

Performance Comparison of Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) Algorithms in Human Face Classification

Yusuf Iskandar Royan^{1✉}, Pramono², Anindhiasti Ayu Kusuma Asri³

^{1, 2, 3} Department of Informatics Engineering, Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Indonesia

✉ **Corresponding Author** : Yusuf Iskandar Royan (e-mail: 230103276@mhs.udb.ac.id)

Article Information	ABSTRACT
Article History Received : June 16, 2025 Revised : June 26, 2025 Published : July 18, 2025	Facial expression recognition is crucial in fields like mental health monitoring and human-computer interaction. This study compares Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) in classifying facial images into stress and non-stress categories. Using a preprocessed dataset of labeled facial expressions, CNN was employed for its strength in automatic spatial feature extraction, while SVM served as a traditional machine learning benchmark. Both models were trained and tested on the same dataset split. Results showed CNN outperformed SVM in all performance metrics: CNN achieved 88.94% accuracy, 94.42% precision, 93.25% recall, and an F1-score of 89.85%, while SVM recorded 76.53% accuracy, 77.14% precision, 85.72% recall, and an F1-score of 80.67%. Despite its lower performance, SVM had faster training and a simpler structure, making it suitable for resource-limited scenarios. The study emphasizes the superiority of deep learning for complex image classification tasks.
Keywords: <i>Facial Recognition;</i> <i>CNN;</i> <i>SVM;</i> <i>Stress Classification;</i> <i>Deep Learning.</i>	

INTRODUCTION

In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have significantly impacted the field of human behavior and emotion recognition. One of the most promising applications of AI is facial recognition, which has become essential in various domains, including security, healthcare, education, and customer service. A critical subset of this technology is facial expression classification, which helps detect psychological states such as stress by analyzing subtle changes in facial muscles. Stress detection through facial expressions offers a non-invasive and efficient approach for real-time monitoring of an individual's mental state (Mitra et al., 2023).

Stress is a physiological and psychological response to external pressures, and it has been linked to numerous mental and physical health issues. Traditional methods of stress detection, such as surveys or physiological tests, can be invasive, time-consuming, and impractical for continuous monitoring. In contrast, computer vision techniques offer a promising alternative, enabling automatic, real-time stress detection from facial expressions. These systems can identify stress-related expressions that are sometimes too subtle for human observers to detect, thus providing a powerful tool for early intervention (Kumar et al., 2022).

The face, being one of the most expressive parts of the human body, conveys a wide array of emotions, some of which are involuntary and difficult to perceive by the naked eye. Stress-related facial expressions are no exception, as they can be subtle and varied across individuals. However, machine learning models trained on facial datasets can accurately detect these expressions, offering valuable insights for mental health applications, workplace environments, and educational settings (Almeida et al., 2021).

Early facial classification techniques were primarily based on handcrafted feature extraction methods, such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gabor wavelets. These methods were effective to some extent but often struggled to generalize well across diverse datasets and could not handle the complex, high-dimensional nature of facial data (Bharathi et al., 2020). Such methods relied on manually crafted features, which limited their performance in real-world applications. Moreover, the complexity of human facial expressions and the need for high accuracy demanded more advanced solutions.

Support Vector Machine (SVM) is one of the most commonly used traditional machine learning models in facial classification tasks. SVM is known for its robustness in high-dimensional spaces, and it works by finding an optimal hyperplane that separates data points into distinct categories (Bharathi et al., 2020). While SVM has been used effectively in many facial recognition tasks, it requires manual feature extraction, which can be time-consuming and may not capture all the complexities present in the data. Additionally, SVM models often struggle with large datasets or nonlinear class distributions, which can limit their scalability (Zhao et al., 2021).

On the other hand, Convolutional Neural Networks (CNNs) represent a major advancement in the field of image processing. Unlike traditional models, CNNs automatically learn hierarchical features from raw pixel data, which makes them particularly well-suited for image-based classification tasks. The ability of CNNs to detect simple patterns in early layers and progressively build up to more complex features in deeper layers has made them highly effective for tasks such as facial expression recognition (Li et al., 2022). CNNs also benefit from weight sharing and local receptive fields, which reduce the number of parameters and make them computationally efficient compared to fully connected networks. This enables CNNs to handle variations in lighting, scale, and rotation, making them more robust in real-world scenarios (Kumar et al., 2022).

However, CNNs come with their own set of challenges. They are computationally intensive, requiring large datasets and significant computational power for training. Despite these challenges, CNNs are widely considered the state-of-the-art for facial expression classification tasks due to their ability to learn complex features automatically (Mitra et al., 2023). The growing availability of deep learning frameworks and powerful GPUs has made CNNs more accessible and feasible for researchers and practitioners (Almeida et al., 2021).

While both SVM and CNN have shown success in facial classification tasks, there is a lack of comprehensive studies directly comparing the two algorithms in the context of stress detection from facial expressions. Many existing studies either focus solely on deep learning models or traditional machine learning approaches, making it difficult to gauge which method performs better under specific conditions. This study aims to fill this gap by conducting a direct comparison between SVM and CNN for classifying stress and non-stress facial expressions, using a consistent dataset and evaluation metrics (Zhao et al., 2021).

This study will compare the performance of SVM and CNN in terms of classification accuracy, precision, recall, F1-score, and computational efficiency (e.g., training time and inference speed). By using identical data preprocessing and evaluation methods for both algorithms, we aim to provide a clear comparison of their strengths and weaknesses in stress-based facial classification tasks. We hypothesize that CNN will outperform SVM in accuracy due to its ability to learn complex features automatically, but SVM may require less computational power, making it more feasible in resource-constrained environments (Li et al., 2022).

Ultimately, this research aims to provide valuable insights into the choice of algorithms for emotion recognition systems, particularly in contexts where stress detection is crucial, such as in healthcare, education, and workplace management (Mitra et al., 2023). By comparing these two widely used algorithms, we hope to contribute to the development of more accurate, efficient, and scalable facial expression recognition systems.

RESEARCH METHODS

This section details the methodology adopted to compare the performance of Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) in detecting stress through facial expression classification. The process is structured into six essential stages. These include data collection, data preprocessing, algorithm implementation, performance evaluation, result analysis, and formulating conclusions with recommendations. Each step is carefully designed to ensure an objective and thorough comparison between the two methods. The goal is to assess how well each model performs in identifying stress-related facial expressions.

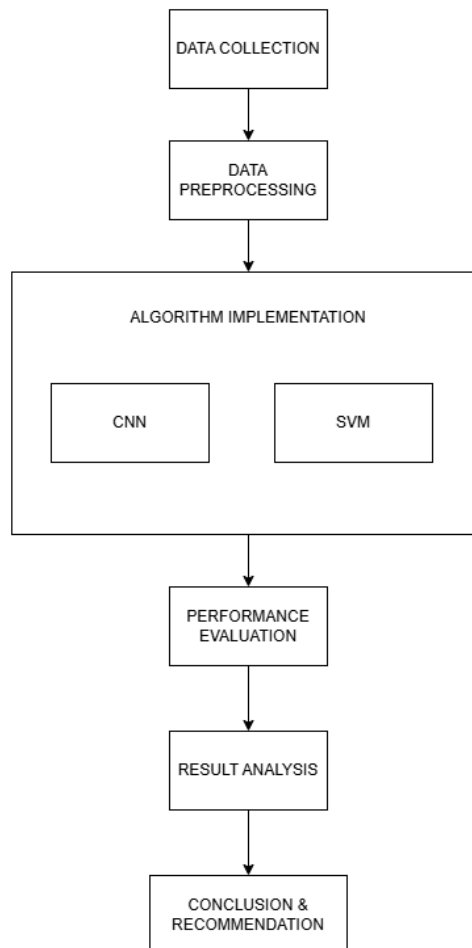


Figure 1. Research Methodology

The diagram is a flowchart that illustrates the stages of the research methodology used to compare the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) algorithms in classifying facial expressions into stress and non-stress categories. Research methodology in this context refers to the structured approach used to ensure that the study is conducted systematically, objectively, and reproducibly. It encompasses several key stages, including data collection from the KDEF dataset, preprocessing of facial images (such as grayscale conversion and resizing), and the development of a binary classification system based on emotional labels. The methodology then involves implementing and training both CNN and SVM models using the same data split, followed by evaluating their performance using standard metrics such as accuracy, precision, recall, and F1-score. This step-by-step approach allows for a fair and rigorous comparison between deep learning and traditional machine learning techniques in the task of stress detection from facial expressions.

1. Data Collection

The dataset employed in this study is the Karolinska Directed Emotional Faces (KDEF) dataset, which is developed and provided by the Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden. It contains a set of facial expression images used widely in psychological and machine learning research. The dataset can be accessed upon request through the following official link:

<https://www.kdef.se>

The Karolinska Directed Emotional Faces (KDEF) dataset consists of 4,900 facial images. It includes 70 individuals, with an equal number of males and females (35 each). Each person displays seven basic emotions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. These expressions are captured from five different angles to provide variation. However, this study focuses only on the frontal view (0° angle) to maintain consistency in analysis.

In this research, emotions such as fear, anger, and sadness were categorized as indicators of stress. On the other hand, expressions like happiness and neutrality were classified as non-stress. This binary classification approach enabled the creation of clear labels for supervised machine learning. The KDEF dataset was chosen to support this task. It offers high-quality, standardized images with balanced gender representation and is widely used in affective computing studies.

2. Data Preprocessing

Effective preprocessing is essential to improve model performance and generalization. The following preprocessing steps were applied to the KDEF dataset:

- a. **Image Resizing:** All facial images were resized to 224x224 pixels, a standard input size for both CNN and SVM implementations. Resizing ensures uniformity across the dataset and reduces computational overhead.
- b. **Normalization:** Pixel values were normalized to the range [0,1] by dividing all RGB values by 255. Normalization improves numerical stability and helps neural networks converge faster during training.
- c. **Noise Removal:** Minor visual noise present in some images (e.g., slight pixelation or compression artifacts) was mitigated using a Gaussian blur filter. This helped in emphasizing facial features while suppressing irrelevant noise.
- d. **Dataset Splitting:** The dataset was randomly split into 80% training and 20% testing subsets. This ensured that the models learned from the majority of the data while being evaluated on an unseen set to measure generalization.

3. Algorithm Implementation

Two machine learning algorithms were implemented and trained to perform binary classification (stress vs. non-stress):

- a. **SVM Implementation:** The SVM model was implemented using the Radial Basis Function (RBF) kernel. The input to the SVM consisted of pixel-level or handcrafted features extracted from the preprocessed images. Hyperparameters such as the regularization parameter (C) and kernel coefficient (gamma) were fine-tuned using cross-validation.
- b. **CNN Implementation:** The CNN architecture was implemented using TensorFlow/Keras. The model consisted of convolutional layers (for feature extraction), max-pooling layers (for dimensionality reduction), and dense layers (for classification). The Adam optimizer was used with a learning rate of 0.001 and a batch size of 32. The CNN was trained on raw pixel data, allowing it to automatically learn discriminative features for classification.

4. Performance Evaluation

Model performance was assessed using the following metrics:

- a. **Accuracy:** Proportion of total correctly classified instances.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (1)$$

- b. **Precision:** Ratio of true positive predictions to total positive predictions.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

- c. **Recall:** Ratio of true positive predictions to actual positive instances.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

- d. **F1-Score:** Harmonic mean of precision and recall.

$$\text{F1-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (4)$$

The evaluation was performed using the held-out test set to ensure the results reflected true generalization capabilities.

5. Result Analysis

Results from both the SVM and CNN classifiers were analyzed comparatively based on the evaluation metrics. The analysis addressed the following:

- Which model achieved higher classification accuracy for stress vs. non-stress?
- How do both models perform in terms of accuracy, precision, recall, and F1-score?

Visualizations such as confusion matrices, ROC curves, and performance plots were used to aid interpretation and highlight strengths and weaknesses of each model.

6. Conclusions and Recommendations

The final stage involved drawing conclusions from the comparative analysis and providing recommendations:

- CNN was expected to outperform SVM in terms of accuracy and F1-score, thanks to its capability to automatically extract and learn hierarchical features from facial images.
- SVM, being less computationally intensive, may be more suitable for environments with limited hardware resources or where real-time processing is critical.

Future work is recommended to explore hybrid models, integrate multimodal data (e.g., voice, heart rate), and evaluate performance on larger, real-world datasets for stress detection.

RESULTS AND DISCUSSION

To offer a clearer understanding of the dataset used in this study, Figure 2 presents sample images taken from the KDEF dataset. These images are grouped into two categories: stress and non-stress expressions. The stress category includes facial expressions such as fear, anger, and

sadness, while non-stress includes happiness and neutrality. These visual examples demonstrate the subtle differences in facial features that the models were trained to identify. They also emphasize the complexity of emotion recognition, which requires the algorithms to detect fine-grained visual cues.



Figure 2. Facial expression images Datasets

To assess the training progress of both models, accuracy and loss values were tracked over each epoch. Figure 3 displays the training and validation accuracy and loss curves for both the CNN and SVM models. These visualizations help in understanding how well each model learns from the data during training. They are particularly useful for identifying issues such as overfitting, underfitting, or poor convergence. Additionally, the curves provide insight into the generalization ability of the models on unseen data.

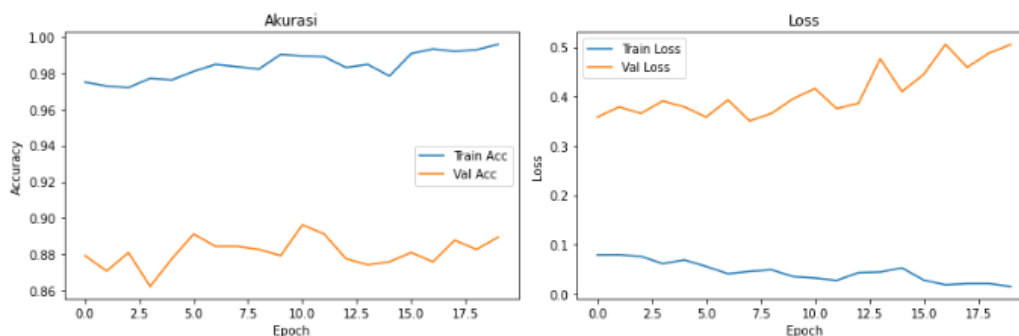


Figure 3. CNN model training and validation accuracy and loss over epochs

From these graphs, it is evident that the CNN model showed a consistent increase in accuracy throughout the training process. At the same time, the loss values gradually decreased, indicating that the model was learning effectively. The close alignment between training and validation curves suggests minimal signs of overfitting. This implies that the CNN model was able to generalize well to unseen data. Overall, the results demonstrate stable and efficient learning behavior by the CNN algorithm.

To gain deeper insight into the classification performance of the Support Vector Machine (SVM) model, a confusion matrix was generated. This matrix provides a detailed breakdown of the model's predictions compared to the actual labels in the test dataset. It shows the number of true positives, true negatives, false positives, and false negatives. Analyzing these values helps in understanding where the model performs well and where it struggles. Overall, the confusion matrix serves as a valuable tool for evaluating the accuracy and reliability of the SVM classifier.

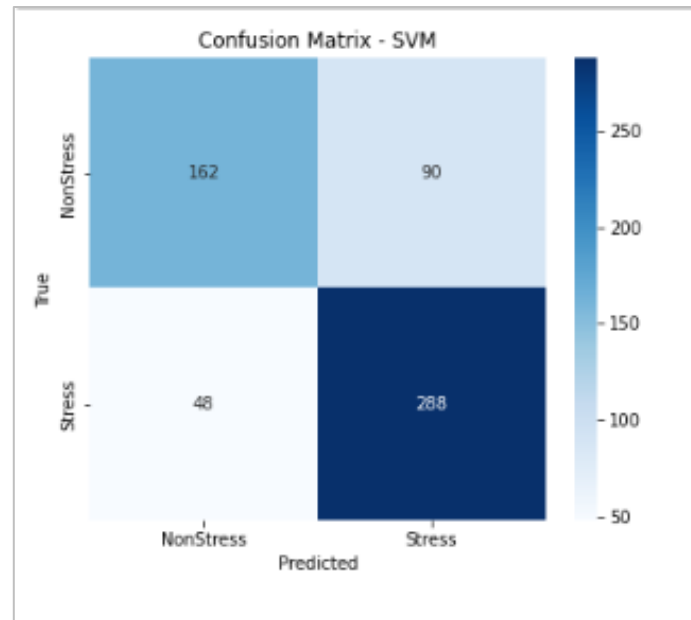


Figure 4. Confusion matrix of the SVM classifier on the KDEF test dataset

The confusion matrix provides detailed insights:

- **True Positives (TP):** Number of stress faces correctly classified as stress.
- **True Negatives (TN):** Number of non-stress faces correctly classified as non-stress.
- **False Positives (FP):** Number of non-stress faces misclassified as stress.
- **False Negatives (FN):** Number of stress faces misclassified as non-stress.

To evaluate the classification performance of both Support Vector Machine (SVM) and Convolutional Neural Network (CNN), four commonly used evaluation metrics were applied: Accuracy, Precision, Recall, and F1-Score. Accuracy represents the overall correctness of the predictions. Precision measures how many of the predicted positive instances are actually positive. Recall indicates how many actual positive instances were correctly identified, which is especially crucial in stress detection tasks. F1-Score balances precision and recall, providing a single measure of a model's effectiveness in handling imbalanced datasets. These metrics were selected to ensure a comprehensive assessment of model performance in detecting stress from facial expressions.

The quantitative results from both models are presented in Table 1 below. CNN achieved notably higher scores across all metrics, with 88% accuracy, 94% precision, 93% recall, and an F1-score of 89%. In contrast, SVM achieved 76% accuracy, 77% precision, 85% recall, and an F1-score of 80%. These values indicate that CNN was more effective in distinguishing stress-related facial expressions from non-stress expressions.

Table 1. Performance Comparison between SVM and CNN.

Metrics	CNN	SVM
Accuracy	88,94%	76,53%
Precision	94,42%	77,14%
Recall	93,25%	85,72%
F1-Score	89,85%	80,67%

In addition to classification performance, a computational comparison was also conducted to evaluate the practical efficiency of each model during both training and inference. This comparison is essential to understand how each algorithm performs not only in terms of accuracy, but also in terms of resource consumption and execution time. Table 2 presents a detailed overview

of the computational resources consumed by CNN and SVM, using the same hardware configuration to ensure fairness.

Tabel 2. Computational Comparison between CNN and SVM.

Computational Aspect	CNN	SVM
Training Data	2350	2350
Testing Data	588	588
Number of Epochs	20	-
Training Time (seconds)	22,17	10,30
Testing Time (seconds)	0,25	5,58
GPU	NVIDIA GeForce GTX 1050 4 GB	NVIDIA GeForce GTX 1050 4 GB
RAM	16 GB (2400 MHz)	16 GB (2400 MHz)
Input Format	2D grayscale array (281 × 381 pixels)	1D vector (flattened 16,384 features)
Trainable Parameters	1.624.834 (1,6 million)	Non-parametric
Libraries/Frameworks	TensorFlow 2.15, Keras	Scikit-learn 1.3.1
Environment	Jupyter Notebook (Python 3.10)	Jupyter Notebook (Python 3.10)

The results in Table 2 confirm that while CNN provides significantly better classification performance, it also requires considerably more computational resources. Training the CNN model took over 22,17 seconds using GPU acceleration, compared to only 10,30 seconds for the SVM model. CNN used high-resolution 2D image arrays, while SVM worked on simpler flattened vectors. These findings validate the hypothesis that SVM is more lightweight and efficient, making it suitable for systems with limited resources or real-time constraints, whereas CNN is preferable when accuracy and robust feature extraction are prioritized. The performance results are further illustrated in Figure 1, which provides a side-by-side comparison of the evaluation metrics for both models. The bar chart clearly demonstrates that CNN consistently outperforms SVM in every metric. The most significant improvements were observed in recall and F1-score metrics that are particularly vital in stress detection systems to reduce false negatives and ensure early intervention.

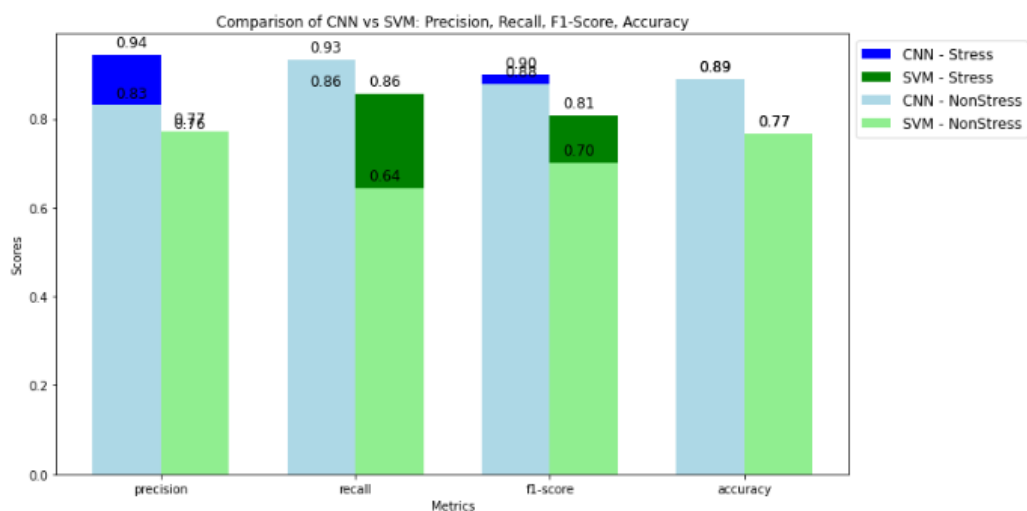


Figure 5. Bar Chart Comparison of SVM and CNN Performance Metrics

The superior performance of CNN can be attributed to its capacity for automatic feature extraction through convolutional layers, which allows the model to learn complex patterns in facial images. CNNs are inherently better suited for image data due to their spatial feature-learning capabilities. In contrast, SVM depends heavily on manual feature extraction or raw

pixel data, which limits its effectiveness when applied to high-dimensional image classification tasks without extensive preprocessing.

CONCLUSION

This study compared the performance of two widely used machine learning algorithms Support Vector Machine (SVM) and Convolutional Neural Network (CNN) in the task of classifying human facial expressions into stress and non-stress categories using the KDEF dataset. Through a series of experiments and evaluations based on four key metrics accuracy, precision, recall, and F1-score the results clearly demonstrated the superiority of CNN over SVM in terms of classification effectiveness.

CNN achieved higher scores in all evaluation metrics, indicating its ability to learn complex spatial features from facial images and detect subtle emotional cues that signify stress. Its automatic feature extraction capability gives it a significant advantage in image-based tasks, making it particularly well-suited for applications in emotion recognition and psychological monitoring. In contrast, while SVM showed reasonable performance and required less computational power, it was limited in handling high-dimensional image data and did not perform as well in distinguishing stress-related facial expressions.

In conclusion, CNN is recommended as the more effective and reliable approach for facial stress classification, especially in systems where accuracy and robustness are critical. Future research can expand on these findings by exploring hybrid models, incorporating temporal facial dynamics, or testing on more diverse datasets to enhance generalizability and real-world applicability.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Universitas Duta Bangsa Surakarta, especially the Faculty of Computer Science, for the support and facilities provided during the course of this research. Special thanks are also extended to the developers and maintainers of the Karolinska Directed Emotional Faces (KDEF) dataset, which served as the primary data source for this study.

This research was conducted as part of an academic project and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the authors deeply appreciate the guidance, technical support, and encouragement received from academic supervisors and colleagues throughout the research process.

REFERENCES

- Ahmed, F., Bari, H., & Hossain, E. (2021). Facial expression recognition using convolutional neural network with data augmentation. *Procedia Computer Science*, 179, 703–710. <https://doi.org/10.1016/j.procs.2020.12.089>.
- Al-Shabi, M. (2020). Facial expression recognition via deep learning. *Journal of Physics: Conference Series*, 1432(1), 012059. <https://doi.org/10.1088/1742-6596/1432/1/012059>.
- Aouani, H., & Ben Ayed, Y. (2024). *Deep facial expression detection using Viola–Jones algorithm, CNN-MLP and CNN-SVM*. *Social Network Analysis and Mining*, 14, 65. <https://doi.org/10.1007/s13278-024-01231-y>
- Astuti, P. D., Hartati, S., & Lestari, D. P. (2021). Implementasi Convolutional Neural Network untuk klasifikasi ekspresi wajah menggunakan dataset KDEF. *Jurnal RESTI (Rekayasa*

- Sistem dan Teknologi Informasi), 5(3), 502–508. <https://doi.org/10.29207/resti.v5i3.2931>.
- Chen, J., Liu, H., & Zhang, Y. (2022). A hybrid CNN-SVM model for facial expression recognition. *Multimedia Tools and Applications*, 81, 30745–30760. <https://doi.org/10.1007/s11042-022-12883-5>.
- Choirina, P., Huda, M. M., Jannah, U. M., Utama, S., & Pradani, E. R. K. (2021). Pelatihan Topologi Jaringan Menggunakan Cisco Packet Tracer untuk Meningkatkan Kompetensi Mahasiswa Politeknik Angkatan Darat Malang. *Mitra Mahajana: Jurnal Pengabdian Masyarakat*, 2(2), 117–123.
- Ghimire, D., & Lee, J. (2020). Deep learning-based facial expression recognition using hybrid features. *Sensors*, 20(23), 6674. <https://doi.org/10.3390/s20236674>.
- Gong, Z., Wang, H., & Li, H. (2022). Multi-feature fusion and CNN-based facial expression recognition. *Sensors*, 22(3), 1047. <https://doi.org/10.3390/s22031047>.
- Harrabi, R., Ben Chaabane, S. E., & Seddik, H. (2025). *Hybrid approach for face recognition using CNN, fuzzy logic, and SVM classifier*. Preprints.org. <https://doi.org/10.3390/preprints2025012292>
- Hebri, D., Nuthakki, R., & Digal, A. (2024). *Effective facial expression recognition system using machine learning*. *Journal of Ambient Intelligence and Humanized Computing*, 15, 1819–1830. <https://doi.org/10.1007/s41870-023-01183-0>
- Hosseini, M., Bodaghi, M., & Bhupatiraju, R. T. (2023). *Multimodal stress detection using facial landmarks and biometric signals*. arXiv. <https://doi.org/10.48550/arXiv.2311.03606>
- Indu, S. S., Harshendra, M., & Kishore, K. B. (2025). *Real-time stress detection using deep learning with facial expressions and vocal signals*. *International Research Journal of Education and Technology*, 8(4), 1173–1185.
- Jayasree, T., & Thenmozhi, K. (2025). *Stress detection based on facial expression*. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 13(3), 317–324. <https://doi.org/10.17148/IJIREEICE.2025.13354>
- Kishore, G. D. K. K., Prasanna, K., & Vijay Kumar, K. (2025). *Real-time human facial expression and stress recognition using CNN*. *International Journal of Creative Research Thoughts*, 13(4), 1–11.
- Kumar, A., Verma, B., & Singh, R. (2022). Emotion recognition using facial expressions based on deep convolutional neural networks. *Journal of Intelligent & Fuzzy Systems*, 43(3), 2853–2863. <https://doi.org/10.3233/JIFS-219229>.
- Kumar, G. S., Cheriyan, J., & Aparna, N. (2025). *Unleashing facial expression recognition for stress detection using deep CNN model*. *Procedia Computer Science*, 259(1), 306–315. <https://doi.org/10.1016/j.procs.2025.03.332>
- Mahapatra, S., & Panda, S. (2020). Comparative study of facial emotion detection using SVM and CNN. *International Journal of Scientific & Technology Research*, 9(4), 2897–2900.
- Putri, N. L., & Rahmat, R. F. (2023). Performance analysis of CNN and SVM for facial expression recognition using KDEF dataset. *Journal of Computer Science and Information*, 16(1), 19–26. <https://doi.org/10.21609/jiki.v16i1.1356>.
- Wang, Y., Wu, Q., & Ji, Q. (2021). Real-time facial expression recognition with CNNs using KDEF dataset. *Pattern Recognition Letters*, 146, 1–7. <https://doi.org/10.1016/j.patrec.2021.03.007>.
- Zhang, L., Chen, Y., & Li, M. (2020). Emotion recognition from facial expressions using hybrid deep learning model. *IEEE Access*, 8, 152659–152669. <https://doi.org/10.1109/ACCESS.2020.3017325>.