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and a traditional Neural Network (NN)—for face emotion identification across six fundamental emotions was compared in this study. With an overall accuracy of 95% and consistently high F1-scores across all emotion classes, the results show that deep learning models, especially the MLP, greatly outperform classical techniques. While the conventional NN had limits in capturing subtle emotional patterns due to insufficient training data or network complexity, traditional models like KNN, RF, and SVM produced modest accuracies (62–84%) and struggled with subtle emotions like melancholy and happiness. A comparison of several age-based datasets showed that while KNN is still competitive for smaller datasets because of its simplicity, deep neural networks generalise well for bigger or more homogeneous datasets. The results demonstrate the promise of AI-powered emotion identification systems for early psychiatric condition detection, facilitating real-time mental health monitoring via wearable technology, telemedicine, and digital health apps. In order to further enhance resilience and practical applicability, future research should concentrate on diversifying datasets, adding multimodal inputs, and refining neural network topologies.

6 References

1. Yang, B.; Cao, J.M.; Jiang, D.P.; Lv, J.D. Facial expression recognition based on dual-feature fusion and improved random forest classifier. *Multimed. Tools Appl.* 2017, 9, 20477–20499.
2. Saba, T.; Khan, S.U.; Islam, N.; Abbas, N.; Rehman, A.; Javaid, N.; Anjum, A. Cloud-based decision support system for the detection and classification of malignant cells in breast cancer using breast cytology images. *Microsc. Res. Tech.* 2019, 82, 775–785.
3. Yousuf, M.; Mehmood, Z.; Habib, H.A.; Mahmood, T.; Saba, T.; Rehman, A.; Rashid, M. A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval. *Math. Probl. Eng.* 2018, 2018, 13.
4. Rehman, A.; Saba, T. Performance analysis of character segmentation approach for cursive script recognition on benchmark database. *Digit. Signal Processing* 2011, 21, 486–490.
5. Haron, H.; Rehman, A.; Adi DI, S.; Lim, S.P.; Saba, T. Parameterization method on B-spline curve. *Math. Probl. Eng.* 2012, 2012, 640472.
6. Nodehi, A.; Sulong, G.; Al-Rodhaan, M.; Al-Dhelaan, A.; Rehman, A.; Saba, T. Intelligent fuzzy approach for fast fractal image compression. *EURASIP J. Adv. Signal Processing* 2014, 2014, 112.
7. Rahim, M.S.M.; Norouzi ARehman, A.; Saba, T. 3D bones segmentation based on CT images visualization. *Biomed. Res.* 2017, 28, 3641–3644.
8. RahimMS,M.; Rehman, A.; Kurniawan, F.; Saba, T. Ear biometrics for human classification based on region features mining. *Biomed. Res.* 2017, 28, 4660–4664
Ahmad,A.M.; Sulong, G.; Rehman, A.; Alkawaz, M.H.; Saba, T. Data hiding based on improved exploiting modification direction method and Huffman coding. *J. Intell. Syst.* 2014, 23, 451–459.

9. Yousaf, K.; Mehmood, Z.; Awan, I.A.; Saba, T.; Alharbey, R.; Qadah, T.; Alrige, M.A. A comprehensive study of mobile-health based assistive technology for the healthcare of dementia and Alzheimer's disease (AD). *Health Care Manag. Sci.* 2020, 23, 287–309.
10. Saba, T.; Javed, R.; Shafry, M.; Rehman, A.; Bahaj, S.A. IoMT Enabled Melanoma Detection Using Improved Region Growing Lesion Boundary Extraction. *CMC-Comput. Mater. Contin.* 2022, 71, 6219–6237.
11. Rehman, A.; Kurniawan, F.; Saba, T. An automatic approach for line detection and removal without smash-up characters. *Imaging Sci. J.* 2011, 59, 177–182.
12. Saba, T.; Rehman, A.; Al-Dhelaan, A.; Al-Rodhaan, M. Evaluation of current documents image denoising techniques: A comparative study. *Appl. Artif. Intell.* 2014, 28, 879–887.
13. Yousaf, K.; Mehmood, Z.; Saba, T.; Rehman, A.; Munshi, A.M.; Alharbey, R.; Rashid, M. Mobile-health applications for the efficient delivery of health care facility to people with dementia (PwD) and support to their carers: A survey. *BioMed Res. Int.* 2019, 2019, 7151475.
14. Ullah, H.; Saba, T.; Islam, N.; Abbas, N.; Rehman, A.; Mehmood, Z.; Anjum, A. An ensemble classification of exudates in color fundus images using an evolutionary algorithm based optimal features selection. *Microsc. Res. Tech.* 2019, 82, 361–372.
15. Keltner, D.; Sauter, D.; Tracy, J.; Cowen, A. Emotional expression: Advances in basic emotion theory. *J. Nonverbal Behav.* 2019, 43, 133–160.
16. Ramzan, F.; Khan, M.U.G.; Iqbal, S.; Saba, T.; Rehman, A. Volumetric segmentation of brain regions from MRI scans using 3D convolutional neural networks. *IEEE Access* 2020, 8, 103697–103709.
17. Jamal, A.; Hazim Alkawaz, M.; Rehman, A.; Saba, T. Retinal imaging analysis based on vessel detection. *Microsc. Res. Tech.* 2017, 80, 799–811.
18. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition-workshops*, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
19. Harouni, M.; Rahim, M.S.M.; Al-Rodhaan, M.; Saba, T.; Rehman, A.; Al-Dhelaan, A. Online Persian/Arabic script classification without contextual information. *Imaging Sci. J.* 2014, 62, 437–448.
20. Nasir, I.M.; Raza, M.; Shah, J.H.; Khan, M.A.; Rehman, A. Human action recognition using machine learning in uncontrolled environment. In *Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Riyadh, Saudi Arabia, 6–7 April 2021; pp. 182–187.
21. Xie, S.; Hu, H.; Wu, Y. Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. *Pattern Recognit.* 2019, 92, 177–191.
- Yu, Z.; Zhang, C. Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, Seattle, WA, USA, 9–13 November 2015; pp. 435–442.
22. Jain, D.K.; Zhang, Z. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. *Pattern Recognit. Lett.* 2020, 139, 157–165.
23. Zhang, K.; Huang, Y.; Du, Y.; Wang, L. Facial expression recognition based on deep evolutionary spatial-temporal networks. *IEEE Trans. Image Processing* 2017, 26, 4193–4203.

24. Chu, W.S.; DeLaTorre, F.; Cohn, J.F. Learning spatial and temporal cues for multi-label facial action unit detection. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 25–32.
25. Lopes, A.T.; De Aguiar, E.; De Souza, A.F.; Oliveira-Santos, T. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognit.* 2017, 61, 610–628.
26. Li, J.; Wang, Y.; See, J.; Liu, W. Micro-expression recognition based on 3D flow convolutional neural network. *Pattern Anal. Appl.* 2019, 22, 1331–1339.
27. Sadad, T.; Munir, A.; Saba, T.; Hussain, A. Fuzzy C-means and region growing based classification of tumor from mammograms using hybrid texture feature. *J. Comput. Sci.* 2018, 29, 34–45.
28. Khan, M.A.; Kadry, S.; Zhang, Y.D.; Akram, T.; Sharif, M.; Rehman, A.; Saba, T. Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Comput. Electr. Eng.* 2021, 90, 106960.
29. Chang, K.Y.; Chen, C.S.; Hung, Y.P. Intensity rank estimation of facial expressions based on a single image. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Manchester, UK, 13–16 October 2013; pp. 3157–3162.
30. Acevedo, D.; Negri, P.; Buemi, M.E.; Fernández, F.G.; Mejail, M. A simple geometric-based descriptor for facial expression recognition. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 802–808.
31. Ghimire, D.; Lee, J.; Li, Z.N.; Jeong, S. Recognition of facial expressions based on salient geometric features and support vector machines. *Multimed. Tools Appl.* 2017, 76, 7921–7946.
32. Al-Agha, S.A.; Saleh, H.H.; Ghani, R.F. Geometric-based feature extraction and classification for emotion expressions of 3D video film. *J. Adv. Inf. Technol.* 2017, 8, 74–79.
33. Reddy, C.V.R.; Reddy, U.S.; Kishore, K.V.K. Facial Emotion Recognition Using NLPCA and SVM. *Traitement du Signal* 2019, 36, 13–22.
34. Siddiqi, M.H.; Ali, R.; Khan, A.M.; Kim, E.S.; Kim, G.J.; Lee, S. Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. *Multimed. Syst.* 2015, 21, 541–555.