Deep Learning Approach for Anxiety and Stress Detection through Facial Emotion Analysis

Houcem REZAIGUIA*
University of Biskra, Biskra, ALgeria.
Abdelhamid DJEFFAL†
University of Biskra, Biskra, Algeria.

Abstract

Facial expressions are crucial in everyday emotional communication since they are an indicator of sentiments and allow a person to convey his emotional state. People can instantly understand another person's emotional condition based on their facial expressions. As a result, information on facial expressions is frequently employed in automatic emotion recognition systems. Anxiety is one of the most prevalent emotions that people experience in various situations. As a result, it must be identified and treated. In this study, we propose a novel framework, named ANet, that utilises the AlexNet Convolutional Neural Networks (CNN) model for anxiety and stress detection through the analysis of facial emotions. The proposed system follows a step-by-step process, starting with obtaining images from a public database and employing the Viola-Jones algorithm for face detection. Subsequently, the detected face region is extracted as the Region of Interest (ROI). During the training phase, the images are trained as feature maps using the AlexNet CNN model. Once the model is trained, input videos undergo frame extraction, face detection, and ROI extraction. The extracted ROI is passed to the classifier for emotion identification, including anger, sadness, happiness, disgust, fear, and neutral. By analysing the combination of detected facial emotions, anxiety levels are determined and categorised as high anxiety, moderate anxiety, mild anxiety, and no anxiety. The proposed ANet framework demonstrates promising results in anxiety detection, providing a reliable and efficient method for early identification and assessment of anxiety based on facial emotions.

Key Words: anxiety, CNN, AlexNet, facial emotions, face detection.

1 Introduction

Understanding human emotion is important for enhancing normal utility and preventing undesirable effects. Affective computing systems that could detect and interpret human non-verbal behaviours, like facial expressions, speech prosody, body movements, and skin conductance, are necessary [13]. The presence of multiple stressors can elicit diverse affective states and physiological responses in individuals. The diagnosis and treatment of stress-related disorders, along with anxiety, have garnered increasing attention from researchers and clinicians. Anxiety detection plays a pivotal role in preventive mental health interventions [9]. Anxiety is widespread, adversely affecting daily functioning, leading to significant suffering, and resulting in substantial healthcare expenses as well as costs linked to decreased productivity [19]. It adversely impacts an individual's health, induces negative feelings, and, in severe cases, may lead to mental health disorders. Statistics on anxiety gathered by Single Care show that the main causes of stress are mostly the same around the world, including money problems, work pressures, and family responsibilities. A novel stressor has arisen since 2020: the COVID-19 pandemic, a divisive political environment, and supplementary factors [8]. Although clinical guidelines are available [16], the handling of these conditions in primary care frequently falls short of optimal standards. These conditions can have a profound impact on an individual's emotional well-being, daily functioning, and overall quality of life. Early detection and intervention play a crucial role in effectively managing and treating anxiety and depression, preventing further deterioration and promoting better mental health outcomes. Conventionally, anxiety and depression diagnoses have mainly depended on clinical interviews, self-report questionnaires, and subjective evaluation by mental health practitioners. Nonetheless, these approaches commonly have inherent drawbacks, including subjectivity, susceptibility to biases, and the inability to conduct real-time monitoring. As such, there is an increasing demand for objective and automatic approaches that can enable early detection and evaluation of anxiety and depression. Computer vision and machine learning advancements have opened up new possibilities for creating automated systems that can quantify and interpret human emotions. Facial expression analysis of emotions has become a viable means of determining the emotional state of a person due to the large amount of information carried by such expressions. Given the current state of technology, we propose

*Computer Science Department , LESIA Laboratory , University of Biskra, Biskra, Algeria. Email: houcem.rezaiguia@univ-biskra.dz.

†Computer Science Department , LESIA Laboratory , University of Biskra, Biskra, Algeria. Email: abdelhamid.djeffal@univ-biskra.dz.

a new framework known as ANet (AlexNet Convolutional Neural Networks model) that aims to detect depression and anxiety through the analysis of facial expressions linked to emotional reactions. The ANet model combines cutting-edge deep learning approaches. For example, it uses the AlexNet convolutional neural network (CNN) model to train itself and subsequently to detect relevant features from salient images of faces that have been extracted and cropped. The AlexNet configuration has been popularly trained with much success for various general applications of computer vision, including object detection and image classification that emphasise non-facial characteristics. Thus, for the detection of anxiety and depression, we applied the AlexNet model to assess whether it, too, could detect and measure emotionality related to anxiety and depression. The suggested framework adopts a systematic approach in detecting anxiety and depression. It starts by acquiring a dataset comprising facial images from an open repository, thereby assuring variation in the emotional states and offering adequate representation. Face detection is then carried out using the Viola-Jones algorithm, enabling appropriate localisation of the area of interest on the face. After detecting the facial area, it is considered the Region of Interest (ROI) and used as the input for the further stages of analysis. The ANet model utilises the AlexNet CNN model to train during the training process and learn network parameters' optimisation. It comes from the ANet dataset, correlating facial expressions with their associated derived emotions—and depression and anxiety, for example. Therefore, once the training is complete, the ANet model can effectively and accurately assign emotions such as anger, sadness, happiness, disgust, fear, and neutrality. To assess degrees of anxiety and depression, the ANet framework uses attributed emotions to determine the degree of anxiety and depression. It can classify from the highest level of anxiety down to minor assessments: high anxiety, moderate anxiety, mild anxiety, and no anxiety. This is beneficial for trained practitioners because it provides them with more data to use in treatment assessments and options. In summary, the ANet model presents an innovative and promising approach to the detection of depression and anxiety through facial emotion analysis. Using deep learning along with the AlexNet convolutional neural network model, the system seeks to offer a clear, automated, and immediate way to identify and assess depression and anxiety. These advances have the potential to transform mental health treatment by enabling early interventions and personalised treatments, thereby optimizing overall health and quality of life in individuals who suffer from a broad array of mental health conditions. This paper is organised in the subsequent way. Section 2 presents a comprehensive examination of the existing literature. The methodology is outlined in Section 3. Section 4 elaborates on the findings, while Section 5 presents the conclusions.

2 Litterature review

Facial expression analysis has recently garnered significant attention as a method for detecting human anxiety. A person's facial expression often tells an observer how they are feeling. The expression shows how someone feels about themselves or how they feel about the person watching [18]. Researchers have been working on automatic anxiety detection for a long time. Studies show that it is possible to automatically recognise worries through facial clues. Some methods are more invasive, including blood or saliva testing, while others are less invasive and include collecting images. Deep Learning (DL) is a type of machine learning. This method uses an artificial neural network as a foundation for training and characterisation [15]. The information gained during deep learning helps us understand data, such as pictures of faces. The first step in using deep learning to detect facial expressions is to choose a data-intensive expression model. The next stage is to use Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Recurrent Neural Networks, and other similar methods. Deep learning is a more advanced form of machine learning that is better at photo identification than typical heuristic classification methods. People use deep learning frameworks for many things, including recognising facial emotions and movements.

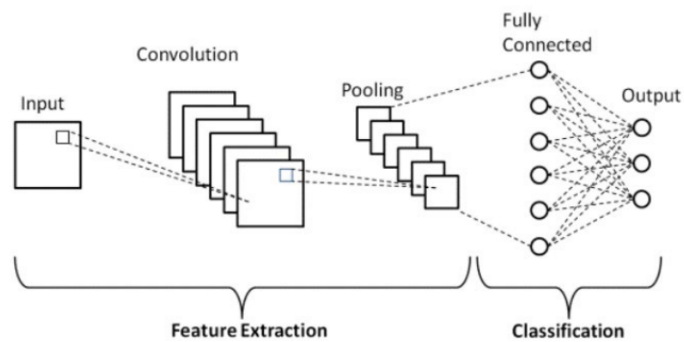


Figure 1: Structure of deep Neural Networks (DNN).

CNNs, or Convolutional Neural Networks, are very advanced neural networks that can respond well to nearby units in their receptive field. These networks have done a wonderful job at processing large images. Figure 1 has additional information. The architecture of CNN has three separate layers. The convolutional layers, the pooling layers, and the fully connected (FC) layers are the three levels. The system has a complete categorisation framework, which makes it possible to find and confirm images. Some of the most well-known convolutional neural network designs are LeNet, AlexNet, VGG, GoogLeNet, and ResNet. The authors [1] suggested using the VGG16, VGG19, and ResNet V2 models to look at facial expression recognition systems that use CNN techniques to find stress. The VGG16 model had the highest accuracy at 0.7665, the VGG19 model had 0.7257, and the ResNet V2 model had 0.8249. CNN, which is a traditional framework in the field of machine learning, has demonstrated remarkable success in a variety of computer tasks, including image enhancement, image processing, and picture identification.

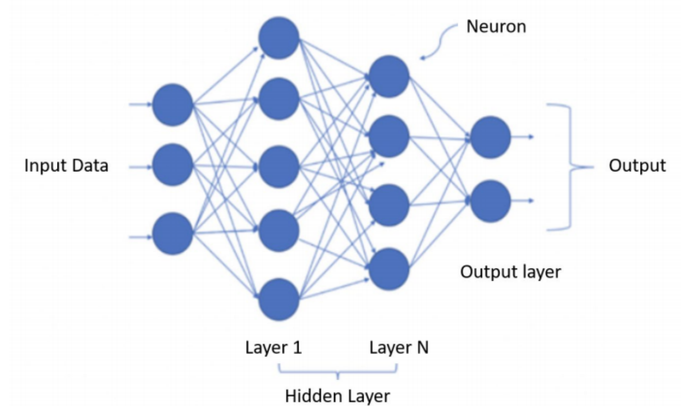


Figure 2: Structure of deep Neural Networks (DNN).

Figure 2 depicts the architecture of a deep neural network. DNN is a deep learning framework distinguished by the inclusion of a minimum of one hidden layer within its neural network design. This type of network is akin to shallow neural networks. DNN enables the modelling of complex, non-linear systems. A DNN is a discriminative model that may be trained using backpropagation methods. The trained DNN model can obtain the estimated ratio, an essential component of facial expression. EmotionNet Nano is the inaugural neural network architecture for face expression identification using DNN. In a study [14] comparing the performance of EmotionNet Nano – A and Nano – B networks, EmotionNet Nano – A achieved accuracy similar to the best larger networks. Conversely, EmotionNet Nano-B exhibits lower accuracy relative to the top-performing networks; however, it still achieves comparable accuracy while being three orders of magnitude smaller in terms of parameter count. EmotionNet Nano variations offer an optimal balance of accuracy and complexity, making them appropriate for embedded applications.

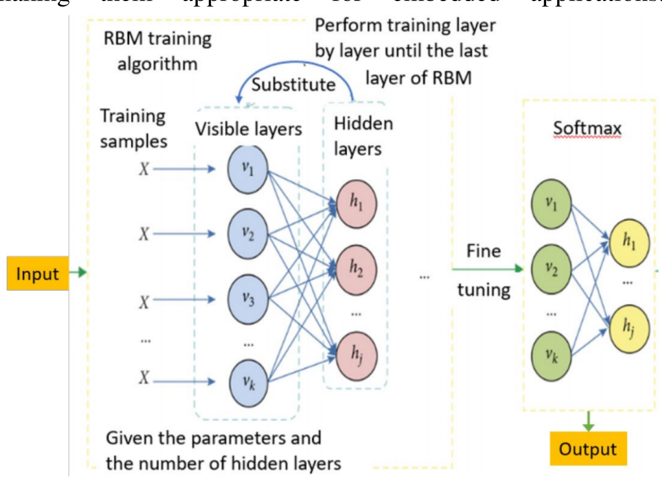


Figure 3: Structure of deep Neural Networks (DNN).

The Deep Belief Network (DBN) method represents a neural network employed in machine learning, applicable in both unsupervised and supervised learning contexts. Figure 3 demonstrates the performance of the Deep Belief Network (DBN) as it processes the input image to produce the output. DBN creates a joint distribution that connects observed data with labels. This technique is different from the traditional neural network approach used by the discriminant model. By adjusting the weights of its neurones to achieve the highest probability, the entire neural network can generate training data. Wu and Qiu [26] did a study to see how well DBN could recognise facial expressions. They used LBP and an improved DBN to get recognition rates in the JAFFE database. In conclusion, the DBN method works better than other algorithms on three datasets. The DBN model uses deep learning networks to find more complex features and boost recognition rates. An RNN sends state information across its network in a systematic way so that it can handle a wider range of time series structural inputs. RNNs can sort image sequences by their time-based relationships. redictions online. This method works best for recognising facial emotions in real time [11], and [20] have made an improved version. However, a Recurrent Neural Network (RNN) is a type of neural network that has trouble dealing with the problem of vanishing gradients in recursive situations. Giannakakis et al in [10] developed a system capable of identifying emotional states of stress and anxiety through the analysis of video-recorded facial expressions. Using many approaches, including Active Appearance Models, Optical Flow, and rPPG, they extracted the most pertinent information for stress classification. A K-NN classifier achieved a classification accuracy of 87.72%. The study by [24] proposed a system that detects symptoms of stress using Facial Action Units (FAUs) derived from movies. The researchers conducted a binary classification employing various elementary classifiers on facial action units (FAUs) taken from each video frame, attaining an accuracy of as much as 74% in subject-independent categorisation and 91% in subject-dependent classification.

3 Materials and Methods

Facial emotion-based anxiety level identification is performed through a series of modules, each contributing to the comprehensive analysis of video data.

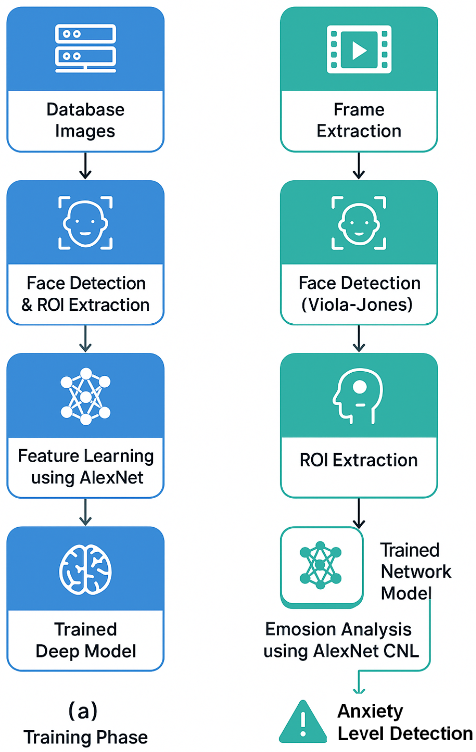


Figure 4: Overall proposed methodology. (a) training Module. (b) testing Module.

a- Training module:

1. **RADIATE facial stimulus collection:** The RADIATE facial stimulus collection includes more than 1,500 distinct photographs featuring a diversified group of more than a hundred models, comprising 25% non-Hispanic White and 75% from minority or ethnic groups. Each model exhibited 16 distinct facial expressions [23], offering a broad spectrum of emotions within a stimulus set that reflects various racial and ethnic backgrounds. The RADIATE stimulus set images were created to allow for integration with other face stimulus sets, utilising a standardised scarf template that accompanies the face stimuli.
2. **Face detection via Viola–Jones algorithm:** Using the Viola–Jones algorithm, face detection was implemented. The detected face region is cropped for the next step, known as ROI extraction. In 2001, Paul Viola and Michael Jones developed the Viola–Jones object detection framework [25], the first object detection system to deliver competitive object detection rates in real time. Although it may be trained to recognise a range of object classes, it was originally driven by the problem of face detection. The open-source implementation of the Viola–Jones detector made it famous. This method is frequently used to detect the face region in an image in order to find an object of unknown size. It has high efficiency and accuracy.

   The Viola–Jones method employs three strategies:

   a. Integral image utilised for the purpose of feature extraction Rectangular in nature, Haar-like features are derived through the use of an integral image.

   b. Adaboost is a face identification algorithm based on machine learning [12]. The term "boosted" refers to the fact that the classifiers at each stage of the cascade is inherently complex, constructed from simple classifiers using one of four boosting approaches (weighted voting). The Adaboost algorithm represents a systematic approach to learning that starts with a weak classification and then learns and constructs a strong classification using the weight value.

   c. A cascade classifier is used to efficiently merge many features. The word "cascade" in the classifier name refers to the fact that the final classifier is made up of numerous smaller classifiers (stages) that are applied one after another to a region of interest until the candidate is rejected or all stages are passed [27]. After cascading each of the strong classifiers, the model may acquire the non-face and face regions.
3. **Training using AlexNet CNN model:** Once the facial region has been extracted, the training phase can begin. Throughout the training phase , images are trained as feature maps using the AlexNet CNN model. AlexNet marked the inaugural use of a convolutional neural network in the LSVRC competition, as noted by Krizhevsky et al. 2012 [3], achieving a significantly higher accuracy than all prior models, including the runner-up. AlexNet employs the graphics processing unit (GPU) to improve performance. AlexNet consists of five convolutional layers, three max-pooling layers, two normalisation layers, two fully connected layers, and one softmax layer within its architecture. Each convolutional layer utilises convolutional filters in conjunction with the nonlinear activation function ReLU. The implementation of max pooling involves the utilisation of pooling layers. The fixed input size is a result of the inclusion of fully connected layers. The input dimensions are commonly stated as 224x224x3; however, due to the application of padding, the true dimensions are 227x227x3. AlexNet comprises a total of 60 million parameters.

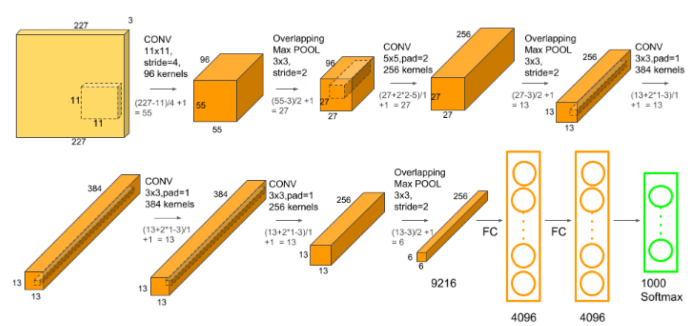


Figure 5: AlexNet architecture.

AlexNet consists of three primary components:

(1) Utilises the ReLU non-linear activation function to

address the vanishing gradient issue more efficiently compared to sigmoid and tanh activation functions found in other neural networks;
(2) Incorporates dropout and data augmentation within the network layer to mitigate overfitting; and
(3) Leverages multiple parallel GPUs to enhance computational throughput during the training process.

Table 1: Architecture of AlexNet CNN

AlexNet Layers	# Kernels	Kernel Size	Stride	Padding	Output Size
Conv1	96	11×11	[4,4]	[0 0 0 0]	$96 \times 55 \times 55$
Mpool1	96	3×3	[2,2]	[0 0 0 0]	$96 \times 27 \times 27$
Conv2	256	5×5	[1,1]	[2 2 2 2]	$256 \times 27 \times 27$
Mpool2	256	3×3	[2,2]	[0 0 0 0]	$256 \times 13 \times 13$
Conv3	384	3×3	[1,1]	[1 1 1 1]	$384 \times 13 \times 13$
Conv4	384	3×3	[1,1]	[1 1 1 1]	$384 \times 13 \times 13$
Conv5	256	3×3	[1,1]	[1 1 1 1]	$256 \times 13 \times 13$
Mpool5	256	3×3	[2,2]	[0 0 0 0]	$256 \times 6 \times 6$
Fc6					$4096 \times 1 \times 1$
Fc7					$4096 \times 1 \times 1$
Fc8 & Softmax					$5 \times 1 \times 1$

b- Testing module:

1. **Video Acquisition:** Videos used for analysis are collected from a publicly available database. The selection of diverse videos ensures representation across different emotional expressions and enables a robust assessment of anxiety levels.
2. **Frames Extraction & Face Detection:** Upon acquiring the videos, frames are extracted from each video to facilitate frame-by-frame analysis. Subsequently, face detection is applied to each frame using Haar Feature-based Cascade Classifiers, specifically the Viola-Jones algorithm. This process accurately localizes facial regions in each frame. After successful face detection, the detected faces are cropped, isolating the Region of Interest (ROI). The ROI extraction step focuses on obtaining the essential facial features necessary for precise emotion analysis.
3. **Facial Emotions Analysis:**The used AlexNet model for facial emotions classification is gained by its ability to pull out important features from pictures of faces. This model is emotion classification-oriented and relies on a classification approach for implementation. The architecture of AlexNet is composed of five convolutional layers along with three fully connected layers.

To address overfitting, dropout is introduced as a regularization technique. The training process is carried out using stochastic momentum gradient descent (SGDM) optimization with an initial learning rate of 0.0003. The training dataset is processed through multiple epochs, indicating the total training time on the complete dataset. The architecture, training dataset, and options for training are predefined before training the AlexNet network. During the classification stage, the trained network classifies each video frame into distinct emotional categories, including Sadness, Anger, surprise, Happiness, Disgust, fear and Neutral. Anxiety Level Identification: The identification of anxiety levels is determined based on the total amount of positive and negative emotions observed throughout the entire video. Positive emotions consist of 'happy' and 'neutral,' while negative emotions encompass 'sad,' 'anger,' fear,' and 'disgust.' By analyzing the balance between positive and negative emotions, the anxiety levels are categorized into four distinct levels, serving as a comprehensive assessment of anxiety.

There are still some restrictions even though the suggested methodology shows encouraging results in determining anxiety levels from emotional facial expressions. For training, the method mostly uses the RADIATE dataset, which is varied but might not adequately represent the unpredictability of real-world situations like shifting lighting, occlusions, or impromptu facial expressions. Furthermore, dividing distinct emotion categories into positive and negative groups is the basis for classifying anxiety levels, which may ignore more nuanced affective cues.

4 Experimental Results

An evaluation of the suggested facial expression recognition system was conducted through experiments employing the Radiate dataset, comprising a total of 1504 instances spanning seven emotion classes. The dataset was partitioned into training and validation sets employing a hold-out validation strategy, where 90 % of the data was allocated for training (imdsTrain) and 10 % for validation (imdsValidation). A pretrained AlexNet convolutional neural network was fine-tuned on the training data using stochastic gradient descent with momentum (SGDM). Data augmentation strategies, including random translation and horizontal flipping, were applied to improve generalization. The network was trained for 200 epochs, and performance was monitored on the validation set throughout the training phase. The confusion matrix indicates that the classifier demonstrates robust performance across all classes, with minimal misclassification. Each row denotes the actual class, whilst each column signifies the anticipated class. The matrix values represent the frequency of instances assigned to each category. For example, class 1 anger achieved 210 true positives out of 218 instances, while class 4 happy achieved the highest true positive count with 315 correct predictions. The overall accuracy on the validation set reached 98.01 %, highlighting the effectiveness of the proposed approach.

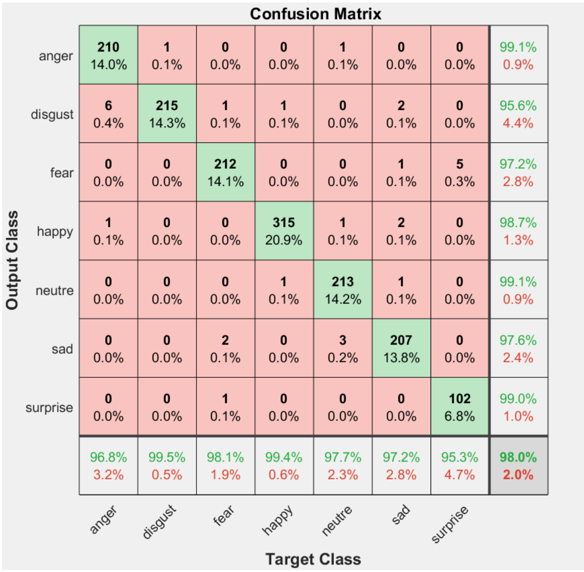


Figure 6: Confusion matrix.

In addition to accuracy, several evaluation metrics were computed to provide a more comprehensive analysis. Table 2 summarizes class-wise and overall performance:

Table 2: Performance metrics

Metric	**Value**
Accuracy	98.01%
F1-score	97.87%
Average Sensitivity (Recall)	97.72%
False Positive Rate	0.34%
Average Specificity	99.66%
Matthews Correlation Coefficient (MCC)	0.9755
Average Precision	98.05%
Cohen's Kappa	0.9186

Comprehensive The performance metrics of the facial emotion-based anxiety level identification system are summarized in Table 2. These indicators offer insights into the efficacy and resilience of the suggested methodology.

Precision: This metric quantifies the model's confidence when predicting a given class. For example, the precision for class 1 was 99.06%, indicating that when the model predicts this emotion, it is very likely to be correct.

Recall (Sensitivity): This measures how well the model captures all true instances of a class. Class 1 showed a recall of 96.77%, meaning that most instances were correctly identified.

F1-Score: The harmonic mean of precision and recall provides a balanced measure., especially useful for imbalanced datasets. For class 1, the F1-score reached 97.90%.

Specificity: This indicates how accurately the model recognizes negative cases, i.e., instances that do not belong to a particular class. The specificity for class 1 was 99.85%, showing excellent performance in avoiding false positives.

Accuracy: Defined as the total proportion of correctly predicted samples across all classes, the overall accuracy achieved was 98.01%.

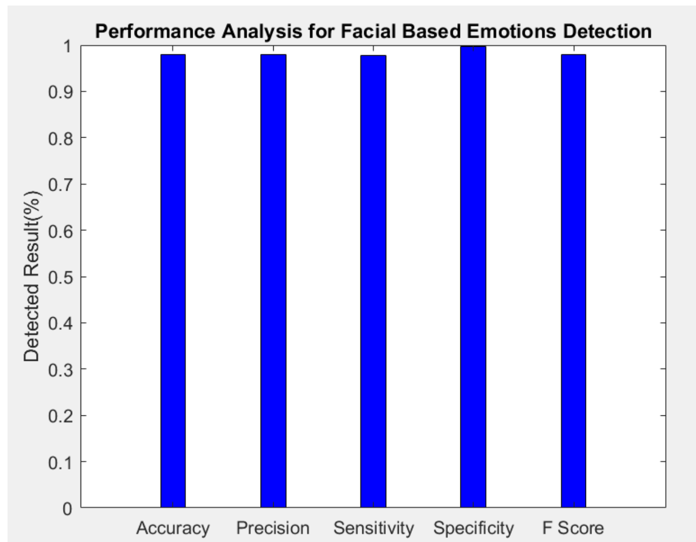


Figure 7: Performance analysis for facial based emotions detection.

The performance Analysis in Fig 7 shows that the model performs well in identifying emotions. The high values for precision, recall, F1-score, and specificity indicate that the model exhibits reliability and is competent at distinguishing correctly among various emotional states. In addition, the high overall accuracy of the model supports this claim of performance.

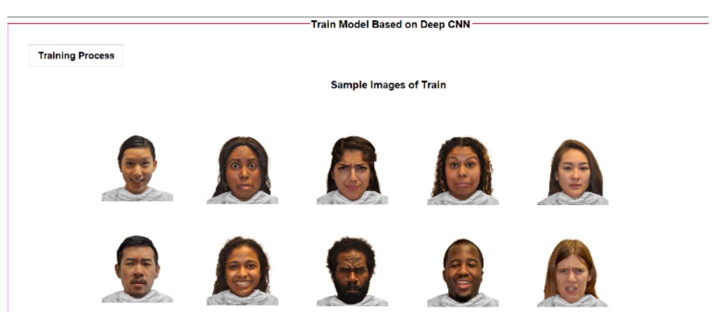


Figure 8: Sample images of train model based on deep CNN.

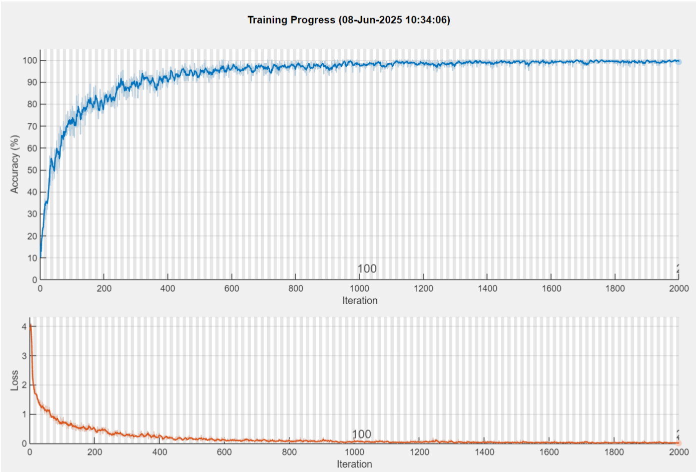


Figure 9: Training process.

Figure 9 illustrates the training progress over 2,000 iterations, showing both accuracy and loss convergence. The model demonstrates rapid learning in early iterations, with accuracy increasing sharply and stabilizing near 98–99%, while the loss gradually decreases and converges towards zero. This indicates an effective optimization process and minimal overfitting, further validating the generalization capability of the model.

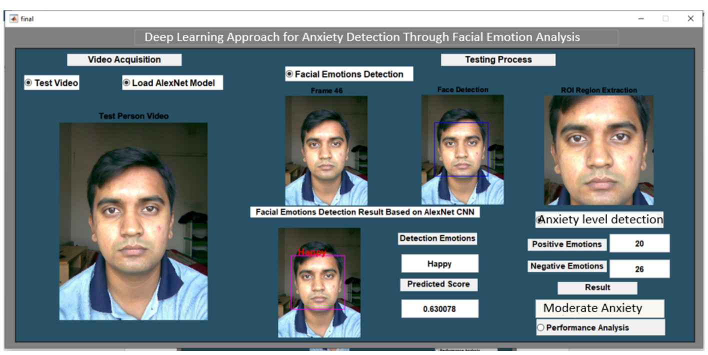


Figure 10: Testing process for person 1.

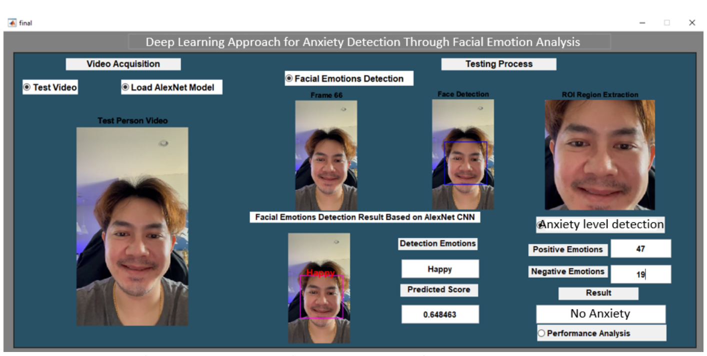


Figure 11: Testing process for person 2.

Table 3: Algorithm Accuracy Comparison

Study	**Algorithm Type**	**Accuracy Range**
[5]	CNNEELM (Convolutional Neural Network with Extreme Learning Machine hybrid)	98% (CK+), 96.53% (FER2013)
[2]	Discriminant Analysis (Quadratic), SVM, Random Forest, K-Nearest Neighbors, Naive Bayes, Decision Tree, Multilayer Perceptron	100% (Discriminant Analysis, SVM, Random Forest); others not reported
[6]	AdaBoost+SVM; LDA	93% (SVM+AdaBoost)
[17]	Radial Basis Function Support Vector Machine (RBF SVM) (with HOG+LBP); others not specified	94% (RBF SVM HOG+LBP); F1: 0.93
Our study	AlexNet Convolutional Neural Networks (CNN) model	98.01% with RADIATE dataset

Table 4: AlexNet Accuracy Comparison Across Different Studies

Study	**Dataset**	**AlexNet Accuracy Range**
[4]	JAFFE	89.2% (Linear Discriminant Analysis), 87.8% (K-Nearest Neighbors)
[22]	FERC-2013	>65%
[7]	CK+ + JAFFE	90%
[21]	FER2013	76%
Our study	RADIATE	98.01%

5 Conclusions

In this work, we introduced ANet, a novel and effective deep learning framework for the automatic detection and assessment of anxiety levels based on facial emotional expressions. Utilizing the AlexNet Convolutional Neural Network (CNN), our approach demonstrates a reliable, structured pipeline that includes face detection, region of interest (ROI) extraction, feature learning, and emotion classification. By mapping classified emotions to anxiety levels, the system enables real-time, non-invasive mental health screening, which is particularly valuable in preventive psychological care. Extensive experiments were conducted on a diverse dataset of 1,504 instances, covering seven fundamental emotions relevant to anxiety inference. The proposed model achieved an impressive accuracy of 98.01%, with strong supporting metrics such as F1-score (97.87%), precision (98.05%), specificity (99.66%), and Cohen's Kappa (0.9186). The accuracy and loss curves show that the training convergence patterns are stable and can be used to make predictions about new data. When compared to traditional methods and existing benchmark datasets, the ANet framework shows that it works better and is more stable. This system is different from traditional diagnostic tools that depend on subjective reporting or clinical interviews. It is objective, automated, and scalable, and it can help clinicians intervene early. The good results show that deep learning and facial expression analysis can be used to find anxiety. The ANet model is a big step forward in the field of affective computing, and it could also be used in bigger mental health assessment systems like telemedicine platforms or digital mental health apps. We want to make the ANet framework more scalable and applicable to more situations by looking at bigger and more varied datasets in future research. We also want to look into how to combine different types of data, like audio and physiological signals, to make anxiety detection systems even more accurate and reliable. The ANet framework is a big step towards meeting the urgent need for objective and automated ways to detect anxiety. Such systems will make it possible for mental health care to offer personalised and timely interventions.

Future research will concentrate on enhancing the ANet framework's robustness and scalability through the use of larger and more varied datasets to improve population-to-population generalisation. To capture more detailed anxiety-related patterns, multimodal fusion—which combines facial analysis with audio and physiological signals—will be investigated. To enhance performance in practical settings, cutting-edge deep learning architectures and transfer learning strategies will be researched. Additionally, privacy-preserving strategies will be used to guarantee ethical and secure deployment in clinical and remote healthcare applications, and explainable AI techniques will be incorporated to improve interpretability.

References

[1] J. Almeida and F. Rodrigues. Facial expression recognition system for stress detection with deep learning. In *Proceedings of the 23rd International Conference on Enterprise Information Systems*, pages 256–263, 2021.

[2] F. Ayache and A. Alti. Performance evaluation of machine learning for recognizing human facial emotions. *Revue d'Intelligence Artificielle*, 34(3):267–275, 2020.

[3] S. H. Babu, S. A. Birajdhar, and S. Tambad. Face recognition using entropy based face segregation as a pre-processing technique and conservative bpso based feature selection. In *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing*, pages 1–8, 2014.

[4] G. Ballikaya and D. Kaya. Facial expression recognition techniques and comparative analysis using classification algorithms. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 12(3):596–607, 2023.

[5] N. Banskota, A. Alsadoon, P. W. C. Prasad, A. Dawoud, T. A. Rashid, and O. H. Alsadoon. A novel enhanced convolution neural network with extreme learning machine: facial emotional recognition in psychology practices. *Multimedia Tools and Applications*, 82(5):6479–6503, 2023.

[6] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 568–573, 2005.

[7] N. R. Calvo-Ariza, L. F. Gómez-Gómez, and J. R. Orozco-Arroyave. Classical fe analysis to classify parkinson's disease patients. *Electronics*, 11(21):3533, 2022.

[8] W. T. Chew, S. C. Chong, T. S. Ong, and L. Y. Chong. Facial expression recognition via enhanced stress convolution neural network for stress detection. *IAENG International Journal of Computer Science*, 49(3):1–10, 2022.

[9] L. Fontes, P. Machado, D. Vinkemeier, S. Yahaya, J. J. Bird, and I. K. Ihianle. Enhancing stress detection: A comprehensive approach through rppg analysis and deep learning techniques. *Sensors*, 24(4):1096, 2024.

[10] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, and M. Tsiknakis. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31:89–101, 2017.

[11] A. Graves, C. Mayer, M. Wimmer, J. Schmidhuber, and B. Radig. Facial expression recognition with recurrent neural networks. In *Proceedings of the International Workshop on Cognition for Technical Systems*, 2008.

[12] R. Karthika and L. Parameswaran. Study of gabor wavelet for face recognition invariant to pose and orientation. In *Proceedings*, pages 501–509. 2016.

[13] S. Latif, H. S. Ali, M. Usama, R. Rana, B. Schuller, and J. Qadir. Ai-based emotion recognition: Promise, peril, and prescriptions for prosocial path. *arXiv preprint arXiv:2211.07290*, 2022.

[14] J. R. Lee, L. Wang, and A. Wong. Emotionnet nano: An efficient deep convolutional neural network design for real-time facial expression recognition. *Frontiers in Artificial Intelligence*, 3, 2021.

[15] S. Lin. Deep learning research and how to get immersed - towards data science. `https://towardsdatascience.com`, 2021. Accessed on March 27, 2022.

[16] M. A. Lumley, J. L. Cohen, G. S. Borszcz, A. Cano, A. M. Radcliffe, L. S. Porter, H. Schubiner, and F. J. Keefe. Pain and emotion: a biopsychosocial review of recent research. *Journal of Clinical Psychology*, 67(9):942–968, 2011.

[17] S. B. Luo, K. T. Nguyen, X. D. Jiang, J. F. Wu, B. H. Wen, Y. Zhang, G. Chierchia, H. Talbot, T. Bourouina, D. L. Kwong, and A. Q. Liu. A high performance of single cell imaging detection with deep learning. In *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, pages 356–360, 2019.

[18] V. Maydych, M. Claus, C. Watzl, and T. Kleinsorge. Attention to emotional information is associated with cytokine responses to psychological stress. *Frontiers in Neuroscience*, 12, 2018.

[19] A. Mehrabian and S. R. Ferris. Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 31(3):248–252, 1967.

[20] A. Mostafa, M. I. Khalil, and H. Abbas. Emotion recognition by facial features using recurrent neural networks. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 417–422, 2018.

[21] T. Eswara Prasad. An efficient facial expression recognition system using novel image classification by comparing cnn over res net. *Journal of Pharmaceutical Negative Results*, 13, 2022.

[22] K. U. Singh, A. Kumar, G. Kumar, T. Singh, S. Kumar, and S. P. Yadav. An autonomous emotion recognition strategy employing deep learning for self-learning. In *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pages 883–888, 2023.

[23] N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. Casey, and C. Nelson. The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3):242–249, 2009.

[24] C. Viegas, S.-H. Lau, R. Maxion, and A. Hauptmann. Towards independent stress detection: A dependent model using facial action units. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, 2018.

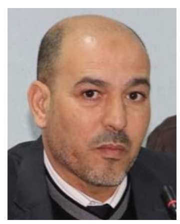
[25] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages 511–518, 2001.

[26] Y. Wu and W. Qiu. Facial expression recognition based on improved deep belief networks. *AIP Conference Proceedings*, 020130, 2017.

[27] H. Yang and X. A. Wang. Cascade classifier for face detection. *Journal of Algorithms & Computational Technology*, 10(3):187–197, 2016.

Authors

Houcem REZAIGUIA is a Ph.D. student and researcher in computer science at Mohamed Khider University, Biskra, Algeria. He is affiliated with the LESIA research laboratory, where his work focuses on automatic anxiety detection from facial expressions using deep learning. His research interests include affective computing, computer vision, and artificial intelligence applied to emotional state analysis.



Abdelhamid DJEFFAL is a Professor in the Department of Computer Science at Mohamed Khider University, Biskra, Algeria. He supervises several research projects in the field of artificial intelligence and signal processing, with a particular interest in emotion recognition, machine learning, and intelligent systems. He is also a member of the LESIA research laboratory.