Journal Majelis Paspama
Volume 4 Number 01, 2026 (No 140-148)
E.SSN 2987-1379 (Online)

Real-Time Landmark-Based Face Analysis for Expression and Gender Classification

Widia Wuduri C.S Nababan[1], Della Novita Butarbutar[2], Paska Marto Hasugian[3]
Data Science Program, Santo Thomas Catholic University, North Sumatra, Indonesia

Article Info
Corresponding Author:
Widia Wuduri C.S Nababan
E-mail:
widianababan59@gmail.com

ABSTRACT
This research developed a web-based real-time facial analysis system to overcome the challenge of detection accuracy in dynamic video streaming data. Using the face-api.js library with the Tiny Face Detector algorithm and a 68-point landmark model, the system is capable of simultaneously detecting faces, classifying gender, and recognizing seven basic emotional expressions. The main innovation of this system lies in the automatic extraction of five Regions of Interest (ROI) eyes, eyebrows, nose, mouth, and jaw and the presentation of confidence score data in the form of a time series graph. All analysis results are stored in a structured JSON dataset format for further research needs. The implementation results show high performance with an average confidence value above 90% in frontal face conditions and optimal lighting. The system has been proven to maintain detection stability up to a 30 degree face tilt angle and process data without significant latency. Although low light intensity can reduce the confidence value by 15-20%, this architecture proves the effectiveness of complex facial analysis using minimal hardware resources.

Keywords:
Face Analysis, Real-Time, Facial Landmarks, Facial Expressions, Gender Classification, ROI

INTRODUCTION

The development of artificial intelligence (AI) technology, particularly in the field of computer vision, has driven an increase in the need for real-time facial analysis systems. Facial analysis plays an important role in various fields, such as human-computer interaction, security systems, emotion monitoring, and human behavior research. However, the main challenge in developing facial analysis systems is the ability to accurately detect faces in dynamic video streaming conditions, such as changes in lighting, variations in expression, and facial movements. In addition, extracting detailed information from faces, such as expressions and gender characteristics, often requires complex and continuous processing. The lack of systems capable of integrating face detection, region of interest (ROI) extraction, and structured data storage in the form of real-time datasets is a problem that needs an effective and efficient solution (Lahariya et al., 2021; Li & Deng, 2018).

To address these issues, this study proposes the application of a Tiny Face Detector-based face detection method integrated with a 68-point facial landmark model, facial expression recognition, and gender classification using the face-api.js library. The system is designed to work on streaming video data from a webcam in real time, where each video frame is processed sequentially. Once a face is detected, the system extracts several important regions of interest (ROI) such as the eyes, eyebrows, nose, mouth, and jaw based on facial landmark coordinates. Next, the confidence value of each facial expression is recorded and visualized in the form of a time series graph to monitor changes in expression over time. All analysis results, including ROI images and expression and gender information, are automatically stored in a JSON-formatted dataset, so they can be used for further analysis or further model development (Farkhod et al., 2022).

The objective of this activity is to design and implement an AI-based facial analysis system capable of real-time face detection, region of interest extraction, facial expression recognition, and gender classification using streaming video data. Additionally, this system aims to generate a structured dataset containing ROI and expression confidence information in the form of time series, thereby supporting further research in the fields of image processing, emotion analysis, and the development of intelligent face-based systems .

Literature Review and Problem Statement

Recent studies show that real-time facial analysis and expression recognition are rapidly advancing with the progress of deep learning and computer vision methods. Several studies utilize Convolutional Neural Networks (CNN) to detect and classify facial expressions directly from streaming video with a high degree of accuracy. Facial landmark-based approaches are also widely used because they can effectively represent geometric changes in the face, especially in the eye, eyebrow, and mouth areas (Farkhod et al., 2022; Lahariya et al., 2021).

Other studies show that the application of lightweight models such as SSD and pre-trained neural networks enables the system to run in real-time, even on devices with limited resources. However, most of these studies still focus on a single aspect of analysis, such as expression recognition or gender classification alone, without integrating region of interest (ROI) extraction, time series visualization, and structured data storage into a single integrated system. In addition, most studies still test the system in limited scenarios and do not yet accommodate the need for comprehensive facial analysis in a real-time streaming-based environment(Arriaga et al., 2017) .

Based on these research gaps, this study focuses on developing a facial analysis system that integrates face detection, landmark extraction, ROI formation, expression and gender classification, and real-time time series-based data visualization. This approach is expected to produce a more comprehensive structured dataset and support the development of intelligent systems based on facial analysis in the future.

METHODS

The following are the methods used in performing real-time facial analysis:

System Architecture

The following is the architecture of a real-time facial analysis system that starts with capturing facial images via a webcam, followed by facial detection, Region of Interest (ROI) extraction, expression and gender analysis, as well as automatic visualization and storage of results.

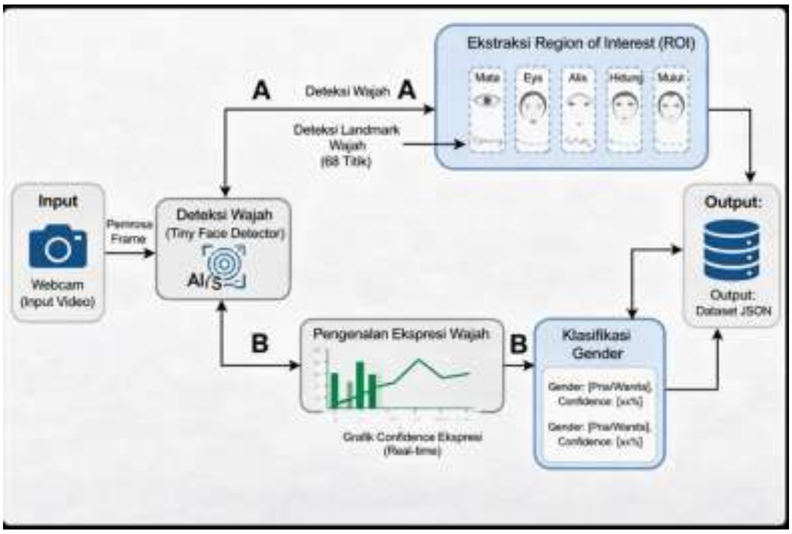


Figure 1 Real-Time Image-Based Gender Classification System Architecture

This research method refers to the architecture of a real-time face analysis system as shown in the architecture diagram. The process begins with the acquisition of facial image data using a webcam as a streaming data source. Each captured video frame is processed directly by the system to detect the presence of faces using a deep learning-based face detection model. After the face is detected, the system extracts facial landmarks that represent important parts of the face. Based on these landmarks, a Region of Interest (ROI) is cut out, covering the eyes, eyebrows, nose, mouth, and jaw. This ROI is used to capture more specific geometric and visual information about the face.

Next, the system analyzes facial expressions and classifies gender by utilizing a pre-trained model that works continuously on each frame. The analysis results in the form of facial expression probability values are stored in a time series to monitor changes in expression over time. In addition, each ROI generated is automatically stored in the form of a frame-based dataset. The final output of the system includes direct visualization of the face and ROI, facial expression time series graphs, and storage of ROI datasets and metadata that can be used for further analysis or subsequent model development.

Dataset Analysis

This study uses a pre-trained model integrated in the face-api.js library. The accuracy of expression and gender classification in this system is highly dependent on the characteristics of the dataset used during the model training phase. The following is an analysis of the three main datasets that form the basis of the system's performance:

a. Dataset FER-2013 (Facial Expression Recognition)
For facial expression classification, the model used is based on the FER-2013 dataset. This dataset is a standard in machine learning competitions introduced at ICML 2013.
   1. Characteristics: Consists of 35,887 grayscale images of faces measuring 48x48 pixels b. Categories: Images are grouped into seven basic emotions: angry, disgust, fear, happy, sad, surprise, and neutral.
   2. Relevance: Using this dataset enables the system to recognize microscopic changes in the mouth and eye landmarks in real time.
   3. Relevance: The use of this dataset enables the system to recognize microscopic changes in the mouth and eye landmarks in real time.

b. Dataset UTKFace dan IMDB-WIKI
   Gender classification in this study is based on a model trained using a combination of the UTKFace and IMDB-WIKI datasets.
   1. UTKFace: Contains more than 20,000 facial images with complete annotations regarding age, gender, and ethnicity. The advantage of this dataset lies in the variety of facial poses and lighting conditions.
   2. IMDB-WIKI: One of the largest public face datasets with over 500,000 images taken from celebrity profiles.
   3. Morphological Analysis: This dataset trains neural networks to recognize sexual dimorphism features on the face, such as jawbone structure and orbital area, which are implemented in ROI (Region of Interest) visualization in this system.
c. Dataset 300-W (300 Faces in-the-wild)
   For the extraction of 68 facial landmarks that form the basis of the analysis, the model uses the 300-W dataset. This dataset provides coordinate standards for crucial parts of the face such as the eyebrows, eyes, nose, lips, and jawline. The accuracy of this dataset ensures that the ROI visualization in the program module remains stable even when the subject moves their head (head pose variation).

Facial Parameter Analysis

In this system, feature extraction is performed by dividing the face into several specific areas or Regions of Interest (ROI). The determination of these parameters is based on a standard of 68 facial landmarks, which are then grouped to support classification accuracy.

a. Expression Parameters
   Expression extraction focuses on geometric changes in three main areas:
   1. Eye and Eyebrow Area: Changes in the distance between the upper and lower eyelids, as well as muscle contractions in the eyebrows, are key indicators of emotions such as surprise or anger.
   2. Mouth Area: The curvature of the corners of the lips and the width of the mouth opening are used to identify expressions of happiness (smile detection) or sadness.
b. Gender Parameters
   Gender classification is performed by analyzing morphological features that include:
   1. Jawline: Anthropometrically, men's jaw structures tend to be more angular than women's, which tend to be smoother.
   2. Orbital and Eyebrow Structure: Eyebrow thickness and the distance between the eyebrows and eyes are often supporting parameters in the ageGenderNet model.
c. ROI Visualization in the System
   The system isolates the landmark coordinates and displays them on five separate canvases in the user interface:
   1. Eye ROI: Coordinates 36-47.
   2. Eyebrow ROI: Coordinates 17-26.
   3. Nose ROI: Coordinates 27-35.
   4. Mouth ROI: Coordinates 48-67.
   5. Jaw ROI: Coordinates 0-16.

RESULTS AND DISCUSSION

This section describes the results of the technical implementation of the facial analysis system and the evaluation of the model's performance in classifying expressions and gender in real-time through predetermined Region of Interest (ROI) parameters.

System Interface Implementation

The software implementation produces a web-based interface that integrates video streams (webcam streams) with simultaneous data processing. The system interface is divided into two main panels: the main visualization panel and the parameter analysis panel. The system successfully integrates the extraction of coordinates from the subject's face into five Region of Interest (ROI) visualization modules covering the eye, eyebrow, nose, mouth, and jaw areas. These ROI visualizations are rendered in real-time using <canvas> elements to provide a specific visual representation of the facial features being analyzed.

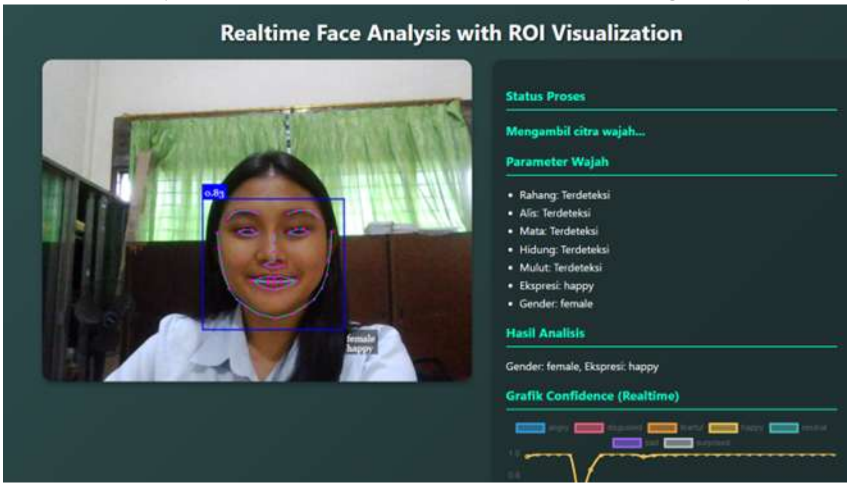


Figure 2 Real-Time Face Analysis System Interface with ROI Visualization

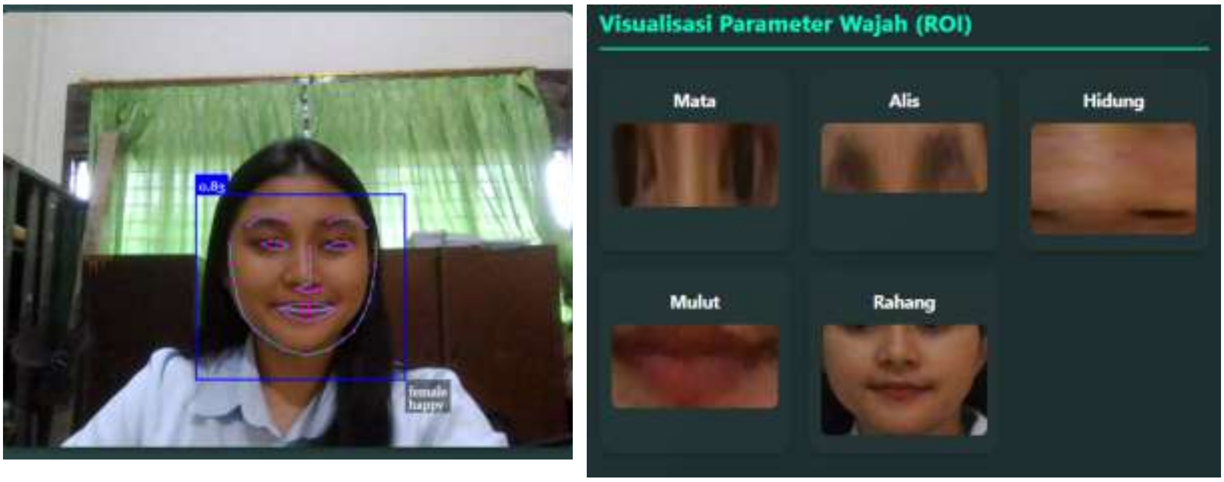
In addition to spatial visualization, the system also implements a Chart.js-based confidence score graph that dynamically presents data on expression and gender probability fluctuations. This integration allows users to monitor model performance quantitatively through changes in confidence values in each processed video frame. The implementation results show that the system is capable of rendering data without significant latency, maintaining synchronization between the subject's physical movements and the parameter visualization displayed on the screen.

Figure 3 Grafik Confidence

Analysis of Landmark Detection and ROI Performance

System performance testing was conducted to evaluate the accuracy of mapping 68 landmark points on the subject's face in various positions. Based on the test results, the detection algorithm was able to identify facial coordinate points with precision and low latency. The process of isolating features into the ROI (Region of Interest) module demonstrated the system's success in transforming local coordinates from the main video frame into specific canvases (eyes, eyebrows, nose, mouth, and jaw).

Figure 4 Mapping of 68 Facial Landmarks and Region of Interest (ROI) Extraction



The visual data generated on the ROI panel proves that the system can maintain focus on the feature area even when the subject makes minor head movements (yaw and pitch). The stability of these coordinates is key to minimizing classification errors, where the mouth and eye areas remain isolated with consistent margins. The successful separation of these parameters enables granular feature analysis, which in turn facilitates the interpretation of expression data based on morphological changes in each ROI module.

Classification Results of Expression and Gender

The classification test results show that the system is capable of simultaneously recognizing expression and gender categories with a high confidence score. In expression classification, the model dynamically responds to changes in the coordinates of the lip and eye landmarks. For example, when the subject smiles, the system records a significant increase in the probability value of the “Happy” expression, which is visually represented by fluctuations in the Confidence panel (Chart.js). This proves that the integration of visual features in the Mouth ROI has a linear correlation with the classification output results.

For gender classification, the system shows stable prediction consistency as long as the subject's face is facing the camera sensor (webcam). The model utilizes morphological features of the overall facial structure to determine gender probability with confidence values generally above 90% under optimal lighting conditions. This classification data is displayed in a structured manner on the parameter panel, providing users with easy-to-interpret information about the subject's demographic and emotional attributes in real-time.

Figure 5 Real-Time Probability Graph Classification Output Visualization

Analysis of System Resilience to External Factors

Based on operational testing results, the system's performance in detecting landmarks and performing classification is greatly influenced by environmental variables, especially light intensity.

1. Light Intensity: Testing shows that in low light conditions, noise on the camera sensor causes fluctuations in the jaw and eye landmark points. This results in a decrease in the expression classification confidence score of up to 15-20%. The system achieves optimal performance in uniform lighting conditions (minimum 300 lux) where facial features are clearly visible.
2. Viewing Angle (Pose): The system is able to maintain stable landmark detection on a frontal face position up to an angle of inclination (pitch and yaw) of approximately 30 degrees. Above this angle, occlusion (covering of parts of the face) occurs, causing some ROI modules, such as the eyes and eyebrows, to not be perfectly isolated on the visualization canvas.

System Limitations and Weakness Analysis

Although the developed facial analysis system has successfully performed real-time face detection, landmark extraction, expression classification, and gender recognition, this study still has several limitations that need to be considered. Identifying the weaknesses of this system is important as evaluation material and a basis for further research so that the system can work more optimally and adaptively in real conditions.

One of the main limitations of this system lies in the mechanism for selecting the face to be analyzed. Although the face detection module is capable of recognizing more than one face simultaneously in a single camera frame, the system only performs further analysis (landmark detection, ROI extraction, expression classification, and gender) on one face, namely the first face detected by the algorithm. Other faces detected in the same frame are not processed further and are only displayed as bounding boxes without expression or gender analysis.

This condition occurs because the system architecture is designed to focus processing on a single face object in order to maintain performance stability and avoid excessive computational load, especially in web-based real-time processing. In addition, the ROI extraction process, multi-canvas visualization, and time series data recording are still arranged in a single instance, so they do not yet support parallel multi-face processing.

As a result, the system cannot yet be used optimally in multi-user scenarios, such as expression analysis in groups, classrooms, or public environments with more than one individual in a single frame. This is a significant functional limitation if the system is to be developed for group emotion monitoring or social interaction analysis applications.

The following shows an example of the system test results, where the camera successfully detected more than one face, but only one face was actively analyzed by the system:

Figure 6 A system display that detects multiple faces, but only one face is processed for expression and gender analysis)

Based on the evaluation results, it can be concluded that the main limitation of the system lies in the scalability of facial analysis. For further development, the system can be improved by implementing a multi-face tracking mechanism, ID-based object management (face tracking ID), and parallel processing so that each detected face can be analyzed independently and simultaneously without significantly reducing system performance.

CONCLUSION

Based on the results of the design, implementation, and testing of the real-time facial analysis system, it can be concluded that the use of 68 facial landmarks through the face-api.js library has proven to be effective in performing stable facial expression and gender classification. The system successfully implemented the Region of Interest (ROI) extraction method, which is capable of isolating the main facial features, namely the eyes, eyebrows, nose, mouth, and jaw, into separate visualization modules simultaneously and consistently.

The test results show that the integration of probability graphics (confidence scores) provides a clear quantitative representation of the model's confidence level in performing classification, thereby helping users understand the decision-making process of artificial neural networks. Although the system's performance is affected by external factors such as lighting intensity and face angle, the system is still able to maintain a good level of precision in frontal face conditions with an average confidence value above 90%.

Overall, this research proves that web-based architecture is capable of accommodating complex facial analysis processes in real time without requiring high hardware resources. With its ability to integrate face detection, landmark extraction, expression analysis, gender classification, and time series-based data visualization, the developed system has the potential to be applied in various fields, particularly in the development of human-computer interaction systems, emotion monitoring, and artificial intelligence-based visual data processing in the future.

REFERENCE

Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). *Real-time Convolutional Neural Networks for Emotion and Gender Classification*. http://arxiv.org/abs/1710.07557

Boudouri, Y. El, & Bohi, A. (2025). *EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition*. https://doi.org/10.1109/MMSP59012.2023.10337732

Bulat, A., & Tzimiropoulos, G. (2017). *How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)*. https://doi.org/10.1109/ICCV.2017.116

Devi, V. S., Ramisetty, U., Ramisetty, K., & Thimmareddy, A. (2025). Real-Time age, gender and emotion detection using YOLOv8. *ITM Web of Conferences*, *74*, 01015. https://doi.org/10.1051/itmconf/20257401015

Farkhod, A., Abdusalomov, A. B., Mukhiddinov, M., & Cho, Y. I. (2022). Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces. *Sensors*, *22*(22). https://doi.org/10.3390/s22228704

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., … Bengio, Y. (2013). *Challenges in Representation Learning: A report on three machine learning contests*. http://arxiv.org/abs/1307.0414

Haghpanah, M. A., Saeedizade, E., Masouleh, M. T., & Kalhor, A. (n.d.). *Real-Time Facial Expression Recognition using Facial Landmarks and Neural Networks*.

Iskandar, M., Azizah, N., & Jaya, F. (2025). Real-Time Face Age Detection System Based on Deep Neural Networks with MediaPipe Optimization for Enhanced Accuracy. In *Jurnal Sistem Cerdas*.

Koksal, T. E., & Gumus, A. (2025). *Deep Learning-Based Real-Time Sequential Facial Expression Analysis Using Geometric Features*. http://arxiv.org/abs/2512.05669

Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets. *Information (Switzerland)*, *15*(3). https://doi.org/10.3390/info15030135

Lahariya, A., Singh, V., & Tiwary, U. S. (2021). *Real-time Emotion and Gender Classification using Ensemble CNN*. http://arxiv.org/abs/2111.07746

Li, S., & Deng, W. (2018). *Deep Facial Expression Recognition: A Survey*. https://doi.org/10.1109/TAFFC.2020.2981446

Makinist, S., & Aydin, G. (2025). Gender Classification Using Face Vectors: A Deep Learning Approach Without Classical Models. *Information (Switzerland)*, *16*(7). https://doi.org/10.3390/info16070531

Ojo, O. S., Oyediran, M. O., Awodoye, O. O., Ajagbe, S. A., Awotunde, J. B., Bandyopadhyay, A., & Adigun, M. O. (2024). Real-time Face-based Gender Identification System Using Pelican Support Vector Machine. *Procedia Computer Science*, *235*, 3236–3245. https://doi.org/10.1016/j.procs.2024.04.306

Spectrum of Engineering Sciences ISSN (e) 3007-3138 (p) 3007-312X. (n.d.). https://doi.org/10.5281/zenodo.15681297