



AI-Driven Prediction of Performance Risks and Team Dynamics in Information Technology Projects

Prepared By

Name: Ammar Mohammad Wassas

ID No.:253002052

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Mohamed Essa

Collage of Informatics, Midocean University, Comoros

Email: 253002052@st.midocean.edu.km

Abstract

The approach presented here is a solution that accounts for the increasing intricacies involved in delivering IT projects and the problem of predicting the performance risk factors within IT project teams by analyzing the behavior and the state of the individuals within these groups.

Conventionally, project management has used manual reporting tools that have the disadvantages of being subjective and having long lead times before the problems occur. To overcome these shortcomings, this study uses artificial intelligence and computer vision technologies, with the help of YOLOv8 and CNN-based emotion recognition tools. These tools use images drawn from the DAiSEE and AffectNet standard image databases to generate data from the workspace snapshots.

The expected outcomes include a high precision score for the proposed risk prediction algorithm that can recognize burnout and disengagement in individuals before they affect the project goals.

Table of Contents

Abstract.....	2
1. Chapter 1 Introduction	7
1.1 Overview.....	7
1.2 Problem Definition.....	11
1.3 Thesis Objectives	11
1.4 Thesis Contribution	12
1.5 Thesis Organization	12
2. Chapter 2 Background and Related Work.....	13
2.1 Background.....	13
2.2 Related Work	20
2.3 Related Work Comparison	25
3. References.....	26

List of Abbreviations

- **AEMT**: Attention-Enhanced Multi-Layer Transformer
- **AI**: Artificial Intelligence
- **ANN**: Artificial Neural Networks
- **ASF**: Attentional Selective Fusion
- **AU**: Action Unit (related to FACS)
- **CBAM**: Convolutional Block Attention Module
- **CNN**: Convolutional Neural Network
- **DAiSEE**: Dataset for Affective States in E-Environments
- **DL**: Deep Learning
- **EGT**: Enhanced GhostNet with Transformer Encoder
- **EM-UDA**: Emotion Detection Using Unsupervised Domain Adaptation
- **FACS**: Facial Action Coding System
- **FER**: Facial Expression Recognition
- **GAT**: Graph Attention Network
- **HOG**: Histograms of Oriented Gradients
- **IT**: Information Technology
- **LBP**: Local Binary Patterns
- **ML**: Machine Learning
- **MTE**: Multi-Layer Transformer Encoder
- **PMO**: Project Management Office
- **QWTR**: Quaternion Wavelet Transformer
- **ReLU**: Rectified Linear Unit
- **ResNet**: Residual Network
- **SE**: Squeeze-and-Excitation (related to ResEmoteNet)
- **ViT**: Vision Transformer
- **W-SCAM**: Window-based Self-Correction Attention Mechanism (related to POSTER++)
- **YOLO**: You Only Look Once (referring to YOLOv8)

List of Figures:

Figure 1.1: The Triple Constraint.....	7
Figure 1.2: Watermelon Effect	8
Figure 1.3: A typical neural network architecture	9
Figure 1.4: Predictive academic performance analysis workflow	9
Figure 2.1: Recognizing human facial expressions with ML	13
Figure 2.2: AI vs ML vs DL	14
Figure 2.3: Visualization of representative AUs From FACS.....	15
Figure 2.4: Histograms Of Oriented Gradients.....	16
Figure 2.5: CNN-based pipeline for FER.....	17
Figure 2.6: ResNet - Deep Learning.....	17
Figure 2.7: A GhostNet-Based Two-Stage Transformer.....	18
Figure 2.8: Architecture of the base vision-transformer model	19
Figure 2.9: Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network - Quang Tran Ngoc et al., 2020.....	19

List of Tables:

Table2.1: Comparative Analysis of State-of-the-Art FER Models on the AffectNet Dataset.....	25
---	----

1. Chapter 1 Introduction

1.1 Overview

The main aim of this study is to design an empirical framework to predict performance risks using real-time observation of team dynamics in the IT setting. The reason for pursuing this objective is that there is a dire need to shift from qualitative human factors analysis to quantitative human factors assessment in modern project management practice.

Until now, project status evaluation was limited to lagging metrics, which document project failure post-factum by tracking cost overruns, schedule deviations, and achieved milestones. According to Holzmann et al. (2022), existing approaches to project management cannot capture the "Human Factor" in real time, thus leaving room for unnoticed fatigue and conflicts among team members until they miss an important milestone [1].

The background of the topic can be traced back to the 20th century, where the main emphasis in project management was laid on the “Triple constraint” (scope, time, and cost).



Figure 1.1: The Triple Constraint: The Project Management Triangle of Scope, Time, and Cost

Although risk management was formally incorporated into PMO standards during the 1990s and 2000s, these practices were still linked to traditional reporting systems and audits. As Felicetti et al. (2024) have posited, such mechanisms can be prone to reporting bias, where the team or management tends to manipulate information and represent an optimistic situation to the stakeholders [2]. Such practice leads to the so-called watermelon effect, where the project looks good from the outside but fails from within.

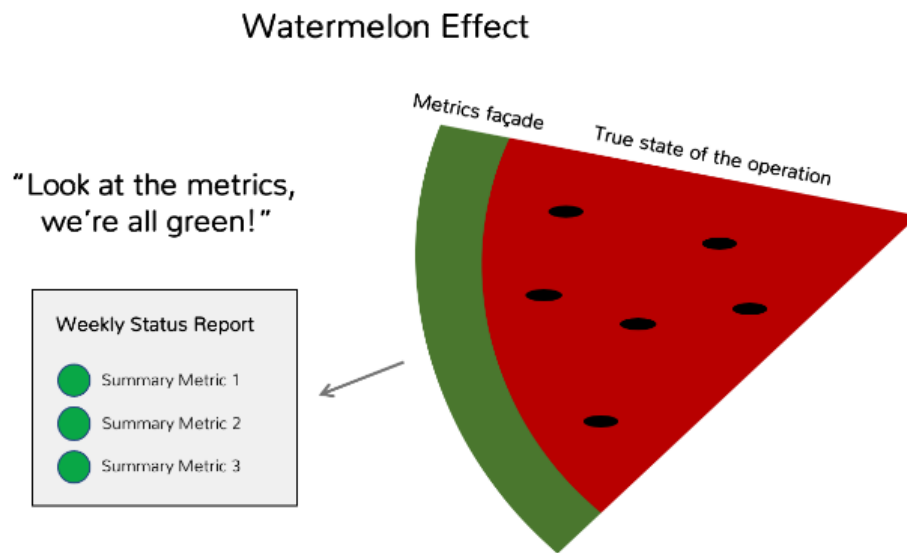


Figure 1.2: Watermelon Effect - Beware of the Watermelon by Nicholas Hartman

As the complexity of IT projects increased along with their adoption of agile project management approach, there was a dire need for leading indicators that can help forecast future performance. Unfortunately, existing indicators cannot determine any precise changes, such as reduction in engagement or increase in stress levels, which are indicative of poor performance in the near future. The proposed study leverages recent developments in the area of Artificial Neural Networks (ANN). It is important to mention that ANN has been selected owing to its unique capability of dealing with high dimensional nonlinear problems [3].

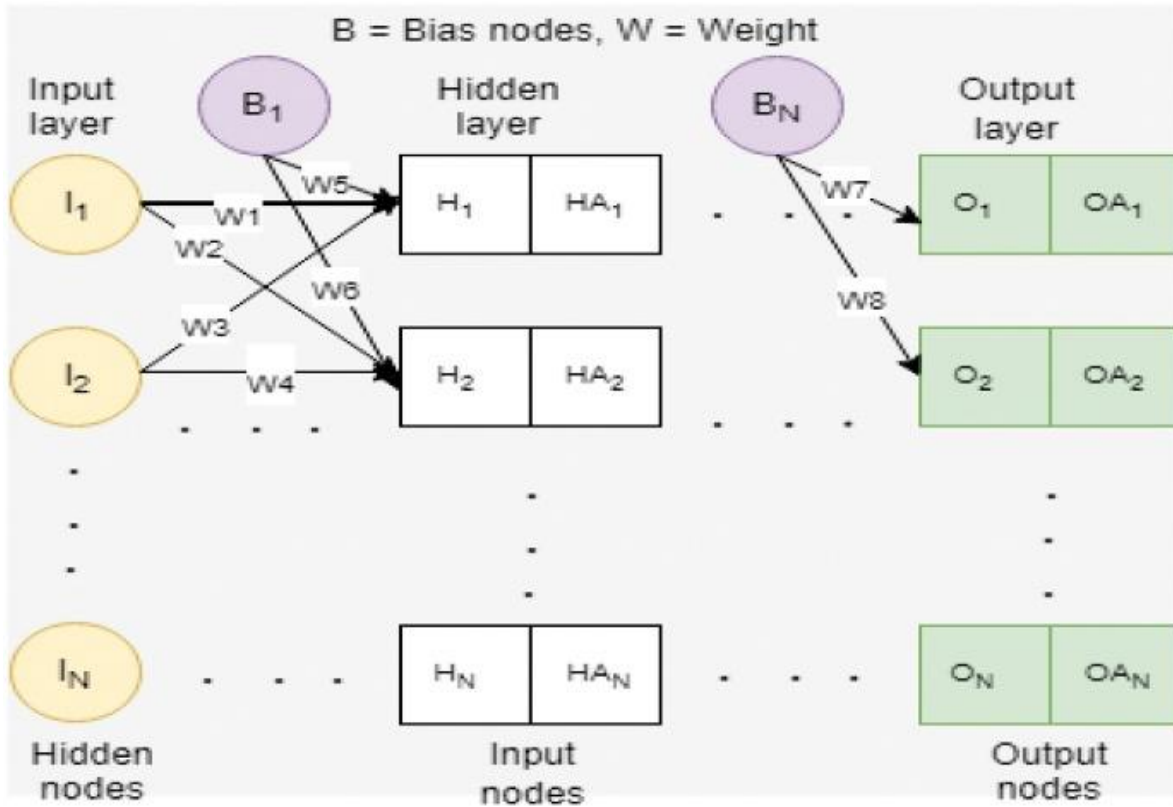


Figure 1.3: A typical neural network architecture by Abiodun et al. (2018)

The use of predictive data mining, as seen through the work done by Abu Saa et al. (2019), makes it possible to determine certain behavioral variables that are correlated with performance results [4]. Following the success of using an "ANN approach," as was done by Adewale et al. (2018) to predict academic performance, this paper seeks to model the same in IT project management scenarios to predict project overrun problems [5].

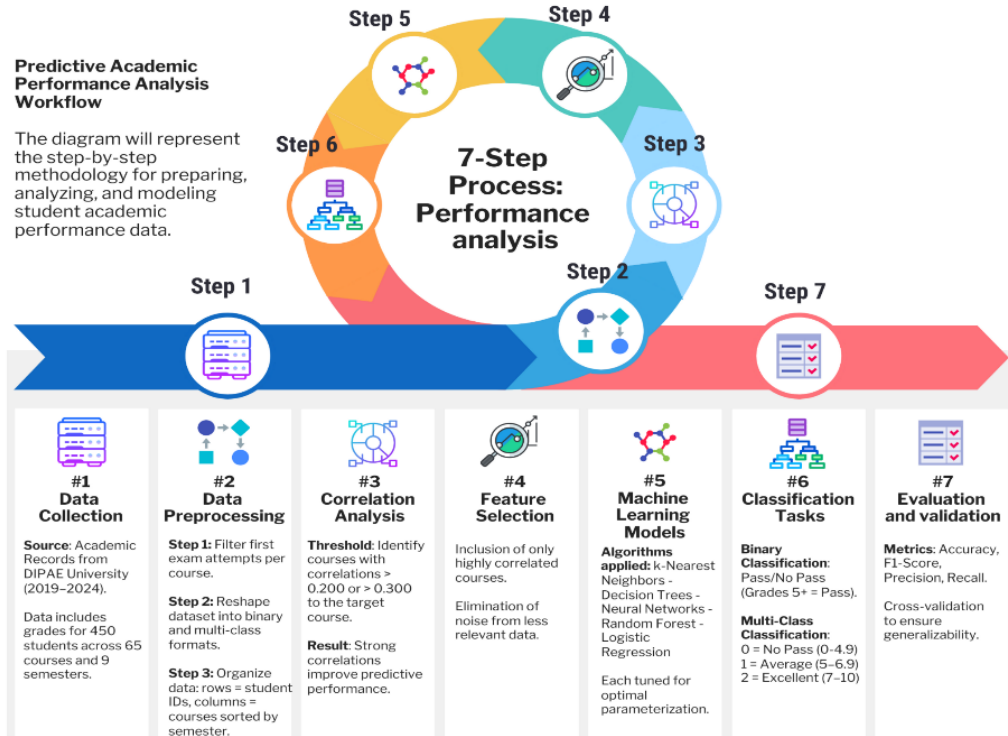


Figure 1.4: Workflow of predictive performance analysis

The present study advocates for a basic paradigm shift in the approach of logging behaviors from being subjective to becoming objective through images of behavior. Through Computer Vision that transforms pictures captured at intervals in workspaces into useful information, the solution offers a means for identifying risks related to human factors at the very early stages of development. In contrast to video surveillance that consumes considerable computing resources, the methodology involves the use of low-complexity techniques for evaluating image frames, such as Yolov8. Finally, the end result is a risk assessment system based on factual data.

1.2 Problem Definition

The current approach to managing IT projects is hampered by a major “Visibility Gap,” requiring managers to depend on subjective self-reports for insights into the health and performance of their teams. This results in the late discovery of human capital risks such as exhaustion and team discord, which become evident only after a project’s deadlines have been missed.

Although modern work environments are equipped with imaging tools, no automatic artificial intelligence system exists that can analyze an image frame and produce relevant information for management purposes.

The purpose of this study is to create a diagnostic imaging solution that gives objective feedback about team states based on imagery, using minimal computing resources.

1.3 Thesis Objectives

- To create an automated pipeline to detect behavioral states within the team and interactions in the workspace.
- To develop a module that helps in quantifying team emotions, particularly focusing on their level of engagement and stress.
- To design a Risk Scoring Algorithm which correlates behavioral states with project metrics to detect any future challenges.
- To test the effectiveness of the proposed framework through the use of high-resolution image data sets, such as AffectNet and DAiSEE video frames.

1.4 Thesis Contribution

This study contributes to the disciplines of AI Application and Project Management in two distinct layers:

➤ **Technical Contribution:**

Development of an analysis pipeline involving multiple stages in which team States (such as Focus, Stress, or Collaboration) are recognized based on environmental snapshots.

➤ **Managerial Contribution:**

An “Evidence-Based PMO Framework” that detects human-based risk factors based on objective evidence rather than subjective reporting.

1.5 Thesis Organization

The organization of this research is aimed at offering a smooth transition from theory to practice:

➤ Chapter 2:

Reviews the theoretical foundation of study and related fields of research.

➤ Chapter 3:

Proposes the methodology, “the technical architecture of the framework.”

➤ Chapter 4:

Analyzes the results of the experiment.

➤ Chapter 5:

Concludes the research and identifies directions for further investigation.

2. Chapter 2 Background and Related Work

2.1 Background

2.1.1 Introduction to Facial Expression Recognition (FER):

FER is an advanced form of affective computing and computer vision, designed to connect human emotions with machine intelligence. Although earlier studies were limited to recognizing emotions such as happiness or anger, current systems attempt to identify more complex psychological states, including frustration, fatigue, and cognitive stress, which play an essential role in psychological and occupational well-being [6,7].

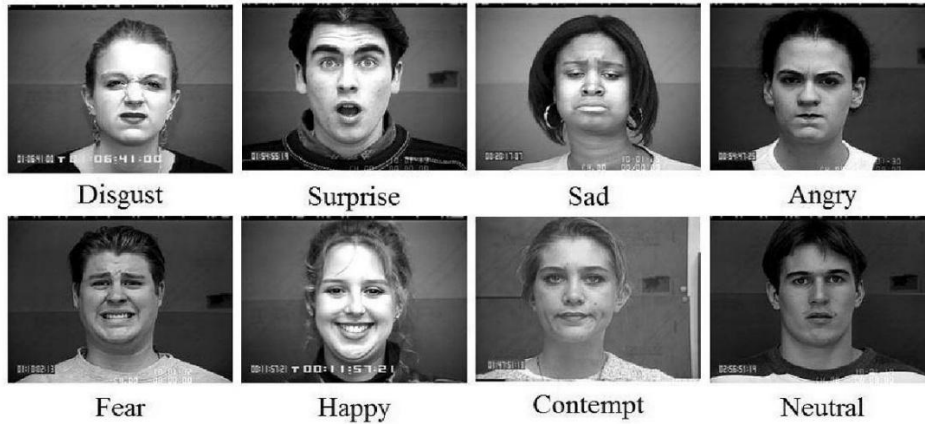


Figure 2.1: Recognizing human facial expressions with ML – Thoughtworks

The initial problem in FER is that the emotions are not only internal but external and expressed by non-linear and physical alterations in one's face appearance, which differ dramatically among individuals and cultural groups [8]. It is easy to assess this phenomenon in laboratory conditions, yet in the “in-the-wild” environment of, for instance, a dynamic IT PMO, such challenges as poor visibility, mask wearing, or the movement of one's head can complicate standard solutions [9]. The progress in this discipline can be traced through three fundamental breakthroughs: from traditional methods of counting features to deep learning, and further to contemporary approaches based on Attention and Transformers capable of perceiving the whole face context.

2.1.2 Machine Learning vs. Deep Learning:

The architecture of this research can only be explained by identifying the different hierarchies within AI. Despite the interchangeability of these terminologies, they represent diverse levels of computational complexity and autonomy.

ML is a subfield of AI that employs algorithms and data to mimic human intelligence and learn from them. In the conventional sense, there is always a requirement for humans to do "feature extraction." In conventional ML, there is a need for human experts who have to manually instruct the computer to examine certain features, such as the length of the mouth or eyebrows. Mathematical models, such as Support Vector Machines or Random Forests, are then applied to classify the information.

Deep Learning is a more refined version of ML based on Artificial Neural Networks with numerous layers. The major distinction is in the process of information handling. Unlike ML, there is no requirement for humans to do feature extraction. The neural network automatically identifies the "features" needed to perceive emotions based on raw pixel images.

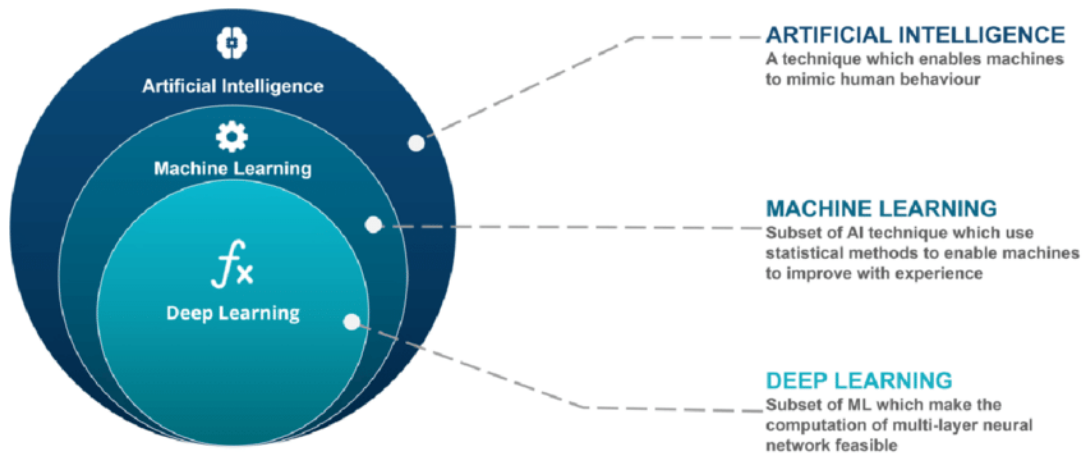


Figure 2.2: AI vs ML vs DL by Prajwal Shrestha

2.1.3 Foundational Techniques: From Manual to Automatic:

The Era of Handcrafted Features:

Prior to the development of Deep Learning techniques, the computer was instructed by hand as to what needed to be detected. Prior to the development of Deep Learning techniques, the researchers would instruct the computer by hand through Geometric Techniques by measuring the distance between particular landmarks, like the eyes and the jawline [10].

To bring consistency to the process, FACS coding technique was utilized where particular muscle movements were attributed to AUs [11]. The expression of emotions was not something abstract, but rather an interaction of several elementary actions, such as AU12 Lip Corner Pull for smiling.



Figure 2.3: Illustration of representative Facial AUs and corresponding muscle movement from the FACS

On the other hand, Appearance-Based Methods focused on the “texture” of skin, and utilized techniques such as HOG and LBP to detect the shadows and lines formed due to smiling or frowning faces [12,13]. Though these methods were quick and simple to comprehend, they were not effective in practical office settings as they could not deal with the “noisy” environment of poorly lit office spaces and low-quality webcams [14].

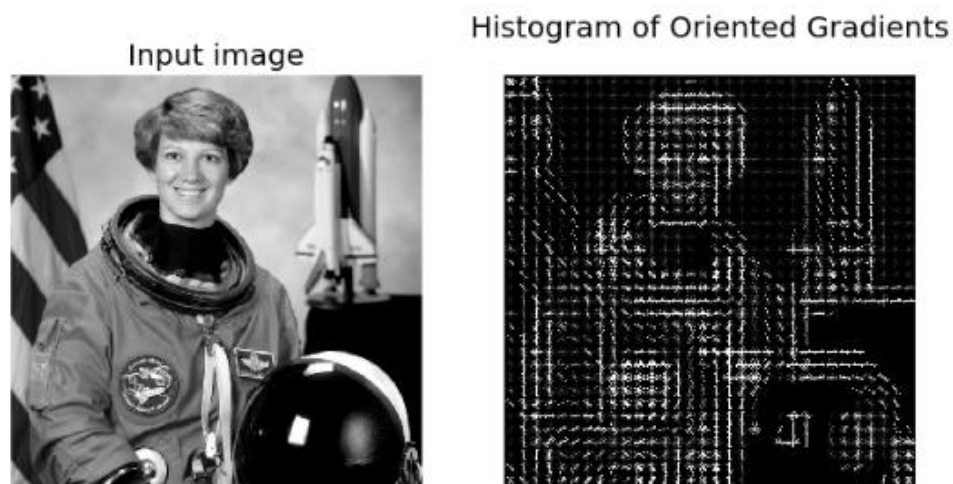


Figure 2.4: Histograms Of Oriented Gradients – By Navneet Dalal et al. (2005)

The Deep Learning Revolution:

That was until the invention of CNNs, in which the definition of “smile,” among other facial expressions, is learned from large amounts of pictures. This works due to convolutional layers that recognize simple facial elements, such as edges, and more complex layers that combine them to create even more complex patterns like a frown [15]. The use of an Activation Function such as ReLU in processing complex data and Pooling Layers that simplify images made CNNs a popular approach to emotional AI [16].

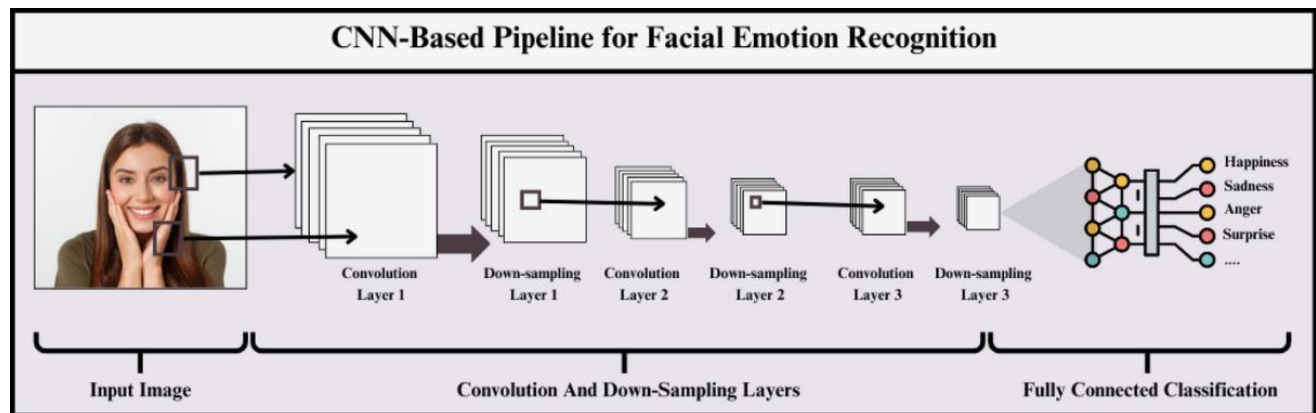


Figure 2.5: CNN based approach for facial expression recognition: an in-depth study of state-of-the-art models, techniques and data sets

2.1.4 Modern Architectures for Real-World FER:

Residual Learning (ResNet)

In attempting to train such networks to detect complex emotions, scientists ran into an important challenge related to the "vanishing gradient" problem in deep neural networks where the learning process could not be completed due to the high number of layers in such a network. The use of residual blocks solved the vanishing gradient problem, making it possible to design a deep neural network like ResNet-50.

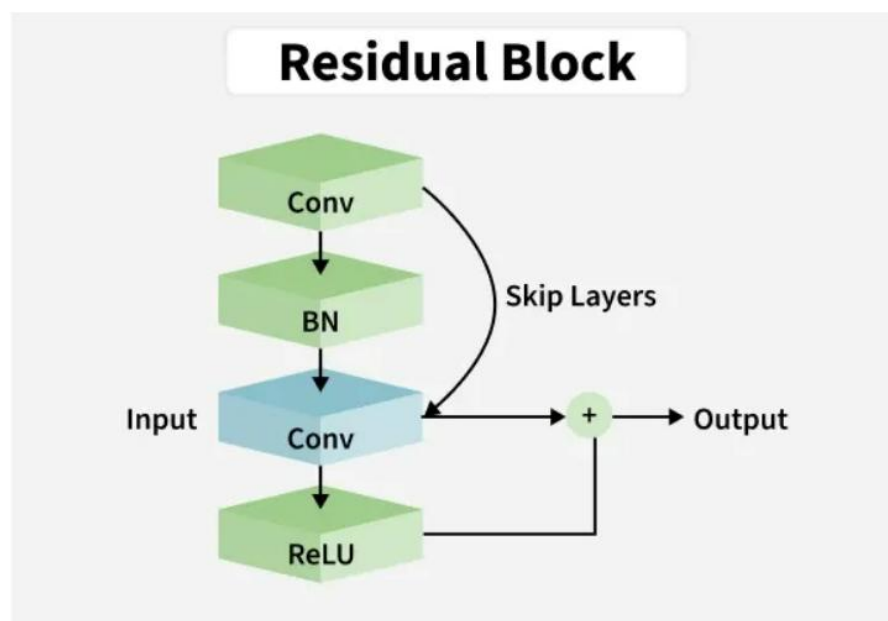


Figure 2.6: ResNet - DL – GeeksforGeeks

Efficiency and Edge Computing:

A PMO dashboard must operate in real-time, therefore it cannot depend on a super computer to work efficiently. Models such as EfficientNet employ compound scaling methods to extract the best performance from hardware [19]. On the other hand, models like GhostNet employ Ghost modules to produce quality feature maps at a relatively lower cost [20]. This ensures that monitoring team wellness is easy on regular office laptops.

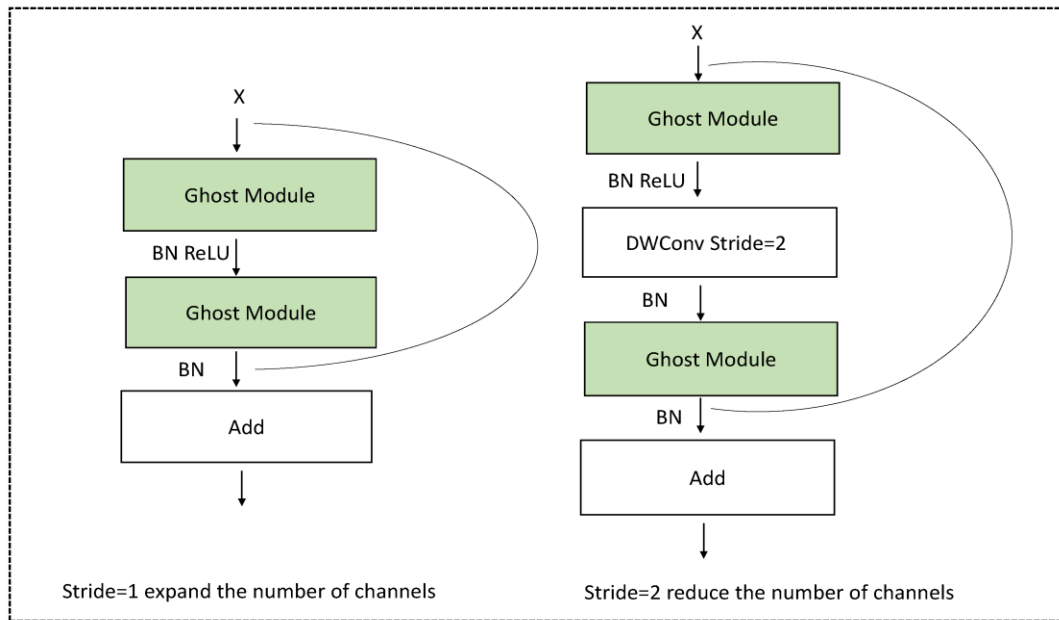


Figure 2.7: Ghostformer: A GhostNet-Based Two-Stage Transformer for Small Object Detection

Attention and Transformers:

The latest breakthrough in FER technology is the implementation of Attention Mechanisms and ViTs. Attention allows for more emphasis to be placed on those facial elements which are key to the classification, thus allowing the algorithm to ignore any superfluous information [21,22]. Whereas in CNNs all information is equally weighted, transformers take into account “the global context”, that is, they can analyze the connection between the movements of the forehead area and the movements in the mouth region [23,24].

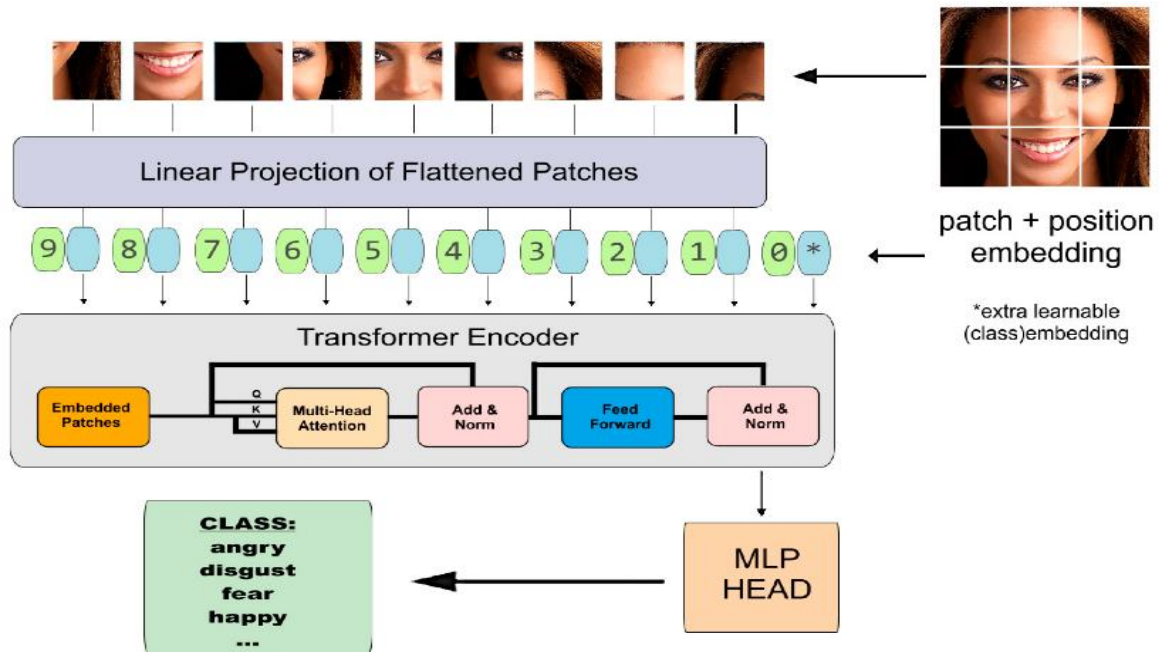


Figure 2.8: Vision-Transformer architecture by: Sukhrob Bobojanov et al. (2023)

Graph-Based Insights and YOLOv8:

Graph-Based Approach, on the other hand, considers a face as a network of interconnected nodes. In this case, using a detector such as YOLOv8 to identify these features will result in a spatial graph that is invariant to head poses [25,26]. This is a crucial step towards accomplishing the objective of this thesis project because it will enable consistent facial expression analysis regardless of the employee's position.

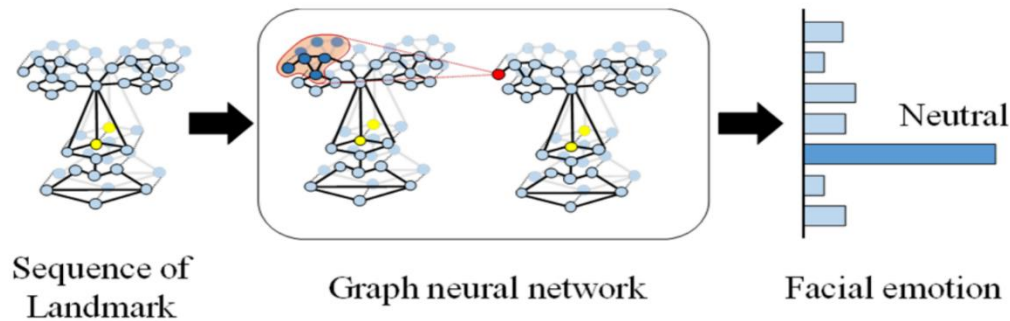


Figure 2.9: Directed Graph Neural Networks for Facial Landmark-based Emotion Recognition: By Quang Tran Ngoc et al. (2020)

2.2 Related Work

Another popular application of ML within computer vision concerns the ability of the machine to "read" human emotions. Scientists have designed many Deep Learning models that attempt to strike a balance between accuracy and computational efficiency but their application to practical issues, like workplace wellness monitoring, still presents challenges. Indeed, an ordinary office environment requires that the system be capable of handling poor lighting conditions, varied head orientations, and individual differences in facial morphology. A number of Deep Learning algorithms have been proposed to address these challenges, attempting to find a balance between accuracy and computational efficiency.

This section provides a review of the key techniques used as building blocks for FER technologies. The research articles discussed in this paper are sorted according to the type of AI architecture used – Residual Networks, Transfer Learning, and Attention mechanisms. For each article, the strengths and weaknesses of the approach proposed will be highlighted. The emphasis will be placed on identifying open questions that have yet to be answered by the scientific community – namely, catching extremely subtle emotion shifts, sustaining high levels of accuracy in real-time streaming environments, and the "Black Box" problem, which occurs when the rationale behind certain AI decisions is unclear. By reviewing the technological landmarks achieved to date, one can better identify the limits of current technology and thus the novel contribution of this thesis.

The AffectNet database was created by Mollahosseini et al. (2017) in order to build an essential benchmark dataset for assessing emotions based on their valences in unrestricted scenarios. Using Deep ResNet, trained on more than 440,000 images, they reached the test accuracy of 58%. One key point about the research conducted by these scientists is the proof provided that large diversity within the data set is crucial for dealing with individual facial differences. The problem with this approach is poor detection of arousal intensity [27].

According to Almeida et al. (2021), Deep Learning and Computer Vision by Transfer Learning with a VGG16 backbone were applied for workplace stress detection on affectNet database through recognition of micro-expression connected with cognitive load for software developers. The outcome showed a prediction rate of 72%. The key advantage of this technique is its capability to measure “unseen” stress, which cannot be determined through self-reporting. One of the major disadvantages of the system is its sensitivity. Sometimes, the algorithm classifies high concentration as “stress,” and without further checks, unnecessary interference occurs [28].

Jain et al. (2024) developed a technique named EM-UDA for unsupervised domain adaptation in order to improve facial emotion recognition in various datasets with few labeled data points. Using adversarial techniques to make feature distribution in source and target domain similar, the model was able to achieve 83.9% accuracy when trained on the combination of the two datasets and tested on CK+. When evaluating on FER2013 dataset, the accuracy was 74.55%. The strength of the paper lies in generalizing the task across multiple datasets without relying on huge amounts of labeled data, which makes it ideal for practical implementation. Nonetheless, one of the weaknesses of the method is the reduction of emotion classes into larger groups like positive or negative emotions [29].

In their work Aly et al. (2025), a mixed deep learning algorithm was developed that merges convolutional neural networks with attention models in order to improve the accuracy of recognizing facial emotions in challenging conditions. With the help of a combination of the extraction of spatial features and contextual learning, this algorithm demonstrated accuracy over 85% on the AffectNet database. The advantages of this paper include the ability to analyze both detailed facial information and overall context. As the disadvantage, one can highlight the computational complexity of the algorithm [30].

A computer vision research was carried out by Ryumina et al. (2022) in order to create a robust Deep Learning Model on the basis of the ResNet-50 framework pre-trained on the VGGFace2 database. Applying the techniques of label smoothing, they used AffectNet Database for fine-tuning and obtained a validation accuracy of 66.4%. An advantage of the work lies in its high level of generalization on different databases. But at the same time, the weakness of the study consists in the use of static image features and ignoring the dynamics of the process under study [31].

In their recent paper, Mao et al. (2023) proposed POSTER++, an efficient architecture using W-SCAM mechanism to address the AffectNet dataset. This approach proved highly effective, having lower costs (8.4G FLOPs) and attaining accuracy of 67.49% for the seven-class problem. The key advantage of this research is the ability to extract multi-scale features with different head poses. Nevertheless, the main drawback of this study is the “Class Imbalance” problem existing within the dataset, making the results less accurate for minority classes such as “Disgust” [32].

Wang, J. (2024) proposed the Attention-Enhanced Multi-Layer Transformer (AEMT), a Deep Learning model to mitigate the effects of motion blur and occlusions in the AffectNet dataset. The model integrates a dual-branch CNN with an ASF module to achieve 71.23% accuracy. The strength of this work is the MTE, which models long-range dependencies. However, a noted limitation was the heavy computational load, which requires a high GPU resource for real-time office deployment [33].

Heidari et al. (2024) presented a ML mechanism to diversify features in deep learning layers for in-the-wild recognition using the AffectNet dataset. The proposed approach reduced feature redundancy, forcing the network to learn a broader range of emotional cues. While it achieved 60.02% accuracy. However, a limitation was the lower performance variance between high-resolution and low-resolution datasets [34].

A novel architecture called ResEmoteNet, which combines Residual blocks with SE blocks, was suggested by Roy et al. (2024) for Deep Learning. The algorithm

was capable of detecting important facial features and minimizing background noise, achieving an accuracy of 72.39%. Nevertheless, it had a drawback in that it heavily depended on initial feature extraction; when the localization of the face was poor, it amplified noise using the Squeeze-and Excitation (SE) block [35].

Xu et al. (2023) presented a transformer-based feature fusion scheme called Global–Local Feature Fusion Transformers, which helped improve the facial expression recognition process in non-constraint conditions. Through the incorporation of both global and local facial information, the algorithm was able to attain a success rate of around 67.46% when applied to the AffectNet database. The contribution of this research work is that it has managed to efficiently recognize difficult facial attributes like pose, occlusion, and lighting. But the drawback of this method is that the computation becomes relatively more complex because of the transformer architecture [36].

Yousafzai et al. (2025) proposed a Multi-scale Simplicial Transformer along with GAT to analyze the AffectNet dataset. The proposed approach makes use of YOLOV8 to create graphs based on the spatial representation of facial features and analyzes them using a Deep Learning transformer, achieving an accuracy of 81.84%. One of the major drawbacks of this approach is that it possesses extremely high mathematical complexity [37].

Ahmad et al. (2025) proposed a Deep Learning framework based on CNN for recognition on the AffectNet dataset in mobile and unconstrained environments. By using pre-processing like greyscale conversion and scaling in the Computer Vision pipeline, they reached 60% accuracy. The power of the framework comes from its practical application for real-time monitoring. One of the drawbacks of the framework is its inability to distinguish fine emotions such as “Fear” and “Neutral” [38].

Zhou et al. (2025) introduced the QWTR, a Deep Learning technique to handle extreme lighting variations in the AffectNet dataset, by using Quaternion

mathematics in the Computer Vision pipeline, they reached 68.37% accuracy. The technique's strength is its mathematical robustness against shadows and artificial yellow lighting. However, its implementation complexity makes it difficult to integrate into standard, non-specialized Python software stacks [39].

A Deep Learning framework that uses ViT for emotion recognition with occluded faces on the masked version of the AffectNet database was proposed by Maghari et al. (2026). The model implemented ML classification re-categorization into five categories to reach an accuracy of 81%. The advantage of this framework is the self-attention feature of this model that can capture context around the eye and forehead regions. However, the major drawback of the high inference time (22ms per image) demands specialized GPU resources [40].

An “Enhanced GhostNet with Transformer Encoder” model was suggested by Sun et al. (2025) as a solution to enhance emotion recognition reliability. This research work combines the advantages of feature extraction provided by GhostNet with the benefits of the Transformer Encoder approach. In addition, the authors incorporated dual attention mechanism into their algorithm. The findings showed that EGT has achieved 85.7% accuracy when working with the AffectNet Dataset. EGT's advantage is its effectiveness due to high accuracy and high computation speed; thus, it can be widely used for real-time surveillance. Nevertheless, there is room for improvement in recognizing micro-expressions in stressful environments [41].

2.3 Related Work Comparison:

In order to present the background behind the proposed framework, Table 2.1 presents a comparative analysis between fifteen major works based on the AffectNet dataset. It is evident from the analysis that the evolution of the technology from simple designs to complex ones has taken place, starting with a model like ResNet from Mollahosseini et al. (2017). Recent developments, especially using transformers, like those developed by Yousafzai et al. (2025) and GhostNet, developed by Sun et al. (2025), indicate that the combination of different technologies can produce an accuracy higher than 80%. Thus, YOLOv8 and ViT have been selected for this study.

Table2.1: Comparative Analysis of State-of-the-Art FER Models on the AffectNet Dataset.

#	Authors	Year	Dataset	Techniques	Accuracy
1	Mollahosseini et al.	2017	AffectNet	ResNet	58.00%
2	Almeida et al.	2021	AffectNet	Transfer Learning (VGG16)	72.00%
3	Jain et al.	2024	AffectNet	unsupervised domain adaptation framework (EM-UDA)	83.9%
4	Aly, M. et al.	2025	AffectNet	Combined CNN and attention mechanisms	85%
5	Ryumina et al.	2022	AffectNet	ResNet-50	66.4%
6	Mao et al.	2023	AffectNet	Window-based Attention (POSTER++)	67.49%
7	Wang	2024	AffectNet	(AEMT (Dual-branch CNN + Transformer)	71.23%
8	Heidari et al.	2024	AffectNet	Feature Diversification	60.02%
9	Roy et al.	2024	AffectNet-7	ResEmoteNet (Residual + SE blocks)	72.39%
10	Xu, R., et al	2023	AffectNet	Global–Local Feature Fusion Transformers	67.64%
11	Yousafzai et al.	2025	AffectNet	Simplicial Transformer + GAT	81.84%
12	Ahmad et al.	2025	AffectNet	CNN (Mobile Optimization)	60.00%
13	Zhou et al.	2025	AffectNet	Quaternion Wavelet Tranformer	68.37%
14	Maghari et al.	2026	AffectNet (Maksed)	Vision Transformer (ViT)	81.00%
15	Sun et al.	2025	AffectNet	Enhanced GhostNet (EGT)	85.70%

3. References

- [1] [Holzmann, V., et al., 2022] Holzmann, V., Zitter, D., & Peshkess, S. (2022). The expectations of project managers from artificial intelligence: A Delphi study. *Project Management Journal*, 53(5), 438–455.
- [2] [Felicetti, A. M., et al., 2024] Felicetti, A. M., et al. “Artificial Intelligence and Project Management”. Elsevier, ScienceDirect. 2024.
- [3] [Abiodun, Oludare Isaac, et al., 2018] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938.
- [4] [Abu Saa, Amjed, et al., 2019] A. Abu Saa, M. Al-Emran, and K. Shaalan, “Factors affecting students’ performance in higher education: A systematic review of predictive data mining techniques,” *Technology, Knowledge and Learning*, vol. 24, pp. 567–598, 2019
- [5] [Adewale, Amoo M., et al., 2018] A. M. Adewale, A. O. Bamidele, and U. O. Lateef, “Predictive modelling and analysis of academic performance of secondary school students: An artificial neural network approach,” *International Journal of Science and Technology Education Research*, vol. 9, no. 1, pp. 1–8, 2018.
- [6] [Picard, R. W., 1997] Picard, Rosalind W. “Affective Computing”. MIT Press. 1997.
- [7] [Owusu, E., et al., 2015] Owusu, Ebenezer, et al. “Facial Expression Recognition-A Comprehensive Review 1.” *International Journal of Technology and Management Research* 2(4), 29-46, 2015.
- [8] [Ekman, P., and Friesen, W. V., 1971] Ekman, Paul, and Wallace V. Friesen. “Constants across cultures in the face and emotion”. *Journal of Personality and Social Psychology* 17(2), 124–129, 1971.
- [9] [Li, S., and Deng, W., 2020] Li, Shan, and Weihong Deng. “Deep Facial Expression Recognition: A Survey”. *IEEE Transactions on Affective Computing* 13(3), 1195–1215. 2020.
- [10] [Tian, Y., et al., 2011] Tian, Yingli, Takeo Kanade, and Jeffrey F. Cohn. “Facial expression recognition”. *Handbook of Face Recognition*, 487–519. 2011.

- [11] [Ekman, P., and Friesen, W. V., 1978] Ekman, Paul, and Wallace V. Friesen. “Facial Action Coding System: A Technique for the Measurement of Facial Movement”. Consulting Psychologists Press. 1978. <https://doi.org/10.1037/t27734-000>
- [12] [Dalal, N., and Triggs, B., 2005] Dalal, Navneet, and Bill Triggs. “Histograms of oriented gradients for human detection”. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1, 886–893. 2005.
- [13] [Ojala, T., et al., 2002] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987. 2002.
- [14] [Churamani, Nikhil, et al., 2023] Churamani, Nikhil, Tolga Dimlioglu, German I. Parisi, and Hatice Gunes. “Continual Facial Expression Recognition: A Benchmark”. arXiv preprint arXiv:2305.06448, 2023.
- [15] [LeCun, Y., et al., 2015] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. Nature 521(7553), 436–444. 2015.
- [16] [Nair, V., and Hinton, G. E., 2010] Nair, Vinod, and Geoffrey E. Hinton. “Rectified linear units improve restricted Boltzmann machines”. Proceedings of the 27th International Conference on Machine Learning (ICML). 2010.
- [17] [He, K., et al., 2016] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition”. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. 2016.
- [18] [Mollahosseini, Ali, et al., 2017] Mollahosseini, Ali, Behzad Hasani, and Mohammad H. Mahoor. “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild”. IEEE Transactions on Affective Computing 10(1), 18–31., 2017.
- [19] [Tan, M., and Le, Q. V., 2019] Tan, Mingxing, and Quoc V. Le. “EfficientNet: Rethinking model scaling for convolutional neural networks”. International Conference on Machine Learning (ICML), 6105–6114. 2019.
- [20] [Han, K., et al., 2020] Han, Kai, Wang, Yunhe, Tian, Qi, Guo, Jianyuan, Xu, Chunjing, and Xu, Chang. “GhostNet: More Features From Cheap Operations”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. <https://doi.org/10.1109/CVPR42600.2020.00165>

- [21] [Hu, J., et al., 2018] Hu, Jie, Li Shen, and Gang Sun. “Squeeze-and-excitation networks”. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7132–7141. 2018.
- [22] [Woo, S., et al., 2018] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. “CBAM: Convolutional block attention module”. Proceedings of the European Conference on Computer Vision (ECCV), 3–19. 2018.
- [23] [Dosovitskiy, A., et al., 2021] Dosovitskiy, Alexey, et al. “An image is worth 16x16 Words: Transformers for Image Recognition at Scale”. International Conference on Learning Representations (ICLR). 2021.
- [24] [Vaswani, A., et al., 2017] Vaswani, Ashish, et al. “Attention is all you need”. Advances in Neural Information Processing Systems (NeurIPS) 30. 2017.
- [25] [Sapkota, Ranjan, and Manoj Karkee, 2025] Sapkota, Ranjan, and Manoj Karkee. “Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition”. arXiv preprint arXiv:2510.09653, 2025.
- [26] [Veličković, P., et al., 2018] Veličković, Petar, et al. “Graph attention networks”. International Conference on Learning Representations (ICLR). 2018.
- [27] [Mollahosseini, A., et al., 2017] Mollahosseini, A., Hasani, B., and Mahoor, M. H. “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild”. IEEE Transactions on Affective Computing 10(1), pp. 18-31. 2017.
- [28] [Almeida, J., and Rodrigues, F., 2021] Almeida, J., and Rodrigues, F. “Facial Expression Recognition System for Stress Detection with Deep Learning”. In Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS). 2021.
- [29] [Jain, P. R., et al., 2024] Jain, Priti R., Quadri, S. M. K., and Khattar, Anuradha. “EM-UDA: Emotion Detection Using Unsupervised Domain Adaptation for Classification of Facial Images”. *IEEE Access*. 2024. <https://doi.org/10.1109/access.2024.3467990>
- [30] [Aly, M., and Alotaibi, N., 2025] Aly, Mohammed, and Nouf Saeed Alotaibi. “A Comprehensive Deep Learning Framework for Real-Time Emotion Detection in Online Learning Using Hybrid Models”. *Scientific Reports*. 2025.

- [31] [Ryumina, E., et al., 2022] Ryumina, E., Dresvyanskiy, D., and Karpov, A. “In Search of a Robust Facial Expressions Recognition Model: A Large-Scale Visual Cross-Corpus Study”. *Neurocomputing* 514, 435-450. 2022.
- [32] [Mao, J., et al., 2023] Mao, J., Rui Xu, Xuesong Yin, Yuanqi Chang, Binling Nie, and Aibin Huang. “POSTER++: A simpler and stronger facial expression recognition network”. *arXiv preprint arXiv:2301.12149*. 2023.
- [33] [Wang, J., 2024] [7] Wang, J. “Evaluation and analysis of visual perception using attention-enhanced computation in multimedia affective computing”. *Frontiers in Neuroscience* 18, 1449527. 2024.
- [34] [Heidari, N., and Iosifidis, A., 2024] Heidari, N., and Iosifidis, A. “Learning Diversified Feature Representations for Facial Expression Recognition in the Wild”. In *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*. 2024.
- [35] [Roy, A. K., et al., 2024] Roy, A. K., Hemant Kumar Kathania, Adhitiya Sharma, Abhishek Dey, and Md. Sarfaraj Alam Ansari. “ResEmoteNet: Bridging Accuracy and Loss Reduction in Facial Emotion Recognition”. *arXiv preprint arXiv:2409.10545*. 2024.
- [36] [Xu, R., et al., 2023] Xu, Rui, Huang, Ailin, Hu, Yafei, and Feng, Xiaobai. “Global–Local Feature Fusion Transformers for Facial Expression Recognition in the Wild”. *Image and Vision Computing*. 2023. <https://doi.org/10.1016/j.imavis.2023.104824>
- [37] [Yousafzai, S., et al., 2025] Yousafzai, S. N., Nasir, I. M., Saidani, O., Ghodhbani, R., Gu, Y., Syafrudin, M., and Fitriyani, N. L. “A multi-scale simplicial transformer with graph attention for facial emotion recognition”. *Ain Shams Engineering Journal* 16(10), 103584. 2025.
- [38] [Ahmad, A., et al., 2025] Ahmad, A., Singh, V., and Upreti, K. “Emotion Recognition Through Facial Expressions: A Machine Learning Perspective in Mobile Multimedia”. *Journal of Mobile Multimedia* 21(1), pp. 87–111. 2025.
- [39] [Zhou, Yu, et al., 2025] Zhou, Yu, Jialun Pei, Weixin Si, Jing Qin, and Pheng-Ann Heng. “Delving Into Quaternion Wavelet Transformer for Facial Expression Recognition in the Wild”. *IEEE Transactions on Multimedia* 27, 3895–3909, 2025.
- [40] [Maghari, A. Y., and Telbani, A. M., 2026] Maghari, A. Y., and Telbani, A. M. “Facial emotion recognition under face mask occlusion using vision transformers”. *International Journal of Electrical and Computer Engineering (IJECE)* 16(1), pp. 395-403. 2026.

[41] [Sun, J., et al., 2025] Sun, Jie, Tianwen Xu, and Yao Yao. “An enhanced GhostNet model for emotion recognition: leveraging efficient feature extraction and attention mechanisms”. *Frontiers in Psychology* 15, 1459446. 2025.