Comparative Analysis of Selected Facial Emotion Recognition Methods as Potential Tool for Emotional Disorders Automated Assessment in Terms of Remote Work Application

Ilona Zimnoch[1], Aleksandra Rokita[2], Magdalena Matuła[2], Wiktor Kuczek[2], Michał Rydzik[2], Mateusz Pomianek[2]

[1]University of Economics and Human Sciences in Warsaw, Poland
[2]Rzeszow University of Technology, Department of Computer and Control Engineering, Poland

Abstract

In recent years, the growing prevalence of mental health issues and emotional disorders—partly associated with remote work and social isolation—has led to increased interest in Facial Emotion Recognition (FER) methods as tools for the early detection of affective disturbances. Assessing the effectiveness of these methods may support the development of solutions aimed at identifying reduced psychological well-being among employees, allowing for timely intervention and professional support.

This publication provides a review of the most frequently used FER techniques based on visible-spectrum imaging systems. The selected methods were implemented and empirically compared using the publicly available FER-2013 dataset. The analysis emphasizes key performance parameters and metrological aspects, with particular attention to their potential applications in psychological, medical, and research settings. Laboratory tests allowed for the identification of practical strengths and limitations of each method, offering a basis for considering their integration into mental health assessment tools and suggesting directions for future research.

Keywords: affective processing, machine vision, emotion recognition, image processing

Introduction

We live in an era where technological advancement often outpaces our adaptive capacities, and automation is permeating increasingly intimate aspects of human experience — including emotions. More and more frequently, we are asking not only whether it is possible, but whether it is appropriate to technologically measure something as elusive as human emotional states. When emotions become data and facial expressions are reduced to clusters of pixels analyzed by algorithms, a deeper reflection on the boundaries of technological intervention in psychological well-being becomes essential. One of the most recent and rapidly evolving areas within this domain is Facial Emotion Recognition, which combines elements of computer science, psychology, and artificial intelligence. FER has applications in medicine, education, and the workplace — especially in contexts where direct human contact is limited, as is often the case in remote work models. In such settings, technology begins to act as an intermediary — not only

recording, but also potentially interpreting emotional states. The use of FER as a tool to support the assessment of remote employees' psychological well-being seems particularly relevant in light of increasing reports about the emotional consequences of professional isolation. The key question, however, is not only whether FER works, but how it works — and whether it works well enough to be trusted in the sensitive area of mental health.

The aim of this article is to compare selected FER methods — both classical approaches based on handcrafted feature extraction and modern deep learning techniques — in terms of their potential usefulness for monitoring emotions in remote work environments. For this purpose, selected models were analyzed using the publicly available FER-2013 dataset, which serves as a standard benchmark in emotion recognition research. This study seeks not only to evaluate the effectiveness of specific technological approaches but also to reflect on their potential applications in psychological and organizational practice. By comparing the performance of classical and deep learning-based FER methods, we examine not only their classification accuracy but also their robustness to noise, implementation feasibility, operational transparency, and ethical implications. The following sections present a theoretical overview of psychological well-being in the context of remote work, a detailed review of FER methods, a description of the materials and methodology, results of the comparative analysis, and a critical discussion of the opportunities and limitations of using such tools to assess employees' emotional functioning. Special attention is given to the ethical dimensions of emotional automation — including concerns related to privacy, surveillance, and cultural diversity.

Psychological Well-Being in the Context of Remote Work

In recent years, there has been a sharp increase in interest in the mental health of employees, largely driven by the widespread adoption of remote work and the accompanying social isolation. The COVID-19 pandemic led to a mass transition to remote work, which, although initially perceived as a temporary solution, has taken on a long-term character in many organizations. However, numerous studies indicate that this mode of work can significantly affect individual psychological well-being. A study conducted among healthcare workers found that feelings of loneliness and professional isolation correlate with decreased mental well-being, even in the presence of social support (O'Hare, 2024: 4). Similarly, Brown and Leite emphasized that the lack of daily interactions in the professional environment exacerbates stress levels and decreases employee engagement (Brown & Leite, 2023: 144). Becker et al. demonstrated that a key risk factor for the psychological well-being of remote workers is low control over tasks, which contributes to a sense of loneliness (Becker et al., 2022: 453). In light of these challenges, there is a growing need to implement technological tools that support the early detection of symptoms related to declining mental well-being and enable timely preventive and intervention measures.

In order to understand the concept of well-being, it is essential to recognize that it is a psychological construct rooted in the tradition of positive psychology, which developed at the turn of the 20th and 21st centuries. Its theoretical grounding in this perspective implies that it is not limited to the absence of suffering or neutral functioning, but rather encompasses a qualitative dimension of life experience—centered on meaning, engagement, interpersonal relationships, and opportunities for growth. In contrast to the notion of happiness, which is often associated with hedonistic

pleasure and a transient emotional state, psychological well-being refers to a relatively stable and multidimensional form of human functioning that integrates both emotional and existential components (Baes, Speagle & Haslam, 2022: 1–2). The psychological literature offers numerous models describing well-being, highlighting its complexity and multidimensional nature. One of the most significant and influential frameworks is the PERMA model, developed by Martin Seligman, the founder of positive psychology. This model identifies five core components of well-being: Positive emotions, Engagement, Relationships, Meaning, and Accomplishment. Each of these elements can independently affect an individual's quality of life, and the development of each contributes to sustained well-being (Seligman, 2011: 47–49). Another notable approach is Carol Ryff's six-dimensional model, which adopts a more eudaimonic perspective on psychological well-being. According to Ryff, well-being is grounded in aspects such as autonomy, self-acceptance, personal growth, purpose in life, environmental mastery, and positive relations with others. This model integrates psychological maturity, life meaning, and relational quality as essential pillars of mental health (Ryff, 1989: 1072–1075). In the Polish context, a particularly compelling concept is Janusz Czapiński's *onion theory of happiness,* which—despite its reference to *happiness*—in essence, describes well-being. Czapiński distinguishes three layers: a deep (core) layer based on the biological will to live, a middle layer involving emotional reactions to everyday experiences, and an outer layer comprising evaluations of satisfaction with various life domains (Czapiński, 2004: 19–22). Each of these models offers a unique perspective on the understanding of well-being, allowing for recognition of its diverse sources and underlying mechanisms.

An increasing body of research confirms that the well-being of remote workers constitutes a significant challenge in the modern world of work. By utilizing reliable diagnostic tools, such as the PERMA Profiler and Carol Ryff's model of psychological well-being, researchers are increasingly identifying decreased levels of well-being among individuals working outside traditional office environments. A study conducted among lecturers at private universities revealed a marked decline in the quality of interpersonal relationships, sense of meaning, and autonomy in those operating exclusively online (Leong, 2022: 19). Similar conclusions were drawn in a study of employees from various industries, where a comparative analysis of the pre-pandemic and pandemic periods revealed declines in PERMA components such as positive emotions, relationships, and accomplishments (Pataki-Bittó & Kun, 2022: 331). These findings were further confirmed in research on teachers working remotely, who reported lower well-being scores as measured by the PERMA Profiler, particularly in the relational and social domains (Berry, 2023: 70). Importantly, this issue is also present in both technological and industrial sectors. For instance, remote workers in the Finnish IT sector reported a decrease in engagement and sense of meaning (Lampinen, 2024: 39), while in South African mining operations, significant deficiencies in relationships and positive emotions were identified (Kau & Flotman, 2025: 9). Even in more specific contexts such as the architecture sector, women working remotely reported lower levels of accomplishment and social connection, both key components of the PERMA model (Rodríguez-Leudo & Navarro-Astor, 2024: 7). Similar patterns were observed in multinational corporate environments, where women experienced a diminished sense of belonging while working remotely (Nozari & Seyedsalehi, 2024: 23).

Studies on well-being also include managerial staff, indicating a strong correlation between the mode of work and the quality of psychological functioning. The

highest levels of well-being were observed among individuals working in a hybrid model, and the lowest among those working fully remotely. Particularly adverse outcomes were noted among remote leaders in the non-profit sector (Zimnoch, 2024: 53–59), which may be attributed to the combination of factors such as limited organizational support, high responsibility with scarce resources, and difficulties in maintaining social relationships. Given that work modality impacts well-being even at the managerial level, its influence on lower-level employees—with less autonomy and control over working conditions—may be even more profound. The breadth of industries examined, the consistency of findings, and the use of standardized research tools all support the conclusion that diminished well-being among remote workers is not an isolated phenomenon but a serious systemic issue.

With the growing prevalence of remote work, the importance of supporting employee well-being—including for those without daily contact with colleagues or supervisors—has become increasingly evident. Since monitoring well-being in such conditions can be challenging, there is a growing interest in new tools that can accurately and rapidly detect early signs of psychological decline—particularly in the context of preventing burnout, reduced engagement, or emotional disturbances. One increasingly popular direction involves the use of emotion recognition technologies, including Facial Emotion Recognition, as a potential means of assessing the psychological well-being of remote workers.

Automatic Emotion Recognition (AER) technologies, particularly those based on FER, are gaining traction in psychological contexts, especially in the fields of mental health and patient care. As demonstrated in a systematic review of studies from 2013 to 2023, real-time emotion recognition can significantly enhance diagnostic processes, emotional state monitoring, and therapeutic interventions—both in clinical and home-based environments (Guo et al., 2024: 5–6). Particularly promising are multimodal approaches—combining facial analysis, speech, and physiological signals—which move beyond subjective self-report methods toward more objective and dynamic measurement tools (Guo et al., 2024: 2–4). In a similar vein, systems employing deep learning to analyze facial expressions during remote psychological consultations are being developed to support therapists in making more accurate clinical decisions, thanks to access to real-time emotional feedback (Hadjar & Hemmje, 2025: 2–3). In the context of work and organizational psychology, however, the application of FER also raises important ethical concerns, prompting reflection on the boundaries of emotional surveillance in the workplace and the need to safeguard employees' rights (Hajric et al., 2024: 3–4).

Facial Emotion Recognition: From Classical Approaches to Deep Learning

Facial Emotion Recognition has been at the center of scientific attention since the 1970s, bringing together researchers in nonverbal communication, psychology, and artificial intelligence. As technology evolved, so too did the methods used to identify emotions — shifting from classical image analysis techniques based on manually defined rules and handcrafted features (Aslam & Hussian, 2021: 2–5) to modern deep learning approaches capable of autonomously learning emotional patterns from large datasets (Rajan, Chenniappan, & Devaraj, 2020: 1373-1374). This transition has not only improved recognition accuracy but has also opened new possibilities in areas such as remote

psychological support, education, healthcare, and well-being monitoring (Hans & Rao, 2021: 11–12). The earliest attempts to automate emotion recognition relied on manual feature extraction, where experts selected facial elements considered relevant to emotion expression (Wegrzyn, Vogt, Kireclioglu, & Schneider, 2017: 5). Two dominant strategies emerged: the geometric approach, which analyzed the positions of key facial landmarks (e.g., the distance between eyebrows), and the appearance-based approach, which focused on textures, gradients, and local patterns (Kas, Ruichek, & Messoussi, 2021: 10).

One of the most well-known systems was the Facial Action Coding System (FACS) developed by Ekman, which decomposed facial expressions into Action Units (AUs), each corresponding to specific muscle movements. While FACS proved accurate and useful in psychological research, it required significant resources — including time, training, and resilience to individual and environmental variability (Kas, Ruichek, & Messoussi, 2021: 11). Several classical FER methods also had technical implementations. Haar-like features, introduced by Viola and Jones, analyzed brightness differences between regions of the image in real time (Viola, 2001: 512). Histograms of Oriented Gradients (HOG) enabled the detection of edges and shapes in images regardless of lighting or noise — a breakthrough in object detection (Dalal, 2005: 887-888). Another important stream involved ensemble methods, such as Bagging and Boosting, which combined the outputs of multiple classifiers. This approach reduced individual model errors and improved classification performance (Dietterich, 2000: 1–4). Despite their many advantages, classical approaches proved difficult to scale. They required expert knowledge, performed poorly on diverse datasets, and were prone to errors in dynamic environments.

The development of deep learning (DL) has radically transformed how emotions are analyzed today. Rather than manually defining where to *look for emotions,* neural networks learn these patterns independently — analyzing thousands or even millions of images. As a result, systems have become more robust to changes in lighting, facial orientation, and individual appearance. The key to this transformation was the introduction of Convolutional Neural Networks (CNNs), which analyze images in layers — from simple patterns to complex emotional configurations (Palaniswamy, 2019: 2). Models such as VGGNet (a deep, symmetrical network with small filters), ResNet (a residual network that learns the *differences* between layers, allowing for deeper architectures), and Inception (a complex structure that analyzes data at multiple scales simultaneously) have become FER standards (He, et al. 2016: 770). In scenarios where emotions evolve over time — such as during video calls — Recurrent Neural Networks (RNNs) and their advanced form, Long Short-Term Memory (LSTM) networks, are used. These models can "remember" key information while ignoring irrelevant details (Hochreiter & Schmidhuber, 1997: 1743).

In recent years, hybrid approaches have gained popularity — combining CNNs with LSTM networks or attention mechanisms, allowing for the capture of both visual structure and temporal context. Transfer learning has also become significant — leveraging pre-trained models (e.g., on the ImageNet dataset) and fine-tuning them for specific tasks, reducing training time and improving performance (Yen & Li, 2022: 4). Thanks to frameworks such as TensorFlow, PyTorch, and Keras, creating, training, and deploying DL models has become accessible not only to scientists but also to practitioners in psychology, medicine, and education (Ismail et al., 2024: 11).

So why are classical methods still in use if deep learning seems superior? The answer is context. Classical approaches are faster, cheaper, and easier to implement. In controlled or educational environments, where high flexibility is not required, they may be entirely sufficient. Deep learning, on the other hand, is better suited to complex, dynamic datasets that require high precision — such as psychological diagnostics, employee well-being assessment, or crisis intervention systems (table 1).

Table 1. *Comparison of Classical and Deep Learning Approaches in Facial Emotion Recognition*

Criterion	Classical FER Methods	Deep Learning Approaches
Feature extraction	Manual, expert-driven	Automated, learned by the network
Data requirements	Low	High – requires large datasets
Noise resistance	Low	High
Result interpretability	High	Low ("black box")
Computational efficiency	Low	High – needs GPU/cloud
Model preparation time	Short	Long (training required)
Applications	Education, offline analysis	Clinical use, remote work, mobile apps, well-being monitoring
Example techniques	FACS, Haar, HOG, Boosting	CNN, ResNet, LSTM, Attention, Transfer Learning

Note. Comparison of classical and deep learning approaches to facial emotion recognition, based on the authors' comparative analysis.

Understanding the differences between classical and modern FER methods enables not only a more precise and context-appropriate selection of analytical tools, but also a more reflective evaluation of their practical applicability, technical limitations, and underlying assumptions. As technological capabilities continue to evolve, researchers and practitioners are increasingly faced with the need to align methodological choices with the specific demands of their domain. This includes considerations such as data availability, computational resources, interpretability requirements, and time constraints. Ultimately, the effectiveness of facial emotion recognition depends not solely on the sophistication of the algorithm, but also on the clarity of its intended use, the conditions under which it operates, and the ethical standards guiding its implementation.

Materials and Methods

This study evaluated the effectiveness of selected machine learning and deep learning models in the context of automatic Facial Emotion Recognition. The FER-2013 dataset was chosen for the experiments, as it is one of the most widely used and broadly recognized datasets in the scientific community—particularly in studies involving deep learning-based emotion recognition. Due to its diversity in terms of facial expressions and image quality, FER-2013 is considered one of the most frequently used benchmarks in this field (Goodfellow, 2013: 62). The FER-2013 dataset played a key role in the ICML 2013 Challenge, which contributed to its standardization in the evaluation of novel algorithms (Mollahosseini, 2016: 1). Additionally, it offers a variety of realistic, "in-the-wild" facial images, making it highly representative and useful for developing systems intended for real-world applications (Mollahosseini, 2016: 2; Minaee et al., 2021: 5). In

numerous literature reviews, FER-2013 has been cited as a benchmark standard in emotion recognition research due to its widespread use and well-established structure (Li & Deng, 2022: 1195). Its popularity and accessibility make it a natural choice for testing the effectiveness of modern emotion recognition models.

To compare different approaches to FER, a set of models was selected to represent both classical and modern deep learning techniques. The models applied in this study included CNNs (LeCun, et al, 1998: 2283), RNNs (Elman, 1990: 7), their advanced variant LSTM networks (Hochreiter, 1997: 9), as well as state-of-the-art deep architectures such as ResNet (He, 2016: 770) and VGGNet (Simonyan, 2015: 2). Additionally, ensemble methods were included, combining the outputs of multiple classifiers to enhance prediction accuracy (Dietterich, 2000: 1). Within the scope of classical emotion recognition methods, feature extraction techniques based on Haar-like features (Viola & Jones, 2001: 512) and Histograms of Oriented Gradients (Dalal & Triggs, 2005: 887-888) were applied. Haar-like features enable the rapid detection of local brightness differences in an image, making them useful for analyzing simple facial patterns with low computational complexity. This technique employs rectangular masks that slide across the image, analyzing contrast between selected regions and capturing typical lighting configurations associated with specific emotional expressions (Viola & Jones, 2001: 512). HOG, on the other hand, allows for precise analysis of local edges and contours by computing gradient directions in small segments of the image and converting them into gradient histograms (Dalal & Triggs, 2005: 888). This method is known for its robustness to lighting variations, which enhances its applicability in visually diverse environments.

The study also incorporated ensemble methods as an extension of classical approaches by combining the outputs of multiple independent classifiers. Techniques such as Bagging and Boosting (Dietterich, 2000: 1–5) improve prediction stability and accuracy by reducing variance and increasing model diversity. Through mechanisms such as voting or weighted prediction averaging, ensemble methods help build more resilient classification systems, particularly in contexts involving incomplete or heterogeneous training data.

In the domain of deep learning methods, Convolutional Neural Networks were applied. These networks consist of convolutional layers, pooling layers, activation functions (ReLU), and fully connected layers (LeCun et al., 1998: 2283). Each module of the network plays a critical role in extracting and classifying facial patterns — convolutional layers process the image locally by filtering features such as edges and textures, while pooling layers reduce dimensionality and increase the model's robustness to spatial distortions. The learning process relies on backpropagation and weight updates to minimize the loss function. The study also included advanced deep learning architectures such as VGGNet and ResNet. VGGNet, with its consistent structure of multiple convolutional layers using small filters (3×3), enables efficient extraction of complex features while maintaining computational stability and limiting the number of parameters (Simonyan, 2015: 2–3). Regularization techniques such as dropout and L2 were implemented to reduce overfitting — a critical factor when working with limited training datasets. ResNet, in contrast to conventional CNNs, introduces residual connections that allow information to bypass certain layers without degradation of the gradient, thereby facilitating the training of very deep networks (He et al., 2016: 771–773). In addition to spatial analysis, sequential architectures were also

applied — specifically RNNs and their extended form, LSTM networks. RNNs are capable of modeling temporally ordered data, which is particularly useful for analyzing sequences of facial images where changes in emotion over time are key (Elman, 1990: 186). LSTM networks, through the use of memory cells with input, output, and forget gates, can selectively retain or discard information, enabling the modeling of long-term dependencies and reducing the vanishing gradient problem (Hochreiter & Schmidhuber, 1997: 1744).

The selection of methods was based on their documented effectiveness and compatibility with the FER-2013 dataset. All models were implemented and validated under consistent conditions, ensuring reliable and comparable results. A comparative analysis was conducted on six models representing both classical feature-based and modern deep learning approaches to facial emotion recognition.

Table 2. *Summary Comparison of All Evaluated FER Method*

Model	Approach type	Computational requirements	Noise resistance	Applications	Advantages	Limitations
VGG16	Deep learning	High (GPU)	High	Psychology, healthcare, research	Well-documented and effective	Large size, needs transfer learning
ResNet	Deep learning	High (GPU)	Very high	Psychology, adaptive systems	High accuracy with deep structures	Complex and GPU-demanding
LSTM	Deep learning (sequential)	Medium–high (GPU)	Medium	Video calls, temporal emotion analysis	Long-term dependency modeling	Challenging to fine-tune
RNN	Deep learning (sequential)	Medium (GPU optional)	Medium	Dynamic emotion modeling	Simplified sequence processing	Lower accuracy than LSTM
HOG	Classical feature extraction	Low	Medium	Fast offline classification	Transparency, light-insensitive	No automatic feature learning
Haar-like features	Classical feature extraction	Very low	Low	Embedded systems, education	Fast processing, simplicity	Low effectiveness and flexibility

Note. Overview of key characteristics, strengths, and limitations of selected FER models, developed by the authors.

These models differ not only in architectural complexity but also in their data requirements, robustness to noise, and interpretability. The following comparison includes key technical and metrological aspects, allowing for an informed assessment of each model's applicability in psychological, educational, and organizational contexts. The table (table 2) presents the most relevant configuration parameters, distinguishing features, and potential applications of all the methods analyzed.

One of the analyzed models was a modified VGG16 architecture, based on the concept presented by Simonyan and Zisserman (Simonyan, 2015: 1), but adapted to the practical requirements of experiments using the FER-2013 dataset. The introduced simplifications were aimed at increasing computational efficiency and reducing training time while preserving the method's functionality in the context of emotion analysis

under remote work conditions. The model utilized a transfer learning strategy—freezing all convolutional layers responsible for feature extraction, while modifying the final classification layer to match the seven emotion classes present in the FER-2013 dataset.

Table 3. *Configuration of the Simplified VGG16 Model Used for FER Tasks*

Configuration Element	Description
Architecture	VGG16 (based on Simonyan & Zisserman, 2015), with modifications
Approach	Transfer learning
Frozen layers	Yes (convolutional layers)
Modified final layer	Yes – 7 outputs (FER2013 emotion classes)
Input image size	224 × 224 px
Data normalization	RGB means and standard deviations from ImageNet
Data split	Training: 80%, Validation: 20%
Number of epochs	10
Batch size	32
Optimizer	Adam
Learning rate (lr)	0.0001
Regularization (dropout/L2)	Not applied
Loss function	CrossEntropyLoss
Evaluation metrics	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Note. Configuration details of the simplified VGG16 architecture tailored for FER, as implemented in this study.

All images were rescaled to a resolution of 224×224 pixels. The data were standardized based on the mean and standard deviation values of the RGB channels used during the original ImageNet training. The dataset was split into a training set (80%) and a validation set (20%), maintaining class balance. The model was trained for 10 epochs using the Adam optimizer (learning rate of 0.0001) and the CrossEntropyLoss function. A batch size of 32 was applied. Adam optimizer is an adaptive learning rate optimization algorithm designed for training deep neural networks. It combines the advantages of two other popular methods: AdaGrad and RMSProp, by estimating both the first and second moments of the gradients (Yi, Ahn, & Ji, 2020: 2).No regularization techniques such as dropout or L2 were used, and no learning rate scheduling was implemented. These decisions were made to simplify the process and allow for efficient experimentation under resource constraints. During each epoch, classification accuracy was monitored on both the training and validation sets. After training was completed, the model's performance was evaluated using metrics such as precision, recall, and the harmonic mean (F1-score). A confusion matrix was also generated to analyze the most frequent classification errors (table 3).

The second analyzed model was the ResNet architecture (He et al., 2016: 771), originally designed to address the vanishing gradient problem in very deep neural networks. In this study, the ResNet-18 version was selected as a compromise between

network depth and computational efficiency, while maintaining the ability to effectively extract features from complex facial expressions.

Table 4. *Configuration of the Simplified ResNet-18 Model Used for FER Tasks*

Configuration Element	Description
Architecture	ResNet-18 (based on He et al., 2016), adapted for FER tasks
Approach	Transfer learning
Frozen layers	Yes (convolutional layers)
Modified final layer	Yes – 7 outputs (FER2013 emotion classes)
Input image size	224 × 224 px
Data normalization	RGB means and standard deviations from ImageNet
Data split	Training: 80%, Validation: 20%
Number of epochs	10
Batch size	32
Optimizer	Adam
Learning rate (lr)	0.0001
Regularization (dropout/L2)	Not applied
Loss function	CrossEntropyLoss
Evaluation metrics	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Note. Experimental setup of the ResNet-18 model adapted to the FER-2013 dataset.

The model was initialized using pre-trained weights from ImageNet and then adapted to the specifics of the FER-2013 dataset by freezing all convolutional layers and replacing the final classification layer with a new fully connected layer featuring seven outputs corresponding to the FER-2013 emotion classes. The data preparation process was identical to that used with the VGG16 model: images were resized to 224×224 pixels and normalized using the RGB means and standard deviations from ImageNet. The dataset was split into training (80%) and validation (20%) subsets. ResNet was trained for 10 epochs using the Adam optimizer (learning rate: 0.0001), without additional regularization. A batch size of 32 was applied. The training strategy prioritized simplicity and consistency with the VGG16 configuration, enabling a reliable comparative analysis. During each epoch, classification accuracy was monitored on both the training and validation sets. After training, the model's performance was evaluated based on accuracy, precision, recall, and F1-score, and a confusion matrix was generated. The results (table 4) allowed for an assessment of ResNet's effectiveness in recognizing facial emotions under conditions of limited data and resources.

The next model analyzed was the Long Short-Term Memory network, an extension of classical Recurrent Neural Networks, designed to retain long-term temporal dependencies in sequential data (Hochreiter & Schmidhuber, 1997: 1744). In the context of emotion recognition, this model enables the analysis of image sequences (e.g., video recordings), allowing for the capture of emotional expression dynamics — a key factor in psychological diagnostics and the analysis of online interactions. The LSTM network was combined with a convolutional feature extraction network (e.g., CNN), where the

CNN served as the feature extractor. The extracted vector representations of each frame (typically from the final convolutional layers) were then processed by the LSTM layers to identify temporal emotional changes. A single-layer LSTM structure was used, consisting of 128 units and a dropout rate of 0.3. A fully connected output layer followed, producing seven outputs corresponding to the emotion classes. Instead of using individual static images, the data were transformed into short sequences of images (e.g., five consecutive frames). Each frame was normalized according to ImageNet RGB means and standard deviations and resized to 224×224 pixels. The sequences were class-balanced across emotion categories. The model was trained for 15 epochs using the Adam optimizer (learning rate = 0.0001), the CrossEntropyLoss function, and a batch size of 16. Validation accuracy was monitored during training, and a full evaluation of classification quality was conducted upon completion. The LSTM model enabled the capture of the sequential nature of emotions, which is particularly valuable in the analysis of video recordings, online conversations, and therapeutic observations (table 5). This model allowed for the identification of subtle changes in facial expressions that might be missed by traditional convolutional models.

Table 5. *Configuration of the LSTM-Based Model for Temporal Facial Emotion Recognition*

Configuration Element	Description
Architecture	Long Short-Term Memory (LSTM)
Integration with CNN	Yes – CNN as feature extractor
Number of LSTM layers	1
Number of LSTM units	128
Dropout	0.3
Final output layer	Fully connected (7 emotion classes)
Input data type	Image sequences (e.g., from video)
Input frame size	224 × 224 px
Number of frames per sequence	5
Batch size	16
Number of epochs	15
Optimizer	Adam
Learning rate	0.0001
Loss function	CrossEntropyLoss
Evaluation metrics	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Note. Design of the LSTM-based model used for analyzing temporal dynamics of facial emotions.

RNNs are among the oldest yet most foundational architectures for sequence processing in the field of machine learning. In this study, a basic version of an RNN was applied to analyze sequences of facial images for the purpose of emotion recognition. Despite its limitations, this model is capable of capturing basic temporal dependencies in sequential image data, which may be useful in analyzing short-term emotional dynamics. The RNN model received as input a sequence of feature vectors extracted from the output of a CNN feature extractor (e.g., convolutional layers). Unlike LSTM,

classical RNNs lack gating mechanisms to regulate information flow, making them more susceptible to the vanishing gradient problem but also less resource-intensive and faster to train. A single-layer RNN with 64 units was implemented, followed by a fully connected output layer with seven outputs corresponding to the emotion classes. The input data consisted of five consecutive frames from image sequences (e.g., extracted from video recordings). Each frame was resized to 224×224 pixels and normalized according to ImageNet standards. The model was trained for 10 epochs using the Adam optimizer (learning rate = 0.0001), the CrossEntropyLoss function, and a batch size of 16. During training, classification accuracy and validation loss were monitored (table 6). Although RNNs are less advanced than LSTM networks, they enabled a preliminary analysis of emotional changes over time. This makes them suitable for projects requiring basic sequence analysis or for use in resource-constrained environments such as mobile devices or educational systems.

Table 6. *Configuration of the RNN-Based Model for Sequential Facial Emotion Recognition*

Configuration Element	Description
Architecture	Recurrent Neural Network (RNN)
Integration with CNN	Yes – CNN as feature extractor
Number of RNN layers	1
Number of RNN units	64
Final output layer	Fully connected (7 emotion classes)
Input data type	Image sequences (e.g., from video)
Input frame size	224 × 224 px
Number of frames per sequence	5
Batch size	16
Number of epochs	10
Optimizer	Adam
Learning rate	0.0001
Loss function	CrossEntropyLoss
Evaluation metrics	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Note. Structure and training setup of the RNN model applied to sequential facial emotion data.

The Histogram of Oriented Gradients is a classical feature extraction technique that was widely used in image analysis prior to the deep learning era. In the context of facial emotion recognition, HOG allows for the capture of local edges, contours, and structural patterns of the face that may be characteristic of specific emotional expressions. The HOG method divides the image into small, regular cells (e.g., 8×8 pixels), within which a histogram of gradient orientations — i.e., brightness changes — is computed. These histograms are then normalized across larger blocks (e.g., 2×2 cells), improving robustness to lighting variations. The result is a feature vector that represents the structure of the image in a form suitable for classification. In this study, a standard HOG configuration was used: 8×8 cells, 2×2 blocks, and 9 histogram bins. Feature extraction was performed using the **scikit-image** library. The resulting feature vectors

were classified using a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel. Images from the FER-2013 dataset were rescaled to 64×64 pixels (due to SVM classifier constraints) and processed using HOG. The data were split into training and test sets in an 80/20 ratio. The SVM classifier was trained using default parameters: C=1.0, gamma='scale', with the RBF kernel. After training, classification performance was evaluated using standard metrics: accuracy, precision, recall, F1-score, and a confusion matrix (table 7). Although HOG is an older technique compared to deep learning models, it offers high interpretability and can be implemented on devices with limited computational power. Its main limitations include low tolerance to data variability (e.g., different face orientations) and the lack of self-learning capability. Nevertheless, in controlled or educational environments, it may provide a fast and efficient solution for basic emotion analysis.

Table 7. *Configuration of the HOG-Based Method for Classical Facial Emotion Recognition*

Configuration Element	Description
Method Type	Histogram of Oriented Gradients (HOG)
Cell size	8 × 8 pixels
Block size	2 × 2 cells
Number of histogram bins	9
Histogram normalization	L2-Hys
Input image size	64 × 64 px
Feature extraction	scikit-image library
Classifier	Support Vector Machine (SVM)
SVM parameters	C=1.0, gamma='scale', RBF kernel
Data split	80% training, 20% test
Evaluation metrics	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Note. Description of the HOG-based feature extraction method combined with SVM, used in the authors' experiments.

The Haar-like features method is one of the earliest and simplest techniques used in image analysis, based on intensity differences between adjacent regions. Popularized by the Viola-Jones face detector, it remains a fast and efficient way to detect structural features, including facial patterns. Haar features use rectangular masks to compare brightness areas, such as contrasting the forehead with the eyebrows. Their computational efficiency results from integral images, enabling rapid pixel sum calculations. In this study, several hundred Haar features were generated using OpenCV, and classification was performed using the AdaBoost algorithm — an ensemble method that improves accuracy by combining weak learners. Images from the FER-2013 dataset were resized to 48×48 pixels and processed into integral images. Feature vectors were classified after 50 AdaBoost iterations. Evaluation included accuracy, precision, recall, and F1-score metrics (Table 8).

Although considered outdated compared to CNN or HOG methods, Haar-like features offer advantages in highly resource-constrained environments, such as embedded systems or mobile applications. Their main strength lies in speed, with limitations in robustness to data variability and complex pattern recognition.

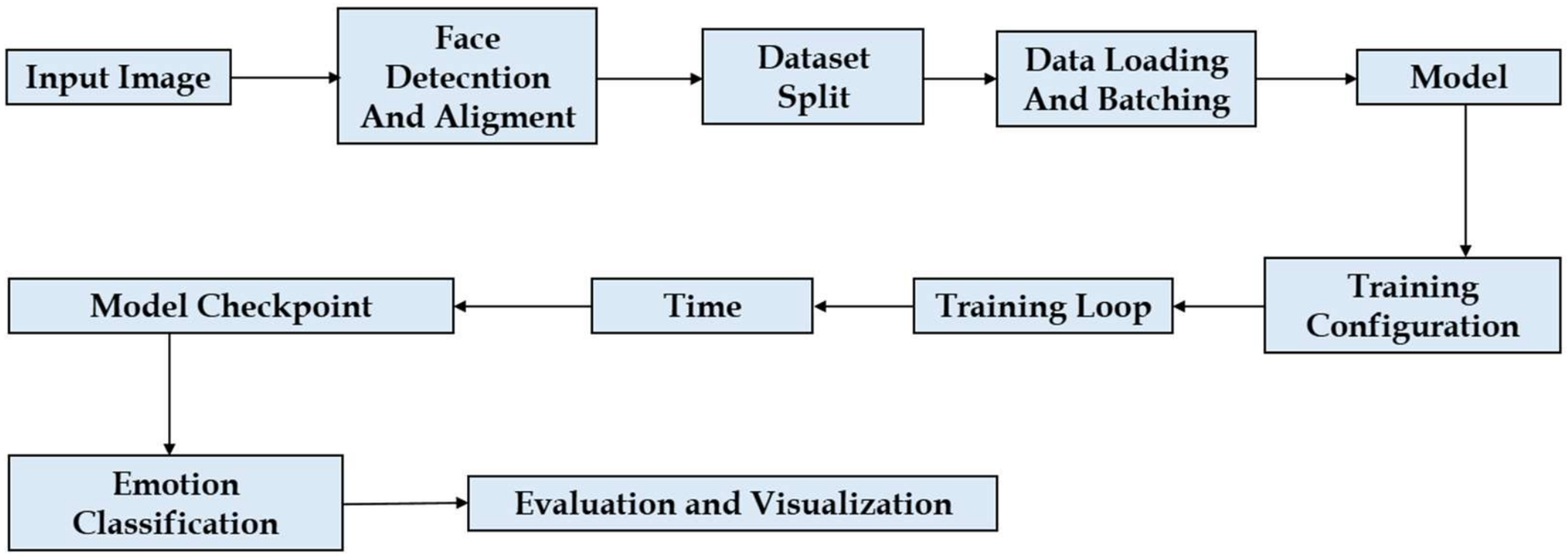
Table 8. *Configuration of the Haar-Like Features Method for Classical Facial Emotion Recognition*

Configuration Element	Description
Method Type	Haar-like features
Operating principle	Analysis of brightness differences between rectangular regions
Feature extractor	OpenCV Haar cascades
Input resolution	48 × 48 px
Image representation	Integral image
Number of features	Several hundred automatically generated features
Classifier	AdaBoost (ensemble method)
AdaBoost parameters	50 iterations, default loss function
Data split	80% training, 20% test
Evaluation metrics	Accuracy, Precision, Recall, F1-score

Note. Implementation of the Haar-like features approach with AdaBoost classifier in the context of classical FER.

All experiments were conducted on a single-workstation rig equipped with an AMD Ryzen 9 9900X CPU, 32 GB DDR5-6000 RAM, and an NVIDIA GeForce RTX 3090 GPU (24 GB GDDR6X). The host operating system was Windows 11 Pro (23H2), and the main software stack included Python 3.13, PyTorch 2.7.0, TorchVision 0.21, and CUDA 12.6. During the training and evaluation procedures, the system consistently operated under 14 GB of VRAM usage and did not exceed 320 W of board power. The entire facial emotion recognition process — from image input to final evaluation — followed a structured and repeatable workflow, ensuring methodological consistency. This workflow is illustrated in Scheme 1.

Scheme 1. *Workflow for Facial Emotion Recognition Model Training and Evaluation.*



Note. Scheme created by the authors based on the experimental setup.

Results

The evaluation of the facial emotion recognition models was conducted using several complementary metrics: accuracy, precision, recall, F1-score, and training time. These measures offer different perspectives on model performance, capturing not only overall correctness but also sensitivity to positive instances and the balance between precision and recall. In machine learning, an epoch refers to a complete cycle in which the model processes the entire training dataset. After each epoch, the model updates its internal parameters (weights) and, if more epochs are set, begins a new cycle with updated weights (Das & Das, 2023). Table 9 presents the models' overall accuracy, weighted averages of precision, recall, F1-score, and the time required to complete training. Precision in this context reflects how many of the predicted positive cases were correct; it is calculated by dividing true positives (TP) by the sum of true positives and false positives (FP) (Hersh, 2005). Recall indicates how many actual positive cases were correctly identified by the model, calculated by dividing TP by the sum of TP and false negatives (FN) (Hersh, 2005). The F1-score, which is the harmonic mean of precision and recall, provides a single measure balancing these two aspects, and gives insight into the model's ability to identify emotions accurately and consistently (Hersh, 2005).

Table 9. *General Performance Metrics of FER Models*

Model	Accuracy	Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score	Time (s)
ResNet	0.44	0.42	0.44	0.40	680
VGG-16	0.60	0.60	0.56	0.59	648
HOG	0.44	0.43	0.44	0.43	155
RNN	0.37	0.30	0.29	0.27	465
LSTM	0.44	0.43	0.44	0.42	494
Haar	0.56	0.55	0.56	0.55	300

Note. Summary of model classification performance based on weighted averages and computational time.

Among the evaluated models, VGG-16 achieved the highest overall accuracy (0.60) and weighted F1-score (0.59), indicating a strong balance between precision and recall. Haar-based detection also yielded robust results (accuracy: 0.56), with notably lower training times (300 seconds), which highlights its efficiency in resource-constrained environments. In contrast, the RNN model demonstrated the weakest performance across all metrics, achieving the lowest weighted F1-score (0.27) and accuracy (0.37), suggesting limitations in recognizing facial emotions effectively. Additionally, the LSTM model, although showing better performance than the RNN, exhibited only moderate improvements, indicating that sequential architectures might require further tuning for optimal FER performance.

A more detailed analysis focusing on recognition performance for specific emotions is presented in Table 10. This table identifies the emotion best recognized (based on precision and recall) and the least effectively detected emotion (based on recall) for each model, providing further insight into their individual strengths and weaknesses across emotional categories.

Table 10. *Summary of Facial Emotion Recognition (FER) Model Performance*

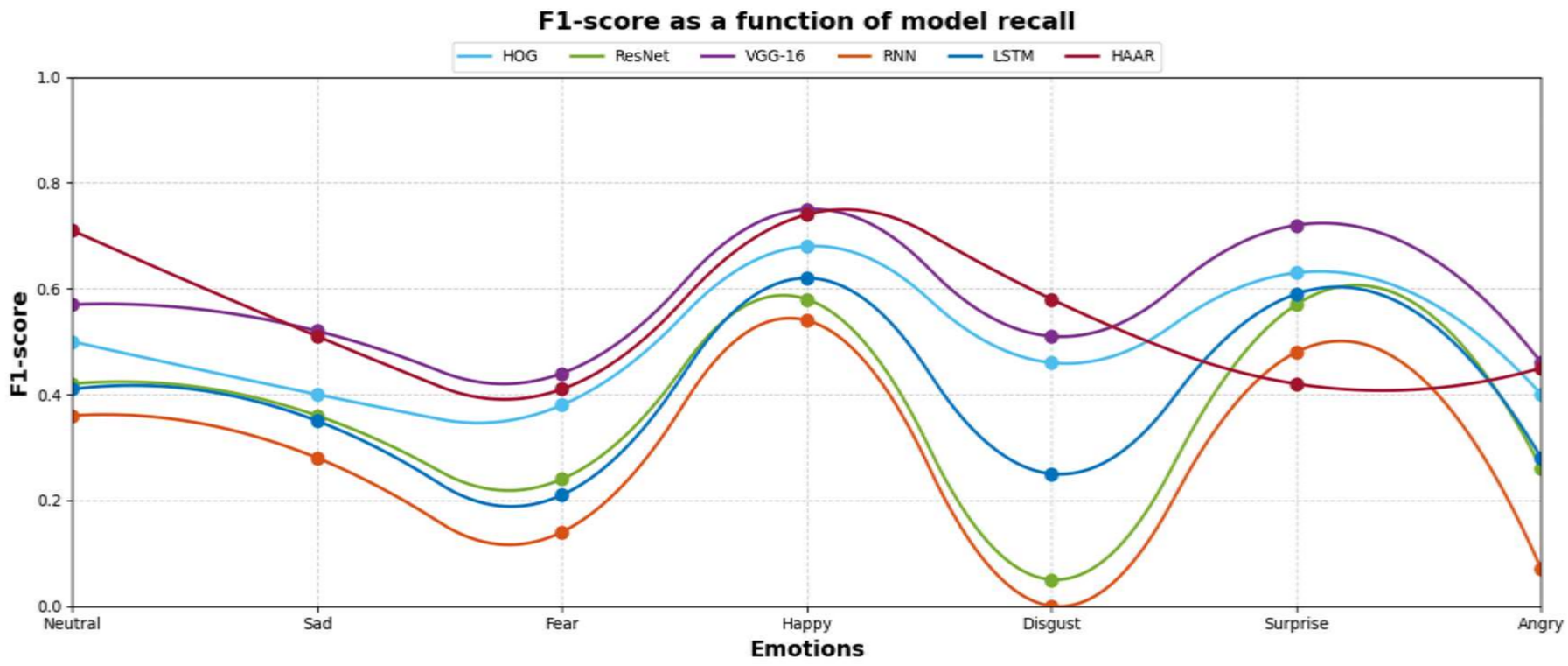
Model	Overall Accuracy	Best Precision (Emotion)	Best Recall (Emotion)	Worst-Recognized Emotion (Recall)
ResNet	0.44	surprise (0.72)	happiness (0.71)	disgust (0.03)
VGG16	0.6	disgust (0.79)	happiness (0.88)	fear (0.34)
HOG	0.44	happiness (0.57)	happiness (0.72)	disgust (0.2)
RNN	0.37	surprise (0.52)	happiness (0.72)	fear (0)
LSTM	0.44	surprise (0.62)	happiness (0.73)	fear (0.16)
Haar	0.56	neutral (0.76)	happiness (0.78)	fear (0.37)

Note. Summary of classification performance across all tested models, based on the results obtained in the present study.

Among the tested models, VGG-16 again stood out, achieving the highest precision (0.79 for disgust) and recall (0.88 for happiness), confirming its capacity to recognize distinct emotional patterns effectively. Haar cascades performed well for neutral and happiness emotions, balancing speed and classification performance. The ResNet model demonstrated good precision for surprise (0.72) and recall for happiness (0.71), but struggled considerably with disgust (recall: 0.03), indicating challenges in capturing more subtle expressions. Classical feature-based methods, such as HOG, achieved stable but modest results, while RNN and LSTM models displayed weaker recognition rates, particularly for fear, with RNN failing to recognize it at all (recall: 0.00).

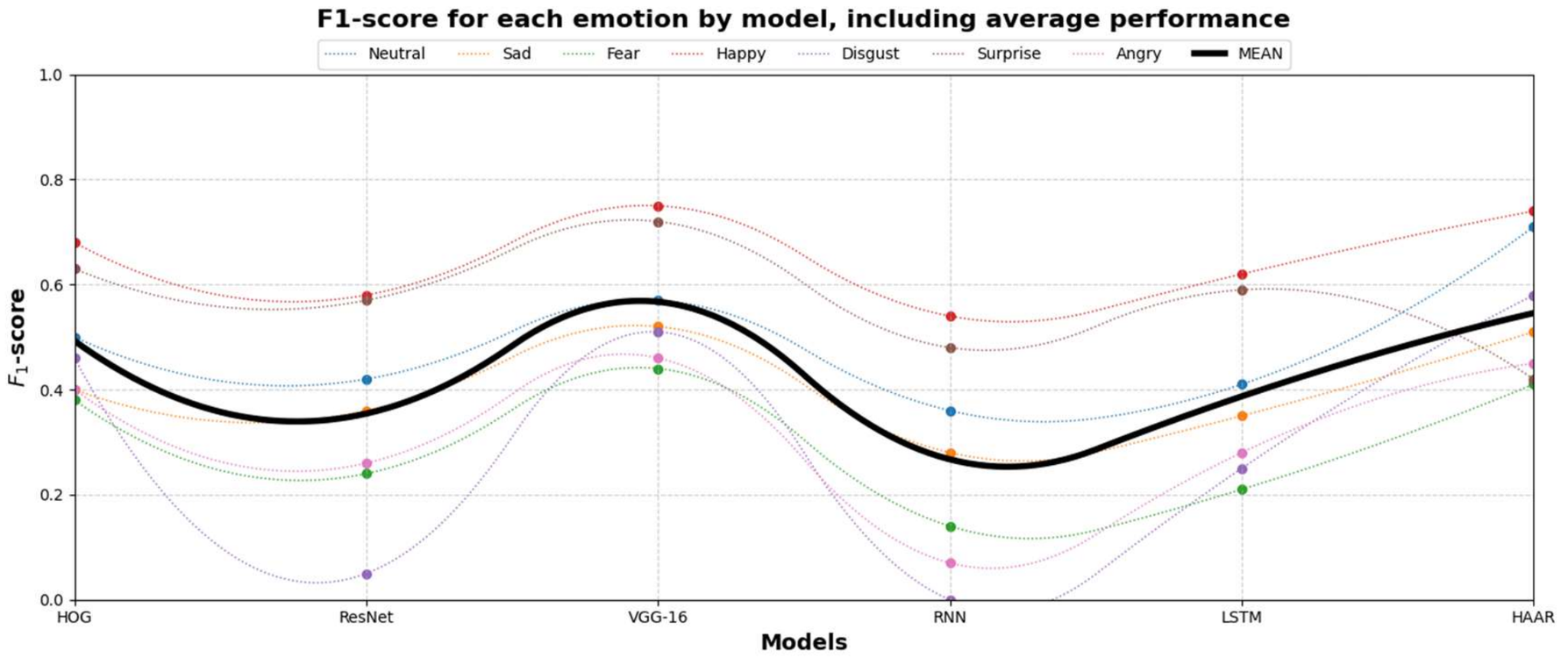
To further explore model performance, three additional visual analyses were conducted.

Figure 1. *F1-score across FER models by emotion category*



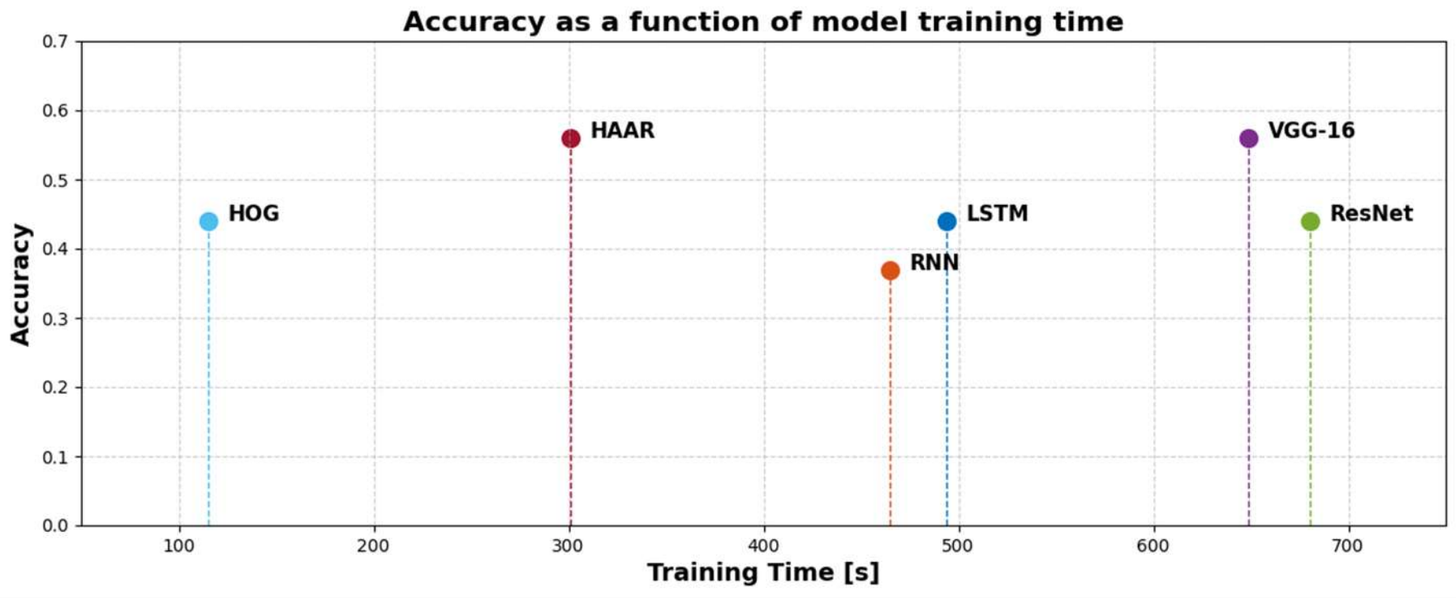
Note. Figure created by the authors based on the current study results.

Figure 2. *F1-score for each emotion by model, including average performance*



Note. Figure created by the authors to compare average and emotion-specific scores.

Figure 3. *Model accuracy as a function of training time*



Note. Figure created by the authors to show the trade-off between accuracy and time.

Figure 1 presents F1-scores across emotional categories for each model, highlighting that happiness and surprise were generally easier to recognize, whereas fear and disgust remained more difficult for all architectures. Figure 2 aggregates F1-scores across models, with the bold curve representing mean performance. VGG-16 and LSTM consistently showed higher overall recognition abilities compared to the other models. Figure 3 illustrates the trade-off between model accuracy and training time, showing that while VGG-16 and Haar achieved better classification results, their training times were significantly longer compared to lightweight methods like HOG.

Discussion

The comparative analysis revealed that certain emotions, such as happiness and surprise, were generally recognized with higher accuracy across models, while others, including fear, disgust, and anger, remained significantly more difficult to classify. These patterns, also observable in Figures 1 and 2, reflect the inherent ambiguity and lower facial distinctiveness of some emotional expressions. Notably, even advanced deep learning models struggled to detect these more nuanced emotions consistently. Moreover, as highlighted in Figure 3, achieving high accuracy often came at the cost of extensive training time, raising important questions about the practical feasibility of deploying complex FER systems in low-resource environments. These observations underscore the need to match model complexity with the intended application context—balancing performance, efficiency, and ethical considerations.

Facial Emotion Recognition systems are gaining increasing significance across a variety of domains, from education and healthcare to smart home technologies. In education, FER can be used to monitor students' emotions in real time—allowing teachers to assess whether a student is frustrated, bored, or engaged and tailor their teaching methods accordingly (Khalfallah & Ben Hadj Slama,, 2015: 276). One example includes a JavaScript-based system that tracks facial landmarks and assesses emotional responses during remote laboratory sessions (Khalfallah, 2015: 279). In the workplace, FER helps monitor employees' moods, supporting psychological well-being and productivity. For instance, Raspberry Pi-based systems can recognize real-time emotions such as joy, sadness, or anger (Rathour et al., 2021: 4). In smart home environments, FER is used to automate surroundings—e.g., by changing a television channel or lowering the volume upon detecting user frustration (Hossain & Muhammad, 2017: 2283). In healthcare, FER is applied in the diagnosis and monitoring of mood disorders such as depression, including in home settings. This technology may facilitate early detection of relapses and assessment of treatment effectiveness (Guo et al., 2024: 6). Recent advancements also include the development of contactless FER systems, which utilize radar sensors, thermal cameras, or Wi-Fi wave analysis instead of RGB cameras. These allow for emotion detection even without direct facial visibility (Khan et al., 2024: 13).

The results obtained in this study suggest that FER systems can significantly support remote work—provided that the technology is appropriately matched to the organization's needs and team structure. The choice of model should not be based solely on classification accuracy but also on computational requirements, robustness to environmental noise, and real-time performance capabilities. For example, the stable VGG16 architecture proves effective in educational settings where teachers conduct online classes and need to monitor student engagement without relying on high-

performance hardware. Due to its predictable structure and consistent results, VGG16 also performs well in organizations that conduct remote training or professional development sessions—especially where capturing clear emotional states such as frustration, joy, or boredom is essential.

The more complex ResNet architecture, with its use of residual connections, proved effective in capturing subtle facial variations. It may be successfully applied in large corporate environments where real-time emotion analysis supports not only mental well-being but also the early detection of burnout or work overload. Such systems may also be valuable for UX and AI research departments, where users' emotional reactions to new digital products are evaluated. Although ResNet requires significant computational resources, it enables deep emotion analysis in visually and culturally diverse settings.

For organizations relying on intensive synchronous communication—such as therapeutic teams, consultants, coaches, or customer service units—the LSTM model may be more appropriate. It enables temporal analysis of emotions, capturing mood changes throughout the course of an interaction. This makes it possible not only to detect emotions at a given moment but also to identify critical turning points—such as when a client begins to experience irritation or anxiety. LSTM is particularly valuable in analyzing continuous interactions, such as video calls, psychological support sessions, or online coaching.

Organizations without access to advanced technological infrastructure, but with a need for speed and interpretability, may benefit from classical methods like Histogram of Oriented Gradients. Due to their relatively low computational requirements, such methods are ideal for small businesses, NGOs, or educational institutions in developing countries, especially when emotion analysis is conducted on recorded material rather than in real time. HOG can also be used in basic feedback systems after remote meetings, aimed at assessing the general emotional tone of participant responses.

In cases where maximum responsiveness and minimal resource consumption are essential—such as in smart homes or mobile devices—techniques like Haar-like features combined with AdaBoost classifiers may be employed. These systems can respond instantly to the user's emotions, e.g., by adjusting lighting or music based on mood. While their performance does not match that of deep learning models, their simplicity and efficiency make them viable for implementation where other methods may be impractical.

Recurrent Neural Networks (RNNs), although less advanced than LSTMs, may be suitable for projects requiring basic emotion analysis over time—particularly in mobile applications, chatbots, or educational systems, where limited resources and fast processing are critical. Their simplicity may be an advantage for rapid prototyping or deploying "good enough" solutions that work effectively in the moment. The diversity of FER methods not only demonstrates the technological potential for emotion analysis but also underscores the importance of a cautious, context-aware implementation strategy. The choice of model should always be preceded by an assessment of technical constraints, organizational goals, and the specific nature of remote work in a given team. When applied thoughtfully, FER systems can support emotional diagnostics and contribute to building healthier, more balanced, and human-centered work environments.

Despite the promising results and the broad range of methods applied, this study has several important limitations that should be considered when interpreting the findings or attempting to generalize them. First, the FER-2013 dataset—although widely recognized as a benchmark in emotion recognition research—has structural and content-related limitations. Its grayscale images are standardized to a low resolution (48×48 pixels), which reduces the ability to capture subtle facial nuances—especially for more complex or less distinct emotions such as surprise or disgust. Moreover, the dataset lacks ethnic diversity; most facial images are of white individuals, which may result in models learning emotion patterns that are typical for this demographic while marginalizing facial expressions from other populations. This raises concerns about the generalizability of results to demographically diverse educational or workplace settings.

A second limitation of the study was the simplified experimental configuration, driven by the need to conduct research under constrained technical conditions. All models were trained for a relatively short time—only 10 to 15 epochs—without applying performance-enhancing techniques such as data augmentation, fine-tuning, dynamic learning rate adjustments, or advanced regularization strategies. While this approach allowed for the efficient comparison of multiple architectures under unified conditions, it may have limited the models' actual potential. This is particularly relevant for more complex architectures, such as ResNet or LSTM, which typically require longer training periods and precise hyperparameter calibration to achieve optimal performance.

Another important constraint was the computational limitation, which necessitated the exclusion of more advanced architectures such as ResNet50, EfficientNet, or deeper LSTM variants. The study was conducted using locally available hardware, which limited both the experimentation with larger models and their testing in production-like environments—e.g., real-time emotion recognition using full video datasets. As a result, the study assumed more of a comparative testing character, similar to educational or prototyping settings, rather than a full-scale industrial validation.

All of these limitations should be taken into account when interpreting the results. The project's goal was not to achieve the highest possible classification accuracy, but rather to explore the practical viability of various approaches in conditions similar to those faced by research, therapeutic, or educational teams without access to advanced computational resources. Although the results are promising, they serve as a starting point for further research—particularly studies involving more diverse datasets, extended training, and models tailored to specific cultural and technological contexts.

Despite the increasing accuracy and availability of facial emotion recognition (FER) systems, their application raises significant ethical concerns—especially in contexts where decisions based on emotion recognition may have real consequences for individuals. Even when such technologies are deployed with good intentions, they may infringe upon fundamental values such as privacy, autonomy, equality, and human dignity.

One of the most frequently cited issues is the violation of emotional privacy. Emotions are inherently personal, and their recognition—especially when done without a person's awareness or consent—can constitute a form of surveillance. Cavoukian emphasizes that privacy is not about hiding information but about having control over what is shared and with whom. FER systems deployed in workplaces, educational

platforms, or public spaces—without the informed consent of users—may undermine this right, contributing to a sense of continuous monitoring (Cavoukian, 2019: 5).

Another serious concern involves system errors and algorithmic bias. Research has shown that FER systems are more likely to misclassify the emotions of individuals with darker skin tones—for instance, labeling neutral expressions as negative emotions such as anger (Buolamwini & Gebru, 2018: 4). This occurs in part because the training datasets used to build neural networks are predominantly composed of images of white individuals. Consequently, these systems may perpetuate stereotypes and lead to unjust outcomes in workplaces, educational institutions, or even legal proceedings trajectory (Boyd & Andalibi, 2023: 5).

Even high-performing systems are not without significant challenges. Emotion recognition is a probabilistic process, and the notion that an algorithm can "know the truth" about a person's emotional state is heavily contested in scientific literature (Barrett, 2017: 12). Studies show that emotions are deeply dependent on cultural, social, and psychological contexts—dimensions that algorithms are fundamentally incapable of fully capturing (Barrett, 2017: 18). Moreover, the illusion of objectivity may lead to FER results being treated as infallible, thereby marginalizing human voice and limiting opportunities for appeal (Keyes, 2019: 14). FER technologies also carry the risk of function creep—the gradual broadening of a system's use beyond its original intent. Systems initially designed to monitor employee well-being may later be used to assess productivity, diagnose mental illnesses, or—in the most controversial scenarios—predict criminal tendencies (Mohammad, 2022: 244). Such applications not only oversimplify the complexity of human behavior but also risk pathologizing natural emotional responses and relinquishing control over how and by whom results are interpreted (Keyes, 2019: 15). FER can also distort power dynamics in professional relationships. Employees subjected to continuous emotional monitoring may experience increased stress, pressure to conform emotionally, and uncertainty about how results may impact their career trajectory (Boyd & Andalibi, 2023: 5) In such contexts, the notion of "voluntary consent" becomes questionable, as the realistic possibility of refusal is often non-existent (Buolamwini & Gebru, 2018: 5).

Ultimately, the advancement of FER raises the question of whether every domain of life should be automated. Emotions are profoundly human phenomena—shaped by history, experience, and socio-cultural relationships. Automating their recognition invites not only technical questions but also philosophical ones: Can a machine truly understand a human being? Should it? As Mohammad (2022: 246) points out, the key issue is not whether we can, but whether we should. Ethical implementation of FER requires more than technical safeguards. It demands participatory design principles, system transparency, acknowledgment of cultural and social diversity, and—most importantly—humility in the face of technological limitations and human complexity (Mohammad, 2022: 247).

Future research should prioritize the use of more diverse datasets—those that include color images in higher resolution and feature individuals from a wide range of cultural and ethnic backgrounds. Additionally, the scope of analysis should be expanded to include multimodal data—combining facial expressions with vocal tone, gestures, or physiological signals—to enhance emotion recognition accuracy in complex, real-world communicative situations. A compelling direction would also be to investigate the

behavioral effects of FER interfaces themselves—for example, whether awareness of being monitored alters the way people express emotions.

Conclusion

This analysis has demonstrated that facial emotion recognition (FER) technologies can serve as a valuable complement to diagnostic tools in the context of remote work and psychological well-being. While classical methods offer fast and cost-effective implementation despite certain limitations, deep learning approaches provide significantly higher accuracy and flexibility in dynamic conditions. However, despite growing technological accessibility, further research is necessary to assess these systems' effectiveness in real-world environments and to ensure their ethical implementation. Future directions should consider integrating FER models into multimodal systems and developing tools that support mental well-being across various organizational levels.

Both our study and the review of existing literature point to a clear potential for applying FER as a technology supporting remote work environments. Most importantly, such systems can systematically and more objectively identify early symptoms of declining mental well-being. This enables the implementation of preventive actions—such as recommending breaks, adjusting work rhythms, facilitating communication with the team or supervisor, or, in more advanced scenarios, signaling the need for psychological consultation. Potential benefits also include adapting the work environment to a user's current emotional state, which may improve concentration, reduce stress, and even enhance efficiency and job satisfaction.

Nevertheless, it is critical to recognize that FER technology touches upon one of the most intimate aspects of the human experience—emotion. Introducing such systems into the workplace carries the risk of both technological and social abuses. The most serious of these is the potential instrumentalization of emotion—treating it as another productivity metric rather than an authentic expression of an individual's mental state. Particularly concerning are potential abuses of power, where emotional data is used without the employee's informed consent or beyond the original purpose for which it was collected. Examples include using FER to evaluate *loyalty*, monitor *motivation* in real time, or even select candidates during recruitment based on their emotional responses.

Several other concerns also emerge: technical ones—related to classification accuracy and potential errors; psychological ones—concerning the impact of constant monitoring on well-being and the authenticity of emotional expression; and social ones—such as cultural differences in emotional expression, which may lead to algorithmic biases and misinterpretations. Moreover, there is a tangible risk of function creep—expanding the original use of FER systems into ethically questionable areas, such as behavior control, diagnosing mental disorders without consent, or even profiling individuals based on perceived emotional risk.

Therefore, as authors, we recommend that any implementation of FER technology in work settings—especially remote environments—be preceded by systemic regulation. This should include not only legal frameworks but also ethical standards, codes of conduct, auditing mechanisms, and a genuine option for individuals to opt out without facing negative consequences. Only an approach based on transparency, voluntariness, and participatory design can ensure that these technologies serve to support humans rather than control them.

This study contributes to the development of practical FER applications in the realities of remote work, demonstrating that even simplified models—when appropriately tuned—can effectively support the monitoring of psychological well-being without the need for advanced or costly systems. In the future, research on multimodal FER systems, combining visual data with voice, text, or physiological signals, may prove particularly valuable. Such approaches could not only improve the accuracy of emotion recognition but also reduce the risk of misinterpretation stemming from the one-sided analysis of facial expressions.

Technology that can perceive emotions has the potential to become a powerful ally—provided it is accompanied by attentiveness to the human being.

REFERENCES

Aslam, A., Hussain, B. (2021). Emotion recognition techniques with rule based and machine learning approaches, pp. 1-27.

Baes, N., Speagle, H., Haslam, N. (2022). Has psychology become more positive? Trends in language use in article abstracts. *Frontiers in Psychology*, 13, pp. 1–8.

Barrett, L.F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. New York: Houghton Mifflin Harcourt.

Becker, W.J., Belkin, L.Y., Tuskey, S.E. (2022). Surviving remotely: How job control and loneliness during a forced shift to remote work impacted employee work behaviors and well-being. *Human Resource Management*, 61(1), pp. 449-464.

Berry, Y. (2023). Examining the workplace wellbeing of elementary school teachers whose students participated in a social and emotional learning program. *Doctoral thesis*, Pepperdine University.

Boyd, K.L., & Andalibi, N. (2023). *Automated emotion recognition in the workplace: How proposed technologies reveal potential futures of work*. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), pp. 1–27.

Brown, A., Leite, A.C. (2023). The effects of social and organizational connectedness on employee well-being and remote working experiences during the COVID-19 pandemic. *Journal of Applied Social Psychology*, 53(1), pp. 134–152.

Buolamwini, J., Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, pp. 1–15.

Cavoukian, A. (2019). *Privacy by Design: The 7 Foundational Principles*. Toronto: Information and Privacy Commissioner of Ontario.

Czapiński, J. (2004). Psychologiczna teoria dobrostanu. In: J. Czapiński (Ed.), *Psychologia pozytywna*. Warsaw: PWN, pp. 19–42.

Dalal, N., Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1(1), pp. 886–893.

Das, N., Das, S. (2023). Epoch and accuracy based empirical study for cardiac MRI segmentation using deep learning technique. *PeerJ*, 11, pp. 1–15.

Dietterich, T.G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science*, 1857, pp. 1–15.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14(2), pp. 179–211.

Goodfellow, I., et al. (2013). Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64, pp. 59-63.

Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., Li, B. et al. (2024). Development and application of emotion recognition technology — a systematic literature review. *BMC Psychology*, 12:95, pp. 1-15.

Hadjar, H., Vu, B., Hemmje, M. (2025). TheraSense: Deep Learning for Facial Emotion Analysis in Mental Health Teleconsultation. *Electronics*, 14(3), pp. 1-17.

Hajric, E., Najar Arevalo, F., Bruce, L., Smith, F.A., Michael, K. (2024). Facial Emotion Recognition in the Future of Work: Social Implications and Policy Recommendations. *IEEE Transactions on Technology and Society*, pp. 1–10.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Hersh, W. (2005). Evaluation of biomedical text-mining systems: lessons learned from information retrieval. *Brief Bioinformatics*, 6(4), pp. 344–356.

Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), pp. 1735–1780.

Hossain, M.S., Muhammad, G. (2017). An Emotion Recognition System for Mobile Applications. *IEEE Access*, 5, pp. 2281-2287.

Ismail, S.N.M.S., Razak, S.F.A., & Aziz, N.A.A. (2024). *Transfer learning for improved electrocardiogram diagnosis of cardiac disease: exploring the potential of pre-trained models*. Bulletin of Electrical Engineering and Informatics, 13(5), pp. 1–15.

Kas, M., Ruichek, Y., & Messoussi, R. (2021). *New framework for person-independent facial expression recognition combining textural and shape analysis through new feature extraction approach*. Information Sciences, 572, pp. 10–25.

Kau, F.S., Flotman, A.P. (2025). The subjective well-being experiences of mine employees in a South African mining organisation. *SA Journal of Industrial Psychology*, 51(0), pp. 1–14.

Khalfallah, J., Ben Hadj Slama, J. (2015). Facial expression recognition for intelligent tutoring systems in remote laboratories platform. *Procedia Computer Science*, 73, pp. 274-281.

Khan, U.A., Xu, Q., Liu, Y., Lagstedt, A., Alamäki, A., Kauttonen, J. (2024). Exploring contactless techniques in multimodal emotion recognition: insights into diverse applications, challenges, solutions, and prospects. *Multimedia Systems*, 30, pp. 1-48.

Keyes, O. (2019). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), pp. 1-22.

Lampinen, A. (2024). Employee Well-Being Management in the Post-COVID Era – Insights from HR Professionals in Large Information Technology Companies in Finland. *Master's thesis*, Aalto University.

LeCun, Y., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp. 2278–2324.

Leong, D.C.P. (2022). Factors Affecting the Psychological Well-Being of Educators: A Study on Private College Lecturers Succeeding COVID-19 Pandemic. *Master's thesis*, Borneo Journal of Social Science & Humanities, pp. 11-31.

Li, S., Deng, W. (2022). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, pp. 1195-1215.

Minaee, S., Minaei, M., Abdolrashidi, A. (2021). Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *Sensors*, 21(9), 3046, pp. 1-16.

Mohammad, S. (2022). Ethics Sheet for Automatic Emotion Recognition. *Computational Linguistics*, 48(2), pp.239–278.

Mollahosseini, A., Chan, D., Mahoor, M.H. (2016). Going deeper in facial expression recognition using deep neural networks. *IEEE Winter Conference on Applications of Computer Vision*. pp.1-10.

Nozari, Zahra & Seyedsalehi, Sadaf (2024). *Building Bridges in Digital Spaces – Enhancing the Sense of Belonging among Remote Employees in a Multinational Company*. Master Thesis, University of Gothenburg.

O'Hare, D., Gaughran, F., Stewart, R., Da Costa, M.P. (2024). A cross-sectional investigation on remote working, loneliness, workplace isolation, well-being and perceived social support in healthcare workers. *BJPsych Open*, 10(1), pp. 1–6.

Palaniswamy, S. (2019). *A robust pose & illumination invariant emotion recognition from facial images using deep learning for human-machine interface*. IEEE International Conference on Consumer Electronics (ICCE), pp. 1–6.

Pataki-Bittó, F., Kun, Á. (2022). Exploring differences in the subjective well-being of teleworkers prior to and during the pandemic. *International Journal of Workplace Health Management*, 15(3), pp. 320-338.

Rajan, S., Chenniappan, P., Devaraj, S. (2020). Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM. *IET Image Processing*, pp. 1373–1381.

Rathour, N., Khanam, Z., Gehlot, A., Singh, R., Rashid, M., AlGhamdi, A.S., Alshamrani, S.S. (2021). Real-Time Facial Emotion Recognition Framework for Employees of Organizations Using Raspberry-Pi. *Applied Sciences*, 11(22), pp. 1-17.

Rodríguez-Leudo, Andrea Lucía & Navarro-Astor, Elena (2024). *Workplace happiness in architectural companies in the city of Valencia: a gender comparison*. Doctoral Thesis, Universitat Politècnica de València.

Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), pp. 1069–1081.

Seligman, M. (2011). *Pełnia życia*. Poznań: Media Rodzina.

Simonyan, K., Zisserman, A., (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICLR)*, ss. 1–14.

Viola, P., Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1(1), pp. 511–518.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition, pp. 1-15.

Yen, C.T., & Li, K.H. (2022). *Discussions of different deep transfer learning models for emotion recognitions*. IEEE Xplore, pp. 1–6.

Yi, D., Ahn, J., & Ji, S. (2020). *An effective optimization method for machine learning based on ADAM*. Applied Sciences, 10(3), 1073.

Zimnoch, I. (2024). Dobrostan kadry zarządzającej w sektorze ekonomii społecznej oraz w sektorze biznesowym. *Master's thesis*, WSB Merito University in Poznań.