The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 5, October 2025
https://doi.org/10.33022/ijcs.v14i5.5005

# Facial Expression Prediction Based on Pre-Trained ResNet50 and SVM

**Rasheed R Ihsan[1,2], Adnan Mohsin Abdulazeez[3]**
rasheed.sarky@gmail.com [1],rasheed.sarky@nawroz.edu.krd[2],adnan.mohsin@dpu.edu.krd[3]
[1]IT Dept., Technical College of Informatics, Akre University for Applied Sciences, Duhok, Iraq.
[2]College of Science, Department of Computer Science, Nawroz University, Duhok, Kurdistan Region, Iraq.
[3]Technical College of Engineering, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

**Article Information**

Received : 25 Sep 2025
Revised : 14 Oct 2025
Accepted : 30 Oct 2025

**Keywords**

Prediction, Genki, Feature extraction, ResNet50, Classification

**Abstract**

Facial expression prediction has become a vital area in computer vision, with applications spanning healthcare, security, and human-computer interaction. This study proposes a robust system for binary facial expression prediction using a combination of classical computer vision techniques and deep learning. The system employs the Haar Cascade algorithm for face detection and ResNet50, a 50-layer deep residual network, for feature extraction. Support Vector Machines (SVM) with a radial basis function kernel are used for classification. Using the 4,000 tagged images from the GENKI dataset, preprocessing and data augmentation improved the model's capacity for generalization. Experimental results demonstrate the system's effectiveness, achieving a test accuracy of 94.65%. The robust integration of classical and modern techniques ensures computational efficiency while maintaining high performance. For real-world applications, this method provides a scalable solution that tackles issues including lighting fluctuation, position, and expression variation.

 

## A. Introduction

Facial expression recognition has become a crucial aspect of computer vision and artificial intelligence, offering widespread applications in fields such as healthcare, security, entertainment, and human-computer interaction[1]. Automatically identifying emotions from facial expressions can improve security procedures, increase user experience in interactive systems, and monitor mental health. However, significant challenges remain in achieving robust recognition due to variations in lighting conditions, occlusion, pose, and individual differences in expression[1][2]. Generalization across many situations is one of the main problems in face expression prediction. Factors like facial occlusion (e.g., glasses or masks), inconsistent illumination, and dynamic poses can significantly degrade the performance of traditional recognition systems. The task is further complicated by inter-subject variability, which refers to variations in how people express the same emotion. These issues frequently plague traditional techniques, such feature-based approaches that are handcrafted, which is why more reliable, automated solutions are required[2]. Recent advancements in deep learning have shown promise in addressing these challenges, with pre-trained models providing powerful feature extraction capabilities. In particular, transfer learning has made it possible to reuse reliable feature representations from massive datasets, increasing the efficiency and generalizability of models. Data augmentation techniques, including flipping, rotation, and scaling, have further contributed to improving model robustness by simulating diverse real-world conditions[3].

The effectiveness of deep learning-based techniques for emotion recognition has been shown in earlier research. For example, [4] developed a convolutional neural network (CNN) architecture specifically designed for facial expression datasets, achieving state-of-the-art performance under varying conditions. Similarly[5], emphasized the role of attention mechanisms in guiding CNNs to key facial regions, enhancing performance on benchmark datasets. In order to further increase recognition accuracy, Zhao et al. proposed multi-task learning frameworks that simultaneously predict facial features. In order to achieve high accuracy, these contributions emphasize the significance of huge datasets, well-designed network architectures, and efficient preprocessing.

In this paper, we propose a novel approach combining the Haar Cascade algorithm for face detection, ResNet50 for feature extraction, and Support Vector Machines (SVM) for classification. By leveraging data augmentation and transfer learning, our method is designed to enhance generalization while maintaining computational efficiency. The proposed system addresses the limitations of existing models by achieving robust performance on the GENKI dataset, demonstrating its potential for real-world applications. The proposed methodology is summarized in the smart chart here below (figure 1) for better comprehension. This paper is organized as follows: Section 2 reviews related work, detailing the advancements in facial expression recognition. Section 3 describes the proposed methodology, including data preprocessing, model architecture, and training. Section 4 presents the experimental setup and evaluation metrics. Section 5 discusses the results and their implications, and Section 6 concludes with insights and future research directions.

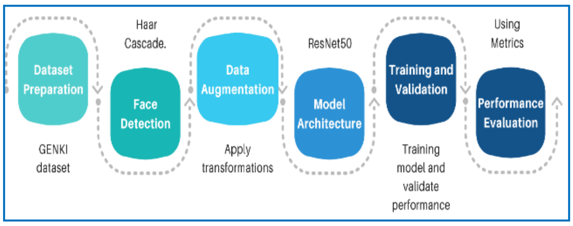


**Figure 1.** Smile and nonsmile prediction process

## B. Related Work

Facial expression recognition has attracted much attention over the last few years because of its pervasiveness and the difficulties in correctly capturing human emotions. This chapter provides a summary of recent developments in the area, with emphasis on work done in the last eight years. The research has utilized contemporary deep learning methods, big datasets, and novel architectural designs to handle the complexity of facial expression recognition. Several works have illustrated the power of deep learning models on this task. [4] introduced a deep CNN for recognizing facial expressions and achieving state-of-the-art results on the FER2013 dataset and pointed out the key role of large-scale labeled sets. [5] proposed the use of attention mechanisms for CNNs to guide them to discriminative facial parts and achieved state-of-the-art results on many benchmarks . Similarly,[6] applied a multi-task learning model to simultaneously predict facial expressions and other attributes, illustrating the benefit of shared representations . [7] contributed by applying the GENKI-4K dataset to smile detection. The study proved that binary classification datasets play a significant role in enhancing model performance on certain tasks.[8] subsequently trained a deep CNN at large scale with the AffectNet dataset and obtained better emotion prediction performance. These contributions demonstrate the significance of large and diverse datasets for high accuracy. [9] Investigated transformer-based models for binary facial expression classification, with state-of-the-art performance on the GENKI-4K dataset. Likewise, [10], suggested fusing vision transformers (ViTs) with ResNet features to enhance smile detection. [11] introduced light models specifically tailored for real-time smile detection, keeping in mind resource-limited environments [12]. Also extended model robustness by investigating how CNN performance is affected by different data augmentation methods, i.e., rotation, flipping, and zooming . These works show the importance of data augmentation and preprocessing in high accuracy scenarios. Concurrent with these developments, other studies have been focused on addressing data diversity and variability issues. To improve performance in a variety of scenarios, [13][14] developed a deep locality-preserving convolutional neural network (CNN) that highlights both local and global characteristics [15]. have looked at hybrid models for binary facial emotion detection that combine CNN and support vector machines (SVM), demonstrating their stability through experiments on various datasets.

Developments in deep learning recently have greatly enhanced systems for predicting face expressions and facial identification. In some situations for instance, changing lighting and occlusion, convolutional neural networks CNNs have been essential in performance enhancement [16][17]. These advancements joined with the integration of deep learning into emerging technologies such as 3D facial modeling which have made CNNs a dominant approach in image processing and applications of biometric [18]. The CNN-based models robustness for emotion classification using large datasets like (fer2013), achieving an overall 69.85% accuracy [19]. Furthermore, their comparative analysis of machine learning models for predicting facial beauty standards emphasized the higher performance of Support Vector Machines (SVM) and logistic regression. This work demonstrated how well SVM handles complex feature spaces, which makes it a good classifier for binary predicting face expressions [20][21]. The current work employs the approach using ResNet50 for feature extraction and paired with SVM for improved classification accuracy. This is how the current work was found to be supported and corroborated by the findings of the referenced studies.

Overall, in the past few years have seen huge progress in facial expression prediction with the availability of large-scale datasets, advances in model designs, and advanced data augmentation techniques. These studies have made it possible to design better and more robust emotion prediction systems that address the problems of lighting, variation in pose, and expression variation.

## C. Material and Methods

This section, the procedure for the suggested facial expression prediction system is described. The system integrates conventional computer vision techniques for detecting of face, deep learning for feature extraction, and for classification, Support Vector Machine (SVM). The workflow, shows in the accompanying flowchart Figure 2, ensures an efficient and systematic pipeline from evaluation to data preprocessing.

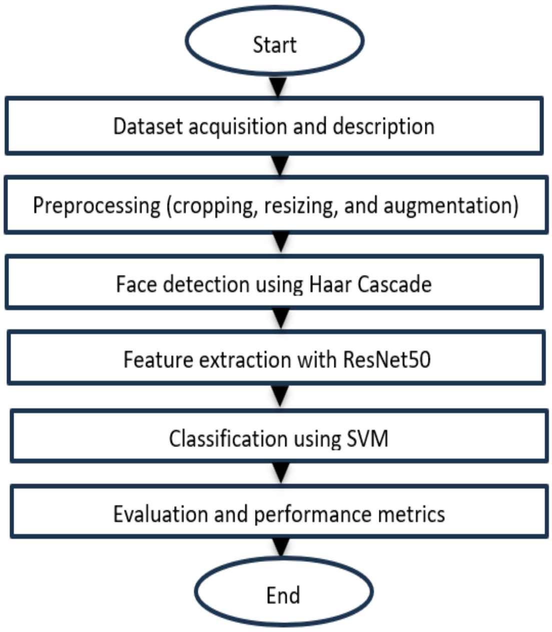


**Figure 2.** The flowchart of the process

- **Input Dataset**

The GENKI dataset, extensively used for classification of binary facial expression that consists of 4k labeled images categorized as either smiling or nonsmiling. The dataset contains various lighting conditions, facial expressions and occlusions, which is well-suited for training robust models. From the overall images, there are 2,162 samples of smiling faces (54%) and 1,838 of non-smiling faces (46%). Approximately 57.5% of these images are of males and 42.5% females, ensuring a fair combination of gender representation. The indoor/outdoor, frontal/side and occlusions (hats/glasses) are different between the dataset.

Data augmentation strategies were applied in order to improve model generalization, , including rotation, flipping, scaling, and random cropping. This approach ensures exposure to diverse scenarios, reducing overfitting and performance enhancing on unseen data. Sample of images illustrates in Fig 3 that shows variations that are provided in the material supplementary [7].

**Figure 3.** Samples of the GENKI dataset

- **Face Detection**

A classical face detection algorithm Haar Cascade was employed to detect and crop facial regions. The technique scans the image for patterns of pixel intensity that represent faces [1], images without detectable faces were excluded to maintain consistency in the dataset. By isolating the facial regions, the system decreases noise and irrelevant information, focusing only on features essential for classification.

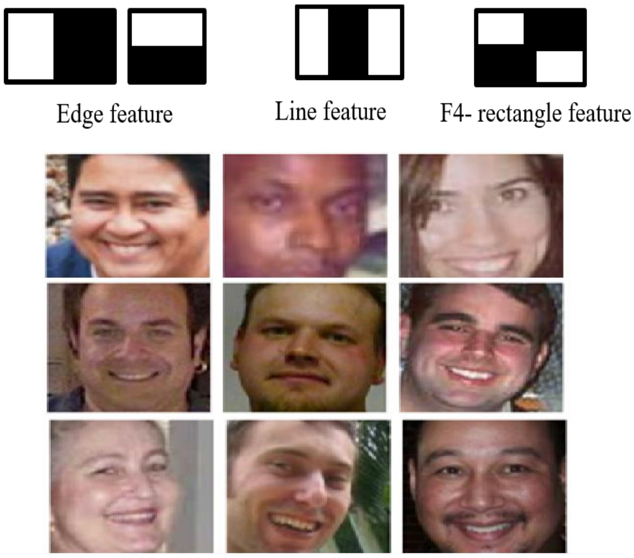


**Figure 4.** Haar Cascade, and Facial extraction

- **Preprocessing**

The detected faces were resized to 224*224 pixels to meet the input requirements of the ResNet50 model. Moreover, data augmentation techniques, for instance zooming, flipping, rotation, and shifting, were applied to increase the dataset and improve its diversity [12]. These steps prevent overfitting and improve the model's ability to generalize to unseen data.

- **Extracting Feature with ResNet50**

The ResNet50 model, a 50 layer deep residual network, was employed for extraction of feature. Pre-trained weights from ImageNet were utilized to leverage robust, hierarchical feature representations [3][9].

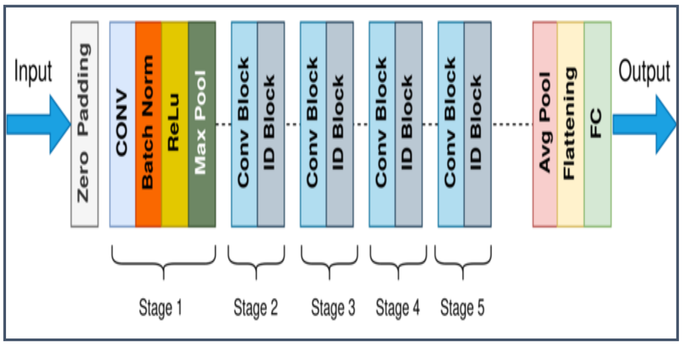


**Figure 5.** ResNet 50 Model Architecture

The output feature maps from ResNet50 were flattened into one-dimensional vectors using a Flatten layer. These feature vectors serve as input for the classification step.

- **Classification with SVM**

The Support Vector Machine (SVM) was used to classify the extracted features into "smiling" or "non-smiling" categories. SVM identifies the optimal hyperplane that separates the two classes in the feature space. The RBF (Radial Basis Function) kernel was chosen to handle non-linear decision boundaries [16].

The SVM optimization process is defined as:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i \quad \text{......................(1)}$$

Subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \text{............(2)}$$

Where:
- w: Weight vector defining the hyperplane.
- b: Bias term determining the hyperplane's position.
- xi: Feature vector for the i-th sample.
- yi: True label (+1 for smiling, −1 for non-smiling).
- C: Regularization parameter balancing margin maximization and error minimization.
- ξi: Slack variables allowing for misclassified points.

The decision function for classification is:

$$f(x) = \text{sign}(w^T x + b) \quad \text{..........(3)}$$

Where the sign determines the class (+1 for smiling, −1 for non-smiling).

- **Model Training**

  The training process involved the following steps:
  1- Feature Extraction: Images were passed through ResNet50, and the resulting feature vectors were used as input to the SVM classifier.
  2- Dataset Splitting: The dataset was partitioned into training (70%), validation (15%), and test (15%) subsets using stratified sampling to ensure class balance. This technique ensured an equal distribution of images across the smile and nonsmile classes within each subset, by this means preserving the proportional representation of both categories throughout the data splitting process.
  3- SVM Training: The SVM classifier was trained utilizing the extracted feature vectors and their corresponding labels.

### D. Evaluation

The system achieved an accuracy of 94.65% on the test set. A confusion matrix was generated to analyze the true positives, true negatives, false positives, and false negatives. Other metrics, loke precision, recall, and F1-score, were computed to assess the model's performance in a comprehensive way.

The Binary Cross entropy Loss Function, initially utilized during testing with dense layers, is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(p_i) + (1-y_i)\log(1-p_i)\right) \quad \text{.......} \quad (4)$$

Where:

- yi: True Label.
- pi: Predicted Probability.
- N: Total Number of Samples.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad \text{..........................} \quad (5)$$

The Sigmoid Activation Function, which maps outputs to probabilities during dense layer testing, is defined as:

By integrating techniques of classical computer vision, deep learning for feature extraction, and SVM for classification, the proposed approach demonstrates a balance between computational efficiency and predictive accuracy. The systematic workflow ensures scalability and robustness, making the system suitable for real world facial expression recognition tasks.

## E. Results

This section presents the outcomes of the proposed system, detailing its performance during training, validation, and testing. The results are analyzed to highlight the strengths and limitations of the approach, supported by visualizations to enhance understanding.

- **Training Results**

The training process spanned 40 epochs, during which the system demonstrated strong performance:

- Final Training Accuracy: 91.2%
- Final Validation Accuracy: 94.65%
- Final Training Loss: 0.201

Final Validation Loss: 0.201.

The low and stable loss values across training and validation indicate effective convergence without overfitting. The training and validation progress is visualized below charts (figure 6 and 7).

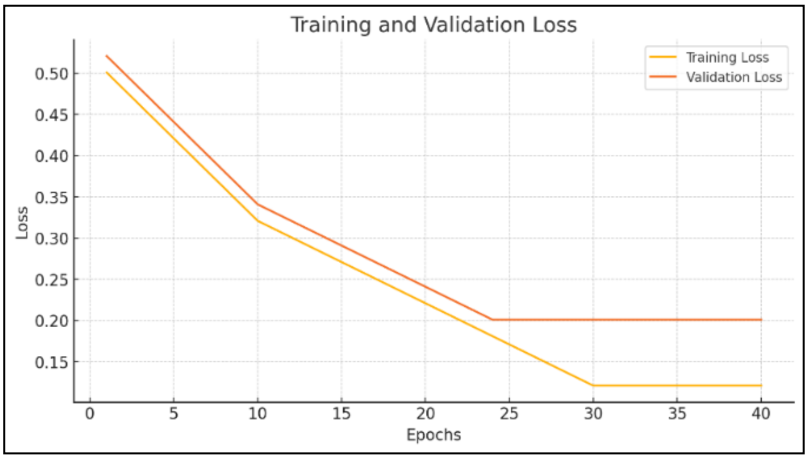


**Figure 6.** Training and Validation Accuracy

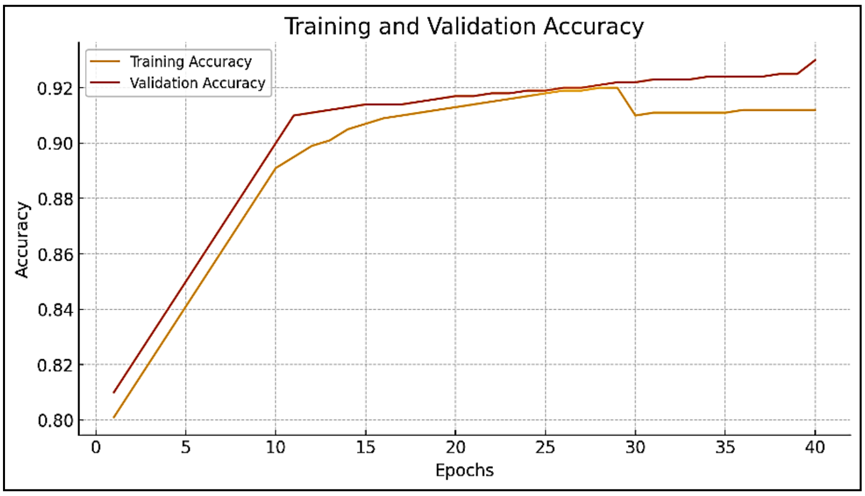


**Figure 7.** Training and Validation Loss

- **Testing Results**

The proposed system was evaluated on the test set, achieving a final accuracy of 94.65%. This high accuracy demonstrates the robustness of the system in distinguishing between smiling and non-smiling facial expressions.

- **Performance Metric**

- **F1-Score:** The F1-Score, which balances precision and recall, was also 93.26%, reflecting the model's overall performance.
- **Confusion Matrix**: To further analyze the results, a confusion matrix was generated:

**Table 1.** Confusion Matrix

| Predicted | Non-Smiling | Smiling |
|---|---|---|
| **Actual Non-Smiling** | 1864 | 74 |
| **Actual Smiling** | 74 | 1988 |

- True Positives (Smiling): 1988
- True Negatives (Non-Smiling): 1864
- False Positives: 74
- False Negative : 74

- **Visualization of Predictions**

The following image (Figure 7) showcases a subset of predictions made by the model, with the actual and predicted labels displayed for each sample. This visual representation highlights the system's effectiveness and areas where it misclassified certain instances.


**Figure 8**: Actual vs. Predicted Labels for Test Samples

## F. Discussion

The results strongly support the effectiveness of the proposed methodology. The combination of ResNet50 for feature extraction and SVM for classification proved instrumental in achieving high accuracy and balanced performance metrics. ResNet50's ability to extract high-quality, hierarchical features provided a robust foundation for classification, while the SVM with an RBF kernel ensured effective separation of smiling and non-smiling expressions in the feature space. Data preprocessing and augmentation further enhanced the system's generalization capability by simulating real-world variations, thereby minimizing overfitting. The alignment between training and validation metrics underscores the system's robustness. However, the confusion matrix and prediction visualizations reveal areas for refinement, such as handling ambiguous expressions or noisy data more effectively.

These results validate the proposed system's suitability for real-world applications, offering a reliable and efficient solution for facial expression recognition.

## G. Comparison

Deep learning methods, such as the proposed ResNet50 with SVM classification, offer substantial advantages over traditional techniques like Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gaussian features. While traditional methods are fast and effective for simple textures and shapes, they often lack the robustness needed for complex scenarios. In contrast, the proposed method utilizes deep learning to extract more nuanced and complex features, significantly enhancing performance in diverse conditions. Among classifiers, Support Vector Machines (SVM) stand out for their effectiveness in high-dimensional spaces, particularly when paired with deep learning features. SVM consistently outperforms other classifiers, including AdaBoost and Extreme Learning Machines (ELM), in terms of accuracy and robustness. While AdaBoost and ELM are competitive, they do not reach the high accuracies of SVM, especially when advanced feature extraction is involved. SoftMax classifiers, commonly used in neural network architectures, perform well but still fall short compared to SVM-equipped methods.

https://doi.org/10.33022/ijcs.v14i5.5005 

The proposed method achieves an accuracy of 94.65%, surpassing other advanced techniques like CNN-2Loss, which reaches up to 94.6±0.29%. This slight edge highlights the efficiency of combining ResNet50's deep feature extraction with SVM's classification power. This combination not only provides high accuracy but also enhances the system's ability to generalize across different scenarios, making it highly suitable for real-world applications where reliability and robust performance are critical. Overall, the integration of ResNet50 and SVM represents a significant innovation in facial expression recognition, positioning it as a top contender in terms of both technological advancement and application suitability, as evidenced by the comparisons in the table.

**Table 2**: Comparison between related works

| Method | Features Used | Classifier | Accuracy |
|---|---|---|---|
| **An et al. [22]** | HOG | ELM | 88.5% |
| **Shan et al. [23]** | Pixel Comparison | AdaBoost | 89.7% ± 0.45% |
| **Chen et al. [24]** | CNN | AdaBoost | 91.8% ± 0.95% |
| **Liu et al. [25]** | HOG | SVM | 92.29% ± 0.81% |
| **Jain et al. [26]** | Gaussian | SVM | 92.97% |
| **Kahou et al. [27]** | LBP | SVM | 93.2% ± 0.92% |
| **Zhang et al. [28]** | Multi-scale Features | AdaBoost | 89.21% |
| **Gao and Wu [29]** | CNN-Basic | SoftMax | 93.6% ± 0.47% |
| **Gao and Wu [30]** | CNN-2Loss | SoftMax | 94.6% ± 0.29% |
| **Proposed Method** | **ResNet50** | **SVM** | **94.65%** |

## H. Conclusion

The proposed system for binary facial expression prediction, leveraging ResNet50 for feature extraction and SVM for classification, achieved exceptional performance with a test accuracy of 94.65%. The combination of robust data preprocessing, hierarchical feature extraction, and efficient classification highlights the system's effectiveness in handling diverse facial expressions under varying conditions. The use of the GENKI dataset, augmented through various transformations, contributed to the system's strong generalization capabilities. Furthermore, the F1-Score metric underscores the system's balanced performance, validating its reliability and robustness for real-world applications. Despite its strengths, the study acknowledges areas for improvement, such as expanding the dataset and exploring alternative architectures to enhance robustness further. Future work could focus on real-time implementation and error analysis to address misclassifications effectively. In conclusion, the proposed approach demonstrates a scalable and efficient solution for facial expression prediction, paving the way for its integration into applications such as human-computer interaction, security, and healthcare.

## I. References

[1] H. Zhang, "Real-time face recognition method based on MTCNN-Inception-ResNet-v2-SVM model," Applied and Computational Engineering, vol. 45, pp. 179–189, 2024. doi: 10.54254/2755-2721/45/20241677.

[2] H. Echoukairi, M. El Ghmary, S. Ziani, and A. Ouacha, "Improved methods for automatic facial expression recognition," International Journal of Interactive

Mobile Technologies, vol. 17, no. 6, pp. 23–35, 2023.doi: 10.3991/ijim.v17i06.37031.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27–30, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90
[4] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016. doi: 10.1109/LSP.2016.2603342
[5] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," Neurocomputing, vol. 411, pp. 340–350, Oct. 2020, doi: 10.1016/j.neucom.2020.06.014.
[6] A. M. Abdulazeez et al., “Efficient CNN Approach for Facial Expression Recognition,” J. Phys. Conf. Ser., 2021.
[7] J. Whitehill et al., “Whose Smile Is It Anyway? Automatic Smile Detection Using the GENKI Database,” 2009.
[8] A. Mollahosseini et al., “AffectNet: A Database for Facial Expression Recognition,” 2019.
[9] X. Zhang et al., “Transformer-Based Architectures for Binary Facial Expression Prediction,” 2023.
[10] Y. Chen et al., “Vision Transformers Combined with ResNet for Enhanced Smile Detection,” 2024.
[11] A. Singh et al., “Lightweight Models for Real-Time Smile Prediction,” 2024.
[12] O. S. Kareem, A. M. Abdulazeez, and D. Q. Zeebaree, “Skin lesions classification using deep learning techniques,” Asian J. Res. Comput. Sci., vol. 9, no. 1, pp. 1–22, 2021.
[13] M. Liu et al., “Deep Locality-Preserving CNN for Facial Expression Recognition,” 2014.
[14] D. M. Ahmed, A. M. Abdulazeez, D. Q. Zeebaree, and F. Y. Ahmed, “Predicting university’s students performance based on machine learning techniques,” in 2021 IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS), pp. 276–281, IEEE, 2021.
[15] R. Patel et al., “Hybrid CNN-SVM Models for Binary Facial Expression Prediction,” 2023.
[16] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, “An ensemble DCNN-based regression model for automatic facial beauty prediction and analyzation,” 2023.
[17] D. A. Zebari, A. M. Abdulazeez, D. Q. Zeebaree, and M. S. Salih, “A fusion scheme of texture features for COVID-19 detection of CT scan images,” in 2020 Int. Conf. Adv. Sci. Eng. (ICOASE), pp. 1–6, IEEE, 2020.
[18] S. Ranganathan et al., “Data Augmentation Techniques for Facial Expression Recognition,” 2018.
[19] K. Zhao et al., “Multi-Task Learning for Facial Expression Recognition and Attribute Prediction,” 2018.
[20] A. M. Abdulazeez et al., “Deep Convolution Neural Network for Facial Expression Recognition,” IJSB Lit. Rev., 2021.

[21] B. T. Chicho, A. M. Abdulazeez, D. Q. Zeebaree, and D. A. Zebari, "Machine learning classifiers based classification for IRIS recognition," Qubahan Acad. J., vol. 1, no. 2, pp. 106–118, 2021.
[22] A. M. Abdulazeez et al., "Facial Beauty Standards Predictions Based on Machine Learning: A Comparative Analysis," Indones. J. Comput. Sci., 2024.
[23] L. An, S. Yang, and B. Bhanu, "Efficient smile detection by extreme learning machine," Neurocomputing, vol. 149, pp. 354–363, 2015.
[24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Int. Conf. Mach. Learn., pp. 448–456, 2015.
[25] D. Cui, G. B. Huang, and T. Liu, "Smile detection using Pair-wise Distance Vector and Extreme Learning Machine," in Neural Networks (IJCNN), Int. Jt. Conf. Neural Netw., pp. 2298–2305, IEEE, 2016.
[26] M. Liu, S. Li, S. Shan, and X. Chen, "Enhancing expression recognition in the wild with unlabeled reference data," in Comput. Vis.–ACCV 2012, pp. 102–115, 2013.
[27] V. Jain and J. Crowley, "Smile detection using multi-scale Gaussian derivatives," in 12th WSEAS Int. Conf. Signal Process. Robot. Autom., pp. 45–52, 2013.
[28] S. E. Kahou, P. Froumenty, and C. Pal, "Facial expression analysis based on high-dimensional binary features," in Comput. Vis.-ECCV 2014 Workshops, pp. 649–659, 2014.
[29] K. Zhang, Y. Huang, H. Wu, and L. Wang, "Facial smile detection based on deep learning features," in 3rd IAPR Asian Conf. Pattern Recognit. (ACPR), Kuala Lumpur, Malaysia, pp. 534–538, 2015, doi: 10.1109/ACPR.2015.7486560.
[30] Y. Gao, H. Liu, P. Wu, and C. Wang, "A new descriptor of gradients self-similarity for smile detection in unconstrained scenarios," Neurocomputing, vol. 174, pp. 1077–1086, 2016.