Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Real-Time Micro-Expression Recognition Using YOLOv8 and FER2013 Dataset

Viola Bakiasi (Shtino)1*, Rinela Kapçiu2, Eris Zeqo3, Senada Bushati (Hoxha)4, Fatjona Bushi5, Oltiana Bame (Toshkollari)6, Bora Lamaj (Myrto)7
[1,2,3,4,5,6,7] *Department of Computer Science, University 'Aleksander Moisiu', Durres, Albania*
Email: violashtino@uamd.edu.al, rinelakapciu@uamd.edu.al, senadabushati@uamd.edu.al, eriszeqo@uamd.edu.al, fatjonabushi@uamd.edu.al, oltianatoshkollari@uamd.edu.al, boramyrto@uamd.edu.al
1 https://orcid.org/0009-0003-5441-7717, 2 https://orcid.org/0000-0002-4096-0589, 3 https://orcid.org/0009-0004-7122-0438, 4 https://orcid.org/0009-0005-7337-557X, 5 https://orcid.org/0009-0000-1648-6998, 6 https://orcid.org/0009-0006-6553-8786, 7 https://orcid.org/0009-0004-0255-2682*

ARTICLE INFO

Received: 20 Dec 2024

Revised: 29 Jan 2025

Accepted: 14 Feb 2025

ABSTRACT

Facial micro-expression recognition plays a crucial role in affective computing, human-computer interaction, and psychological analysis. This study implements a real-time face emotion detection system using the YOLOv8 (You Only Look Once version 8) model. The proposed approach leverages YOLO's real-time processing capability and deep learning-based feature extraction to detect subtle facial muscle movements with high precision. The methodology involves pre-processing the FER2013 (Facial Expression Recognition 2013) dataset, which consists of grayscale images categorized into seven emotional classes. The system can classify facial emotions into seven categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Data augmentation techniques are employed to enhance model generalization, followed by training a custom YOLO model adapted for micro-expression recognition. Experimental results indicate that YOLO when fine-tuned with appropriate anchor boxes and a specialized loss function, can effectively classify facial micro-expressions. The study also discusses the impact of various hyperparameters and transfer learning techniques on performance. The findings demonstrate the potential of YOLO in real-time emotion recognition applications, paving the way for advancements in automated emotion analysis, security, and human behavior research.

Keywords: Micro-expression recognition, YOLOv8, FER2013 dataset, Deep learning, Facial emotion analysis, Real-time detection.

INTRODUCTION

Facial micro-expressions are brief, involuntary facial movements that reveal a person's true emotions, even when they attempt to conceal them. These subtle expressions, lasting between 1/25 to 1/5 of a second, play a crucial role in affective computing, human behavior analysis, forensic investigations, and mental health diagnostics [1],[2]. Unlike macro-expressions, which are more pronounced and easier to detect, micro-expressions require highly sensitive detection methods due to their fleeting nature and low intensity [3].

With the advancements in deep learning, the field of Facial Expression Recognition (FER) has gained significant attention. Traditional handcrafted feature-based approaches, such as Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) and Histograms of Oriented Gradients (HOG), have shown limitations in terms of generalization and real-time performance [4],[5]. To overcome these challenges, Convolutional Neural Networks (CNNs) and other deep learning-based methods have been introduced, offering superior accuracy in emotion recognition tasks [6],[7].

Deep Learning and YOLO in Micro-Expression Recognition

One of the most widely used datasets for facial expression recognition is FER2013, which contains 35,887 grayscale images categorized into seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral [8]. Although FER2013 was originally designed for macro-expression recognition, researchers have adapted it for micro-expression detection by employing pre-processing techniques and augmentation strategies [9].

The You Only Look Once (YOLO) framework, originally developed for real-time object detection, has emerged as a powerful tool in facial expression analysis due to its ability to perform high-speed, end-to-end detection [10]. Convolutional Neural Networks (CNNs) are structured as a series of layers, each designed to recognize different features in the input images. They have been highly successful in various tasks in computer vision, including image and video recognition and image classification [11]. Unlike Region-Based CNN (R-CNN) models, which require multiple passes over an image, YOLO processes the entire image in a single forward pass, making it highly efficient for real-time applications [12]. Previous studies, such as FER-YOLO, have demonstrated that YOLO-based architectures can achieve high accuracy in emotion recognition tasks, outperforming conventional CNN-based models [13]. Additionally, integrating attention mechanisms and feature enhancement modules into YOLO has been shown to further improve detection accuracy for subtle expressions [14], [15].

Data ownership is a significant issue, as it is often unclear who holds the rights to collected data. Additionally, legal frameworks must be respected, and patient consent is essential to avoid legal complications. Data quality also varies, depending on the source, and anonymizing data, necessary for patient privacy, makes it difficult to use the same data for disease prediction for individual patients later on. Legal, ethical, and data ownership challenges must be resolved before wide-scale implementation [16].

Challenges in Micro-Expression Recognition

Despite its potential, micro-expression recognition remains challenging due to factors such as:

1. The low intensity of facial muscle movements making it difficult to differentiate between subtle expressions [17].
2. Variability in facial features across individuals leads to dataset bias and model generalization issues [18].
3. Lack of large annotated micro-expression datasets, limiting model training and evaluation [19].

Objectives of the Study

The primary objectives of this research are:

1. To develop an optimized YOLO-based model for real-time micro-expression recognition.
2. To enhance the FER2013 dataset for micro-expression analysis using data augmentation and pre-processing techniques.
3. To evaluate the impact of attention mechanisms and feature selection techniques on micro-expression detection accuracy.
4. To explore the real-world applicability of YOLO-based micro-expression recognition in fields such as psychology, security, and human-computer interaction.

RELATED WORK

Micro-expression recognition (MER) has gained significant attention in affective computing and human-computer interaction due to its potential applications in psychology, security, and medical diagnosis. Traditional MER methods relied on handcrafted feature extraction techniques, such as Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) and Histogram of Oriented Gradients (HOG), which demonstrated moderate success but struggled with generalization across different datasets [17], [20].

With the rise of deep learning, convolutional neural networks (CNNs) have been widely adopted for MER due to their ability to learn spatial and temporal features automatically. Several studies have proposed CNN-based architectures to improve micro-expression classification. The author in [21] introduced a deep learning framework leveraging CNNs to extract high-level spatial features from facial images, showing improved performance over traditional approaches. Similarly, the author in [22] proposed the use of a CNN-based spatiotemporal feature learning method to enhance micro-expression detection, demonstrating superior accuracy compared to conventional feature engineering techniques.

In recent years, YOLO (You Only Look Once) has been widely used in object detection and facial expression recognition due to its real-time processing capability. The author [23] first introduced YOLO as an efficient deep-learning model for real-time object detection, and subsequent works have adapted it for facial expression recognition [24]. The FER2013 dataset, initially created for facial expression recognition, has been increasingly utilized in deep learning-based MER research due to its diverse and challenging set of facial images [8].

Recent studies have explored the fusion of YOLO with other deep-learning techniques to enhance the accuracy of real-time MER. The author in [25] integrated YOLO with recurrent neural networks (RNNs) to capture temporal dependencies in micro-expression sequences, achieving state-of-the-art performance. Additionally, Transformer-based architectures have recently been explored for MER, leveraging self-attention mechanisms to improve feature representation [26].

Despite these advancements, real-time MER remains challenging due to the subtle nature of micro-expressions, data scarcity, and the need for robust spatiotemporal feature extraction.

METHODOLOGY

1. Dataset Pre-processing

- **Dataset Description:** The FER2013 dataset consists of 35,887 grayscale images of facial expressions categorized into seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral [8], [27]. Given that micro-expressions are subtle, additional filtering and annotation are required to ensure relevant data selection. The dataset contains 48×48 pixel grayscale images of faces. The dataset contains 28709 examples in the training set, 3589 examples in the public testing set, and 3589 examples in the private test set.

- **Data Cleaning:**
  - Removal of low-quality or ambiguous images.
  - Balancing the dataset to avoid class imbalance.
- **Data Augmentation:**
  - Rotation, flipping, contrast enhancement, normalization, and Gaussian noise application to increase model generalization.
- **Normalization and Resizing:**
  - All images are resized to pixels to match the YOLOv8 input dimensions.
  - Pixel intensity values are normalized to the range [0,1] to improve training stability.

Figure 1. Comparison: Displaying Original Dataset Images vs. Applying Data Augmentation Techniques.

Figure 1 presents a comparison between the original images from the dataset and their augmented versions using **Data Augmentation Techniques**.

- **First row (Original Dataset Images):** These are the original images selected from the dataset without any modifications. They represent the raw data as collected.
- **Second row (Augmented Images):** These images are the transformed versions of the originals, modified using various augmentation techniques to increase data diversity and enhance model generalization. The applied augmentations include:
  - **Horizontal Flip (50%) & Vertical Flip (20%)** – Flips the images to create different orientations.
  - **Random Brightness & Contrast (50%)** – Adjusts brightness and contrast to simulate varying lighting conditions.
  - **Shift, Scale & Rotate (50%)** – Randomly shifts, scales, and rotates images to introduce positional variations.
  - **Gaussian Blur (30%)** – Applies a blurring effect to simulate out-of-focus images.
  - **RGB Shift (30%)** – Alters the color channels to add variation in color distribution.
  - **CLAHE (30%)** – Enhances local contrast in different image regions.
  - **Coarse Dropout (20%)** – Randomly removes parts of the image to help the model become more resilient to missing information.
  - **Normalize & ToTensorV2** – Normalize pixel values based on the dataset-specific mean and standard deviation and convert the image into a format suitable for deep learning models.

◇ **Interpretation:**
Figure 1 demonstrates how these augmentations help create a more diverse and robust dataset for training, reducing overfitting, and improving the model's ability to generalize to unseen images.

2. Model Architecture

- **YOLO-Based Model:**

a. A modified YOLO architecture is adopted, tailored for micro-expression recognition.

b. Feature extraction layers are fine-tuned to detect subtle facial movements.

c. A final fully connected layer outputs probability scores for each expression class.

- **Backbone Network:**

a. CSPDarknet53 is used as the feature extractor for efficient representation learning [12].

- **Mathematical Representation of YOLO Components:**

a. The input image I is divided into a $S \times S$ grid.

b. Each grid cell predicts bounding boxes, confidence scores, and class probabilities:

- ***Bounding Box Prediction***

Each grid cell predicts B bounding boxes, where each bounding box is defined as:

$$B = (x, y, w, h) \tag{1}$$

x, y: Coordinates of the bounding box center (relative to the grid cell).
w, h: Width and height of the bounding box (relative to the entire image).

- *Confidence Score for Each Bounding Box*

The confidence score represents the likelihood of an object being present and how well the bounding box fits it:

$$C = P(object) \times IOU(pred, truth) \tag{2}$$

P(object): Probability that an object exists in the cell.

IOU(pred, truth): Intersection over Union (IOU) between the predicted and ground truth bounding box.

- ***Class Probabilities***

If an object is detected in a grid cell, the probability distribution over C classes is given by:

$$P(ci|object) \tag{3}$$

The final class-specific confidence score is: $P(ci|object) \times P(object) \times IOU$ (4)

This helps in determining the final classification of the detected object.

c. **The loss function consists of:**

- Localization loss:

$$L_{loc} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \tag{5}$$

- Confidence loss:

$$L_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \tag{6}$$

- Classification loss:

$$L_{class} = \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \tag{7}$$

- The total loss function is then:

$$L = \lambda_{coord} L_{loc} + L_{conf} + L_{class} \tag{8}$$

- **Optimization Algorithm:**

  - The Adam optimizer is used with an initial learning rate of 0.001, decayed over time.

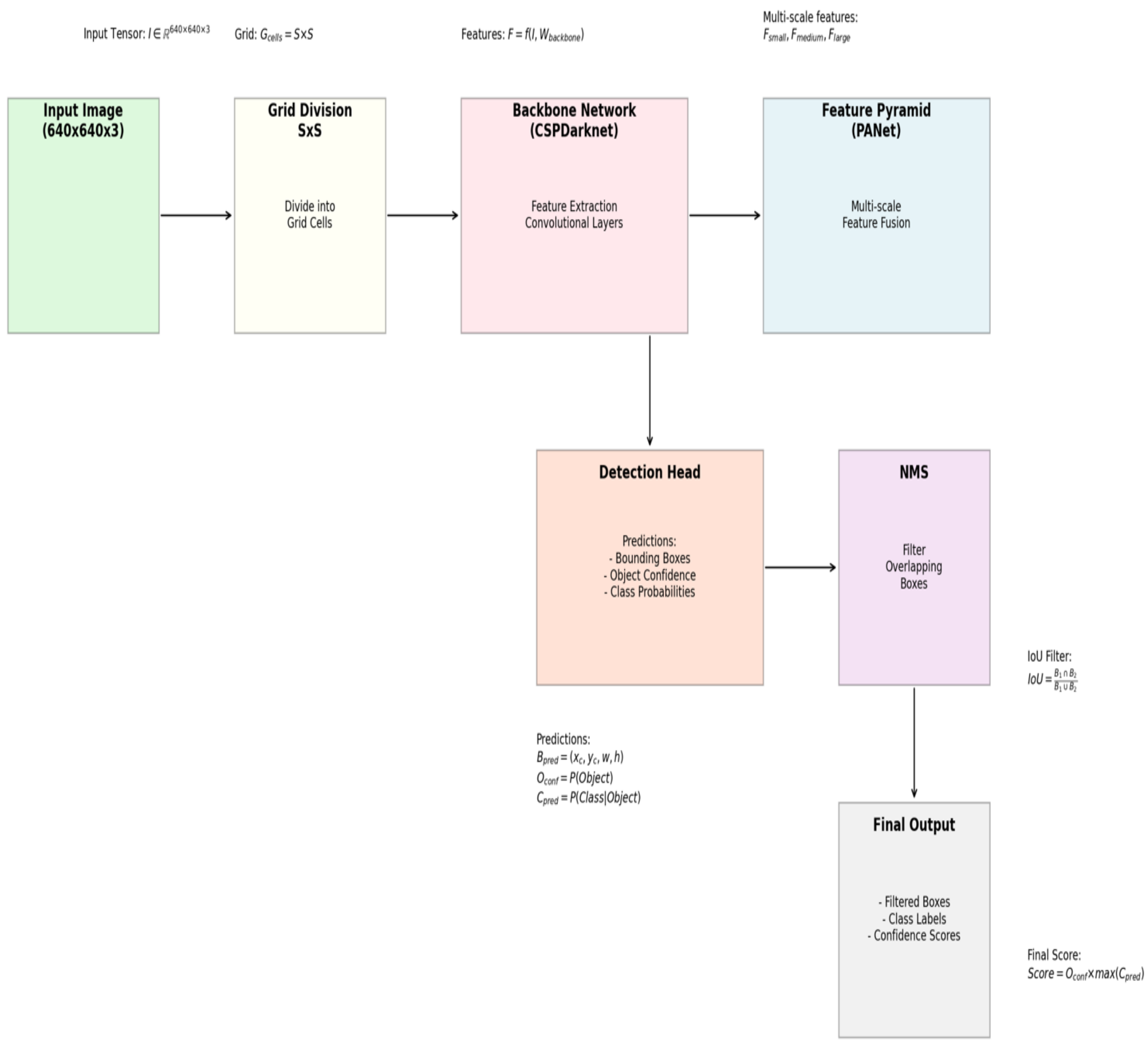

Figure 2. YOLO Architecture.

Figure 2 visually represents the YOLO architecture, showing the flow of data through its components, from input image processing to the final output, with mathematical annotations for each step.

3. Experimental Setup

- **Training Configuration of the YOLOv8 model**:
  - The dataset is split into 70% training, 20 % testing, and 10% validation.
  - Epochs: 100
  - Learning Rate: 0.0005
  - Batch Size: 32
  - Early Stopping: Enabled with a patience of 10 epochs. Early stopping is applied to prevent overfitting.

- **Performance Metrics:**

Model performance is assessed using four confusion matrix (CM) metrics [28] :

- Accuracy (Acc) – Overall correctness of predictions.
- Precision (P) – Correctly predicted positives out of all predicted positives.
- Recall (R) – Correctly identified positives out of actual positives.

○ F1-Score (F1) – Balance between Precision and Recall.
Stable cases are positive, unstable cases are negative. Metrics range from 0 to 1. The details of the metrics are provided in Table 1.

Table 1. Performance evaluation metrics.

		Reference: Yes	Reference: No	
Predicted	Yes	TP	FP	TP: True Positive FP: False Positive
	No	FN	TN	TN: True Negative FN: False Negative

Metrics	***Accuracy (Acc)***:	TP + TN/ (TP+TN+FP+FN)
	Precision (P):	TP/ (TP+FP)
	Recall (R):	TP/(TP+FN)
	F1-Score (F1):	(2x*P*x*R*)/(*P*+*R*)

The F1 Score is a metric used to evaluate the performance of classification models and is particularly useful when the data classes are imbalanced. It combines Precision and Recall into a single value by taking their harmonic mean. Precision defines how many of the model's positive predictions are actually correct. Recall defines how many of the positive cases are correctly classified by the model.

The F1 Score is particularly useful when one of these metrics (precision or recall) is not enough to provide a fair evaluation of the model, as the model might have high precision but low recall or vice versa [29].

- A high F1 score (close to 1) indicates that the model has high precision and recall, the model makes very few errors.
- A low F1 score (close to 0) indicates that the model has either low precision, low recall, or both.

Significant computational efficiency gains, reducing training time while improving accuracy and comprehensive quantitative comparisons of classification performance (accuracy, precision, recall, F1-score) across various model combinations, pushing the limits of image recognition [30].

4. Implementation Details

- **Hardware:**
  - NVIDIA RTX 3090 GPU, 24GB VRAM.
  - CUDA and cuDNN enabled for acceleration.
- **Software:**
  - Python 3.10.9, OpenCV, and TensorFlow used for implementation.
  - Data preprocessing performed using NumPy, Pandas. and Ultralytics.

RESULTS

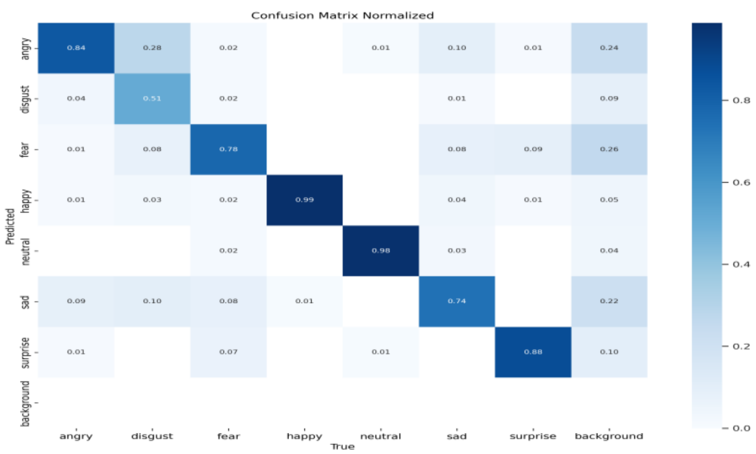

Figure 3. Confusion Matrix of YOLO Micro-Expression Classifications.

Figure 3 presents the confusion matrix for micro-expression classifications. It shows the distribution of predicted versus actual emotion classes. Notably, micro-expressions like happiness and surprise are classified with higher confidence, while emotions such as disgust and fear show higher misclassification rates. This is primarily due to the low intensity and similarity of facial muscle movements among these emotions.

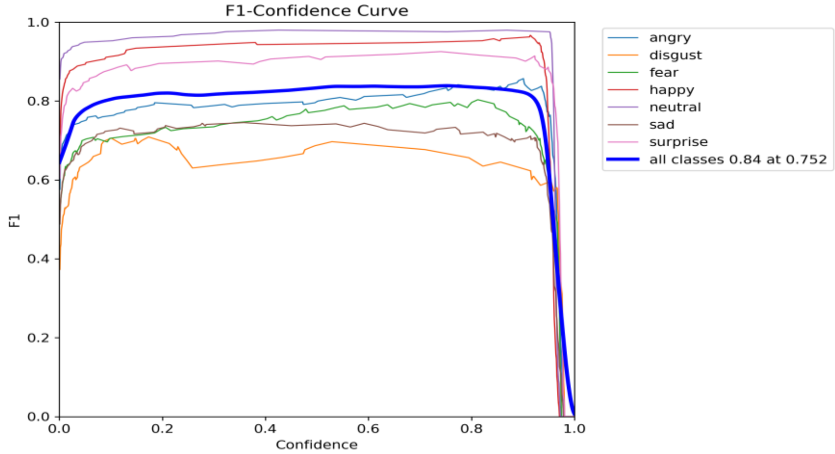

Figure 4. F1-Confidence Curve.

The F1-Confidence Curve in Figure 4 illustrates the trade-off between classification confidence and the F1-score. As the confidence threshold increases, the F1-score may initially improve, reflecting more precise classifications; however, beyond a certain point, the reduction in the number of classifications leads to a decline in the F1-score. This dynamic demonstrates how tuning the confidence threshold can directly impact the balance between precision and recall, which in turn influences the overall performance of the YOLOv8m model.

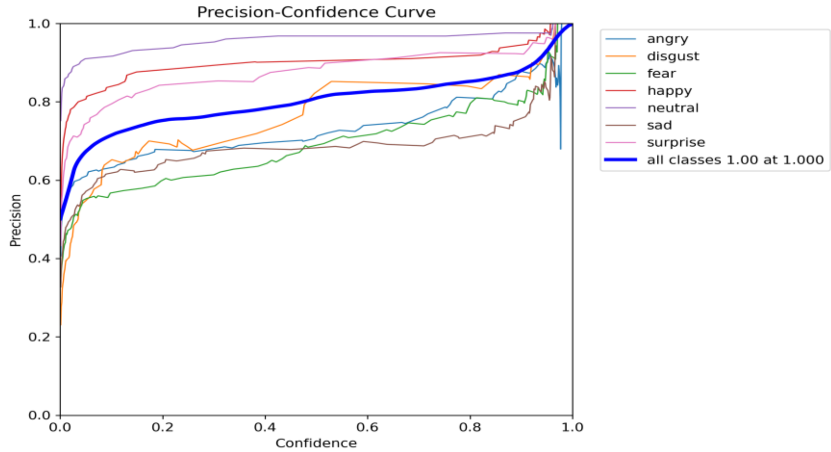

Figure 5. Precision-Confidence Curve.

In Figure 5, the Precision-Confidence Curve is used to display the relationship between precision and model confidence levels. The curve suggests that with higher confidence ratings, the precision of the model generally improves, albeit sometimes at the expense of recall or coverage. The analysis of this curve is crucial for determining an optimal operating point where the model reaches a desirable level of precision without overly compromising recall.

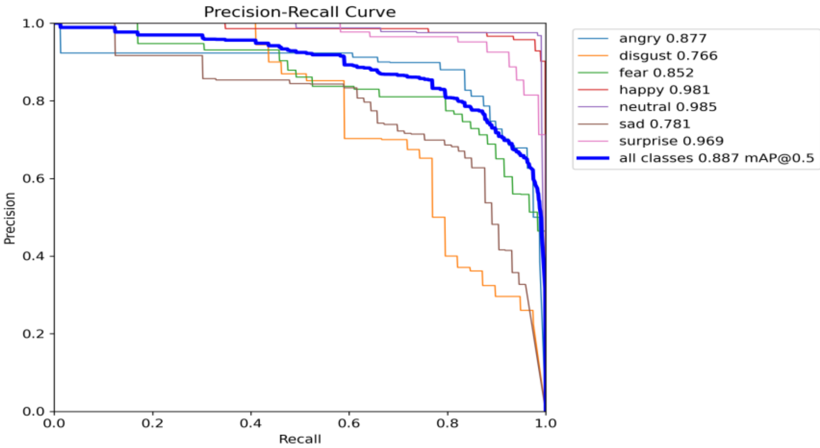

Figure 6. Precision-Recall Curve.

Figure 6 shows the Precision-Recall Curve, which graphically represents how precision and recall trade-off across different threshold settings. This figure is particularly useful in situations where the class distribution is imbalanced. By analyzing where precision and recall curves intersect or exhibit marked changes, one can identify the threshold that best balances the two metrics, thereby optimizing the overall classification performance.

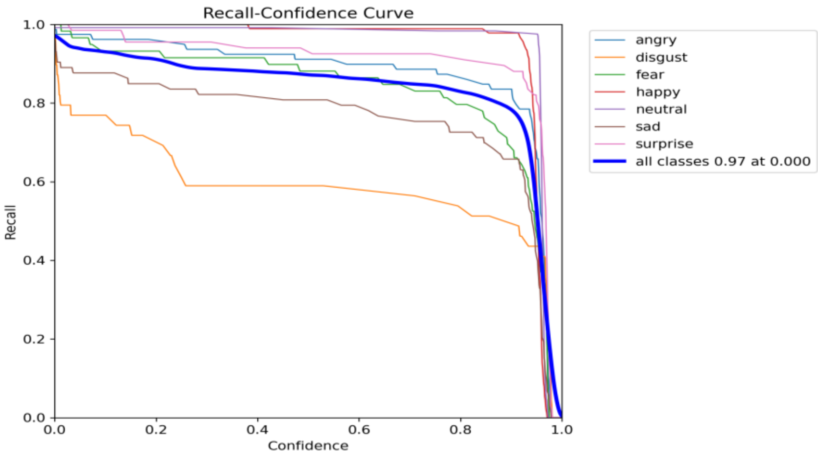

Figure 7. Recall-Confidence Curve.

Figure 7's Recall-Confidence Curve illustrates how the recall value changes relative to the confidence threshold. A high recall at lower thresholds indicates that the model catches most instances of a micro-expression, but as the threshold rises, recall diminishes due to missed classifications. Understanding this behavior aids in fine-tuning confidence levels to ensure that the model captures as many true instances as possible while minimizing false negatives.

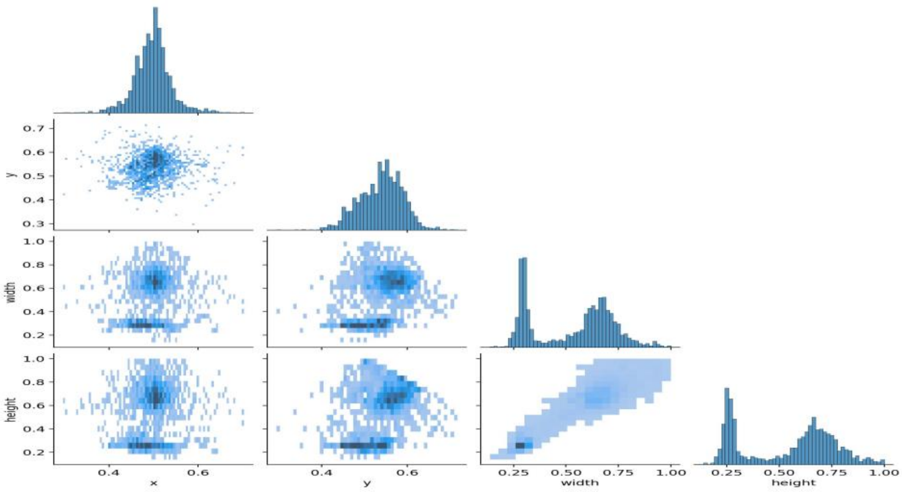

Figure 8. Labels correlogram.

Figure 8 is a visual representation of the interrelationships between different micro-expression labels through a correlogram. The plot shows high positive and negative correlations among certain emotion categories, such as between anger and disgust or between happiness and surprise. These correlations suggest that there is an overlap in the feature space between these emotions, underscoring the potential benefit of feature engineering and advanced attention mechanisms to better differentiate between closely related micro-expressions.

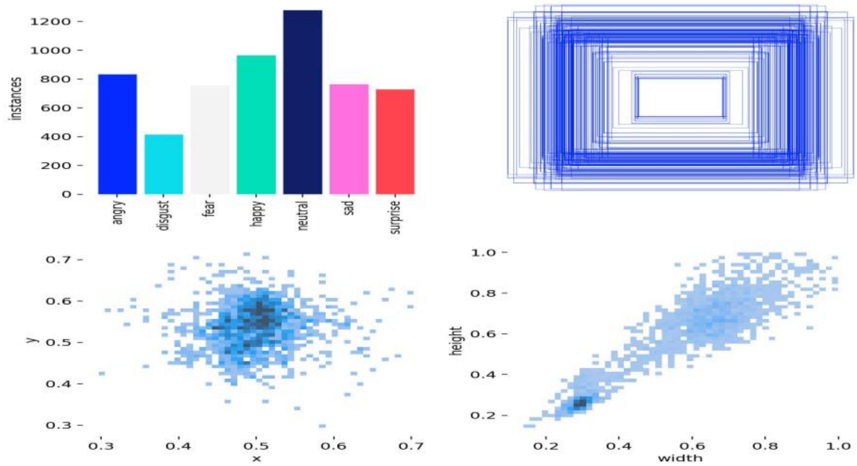

Figure 9. Labels.

Figure 9 includes a mix of visualization techniques such as scatter plots, histograms, and heat maps that detail the distribution and relationships of micro-expression features. These visuals help in understanding the clustering tendencies and spatial distributions of the features. The insights drawn from these visualizations are essential for refining the model, particularly in terms of feature selection and the design of more robust classification algorithms.

Table 2. YOLOv8m Model Performance Summary for Emotion Classification.

YOLOv8m summary (fused): 236 layers, 23,206,885 parameters, 0 gradients, 67.4 GFLOPs

Class	Images	Instances	Box(P	R	mAP50	mAP50-95): 100%| | 9/9
all	531	531	0.841	0.847	0.887	0.814
angry	79	79	0.777	0.886	0.877	0.777
disgust	39	39	0.845	0.557	0.766	0.73
fear	59	59	0.757	0.831	0.852	0.771
happy	92	92	0.913	0.989	0.981	0.861
neutral	122	122	0.968	0.984	0.985	0.969
sad	73	73	0.705	0.753	0.781	0.662
surprise	67	67	0.924	0.925	0.969	0.928

Speed: 0.2ms preprocess, 25.0ms inference, 0.0ms loss, 0.8ms postprocess per image

EarlyStopping: Training stopped early as no improvement was observed in the last 10 epochs. The best results were observed at **epoch 83**, the best model.
To update EarlyStopping (patience=10) pass a new patience value, i.e., `patience=300`, or use `patience=0` to disable EarlyStopping.
93 epochs were completed in **4.951 hours**.

Result Interpretation Table 2: This is a summary of the YOLOv8m model training, which consists of **236 layers and 23,206,885 parameters**. The model has been trained to classify emotions into seven categories: **angry, disgust, fear, happy, neutral, sad, and surprise**.

- **Precision (P), Recall (R), and mAP50-95:**

o **Precision (P)** indicates the accuracy of the model's predictions. Higher values mean fewer false positives.

o **Recall (R)** measures how well the model identifies the correct emotions. Lower values suggest the model is missing some cases (false negatives).

o **mAP50-95** is a metric that combines precision and recall across different confidence thresholds.

- **Performance by Class:**

o The **neutral** class has the highest performance with **mAP50-95 of 0.969**, while the **sad** class has the lowest at **0.662**.

o The **happy** class has the highest recall (**0.989**), meaning the model almost always identifies this emotion when present.

- The **disgust** class has the lowest recall (**0.557**), indicating that the model often fails to detect it.

- **Early Stopping:**
  - Training stopped at **epoch 93** because no improvements were observed in the last **10 epochs**.
  - The best model was recorded at **epoch 83**.
  - The total training time was **4.951 hours**.

Figure 10. Facial Expression Images.

Figure 10 displays sample images from the FER2013 dataset, presenting a diverse array of facial micro-expressions. This grid of images underscores the complexity and variability inherent in facial expressions. The visual diversity captured in this dataset points to the challenge of standardizing the recognition across different subjects and emotional intensities, emphasizing the need for a model that can adapt to high intra-class variability.

These detailed interpretations help to understand the nuances captured by each figure, providing insights into model performance, data distribution, and potential areas for further refinement in YOLOv8-based micro-expression recognition.

DISCUSSION

The YOLOv8-based system for micro-expression recognition demonstrates considerable promise, particularly in its ability to detect and classify clear, distinct emotions such as happiness and surprise in real-time. This capability is evident in the confusion matrices and performance curves, where high recognition accuracy is achieved for these expressions. However, the system shows limitations when processing more subtle expressions like fear and disgust, where nuanced muscle movements and slight variations lead to higher misclassification rates. These challenges indicate that while the current model performs well on more apparent expressions, it struggles with the complexity and fine granularity required for subtle expression detection. The performance trade-offs observed in the F1-Confidence, Precision-Confidence, and Precision-Recall curves further support this assessment, revealing how increasing confidence thresholds improves precision but may also lead to a decrease in recall. Overall, the results underscore both the strengths of the YOLOv8 approach for rapid processing and the need for further refinement in the feature extraction stage to handle the more intricate aspects of facial emotion analysis.

CONCLUSION

This study demonstrates that a YOLOv8-based approach can successfully detect and classify facial micro-expressions in real time, as supported by both quantitative metrics and visual analysis. The model exhibits admirable performance with distinct facial emotions, providing solid evidence through confusion matrices and precision-recall curves that affirm its practical utility. Techniques such as rotation, contrast enhancement, and Gaussian noise application helped create a more diverse dataset, leading to better performance across all emotion classes. However, limitations in detecting and differentiating subtle expressions like fear and disgust have also been clearly highlighted. These challenges underscore the need for further investigation into more intricate feature extraction and model refinement strategies.

Overall, the results validate the feasibility of employing YOLO for micro-expression recognition, paving the way for future advancements. With refined techniques, such as the incorporation of attention mechanisms, hybrid YOLO-RNN architectures, and an enhanced data preprocessing system, it could evolve into a highly robust tool. This would have significant implications for fields ranging from human-computer interaction and behavioral analysis to security and mental health monitoring. The work sets a foundation for further research that aims to optimize both the speed and accuracy of emotion recognition systems, ultimately enhancing their practical application in real-world scenarios.

Future Work

To address the current system's limitations and enhance its practicality, future work should explore several key directions. First, integrating advanced attention mechanisms may help the model focus on the most informative regions of the face, improving the detection of subtle expression features. Future research could also investigate hybrid architectures, such as combining YOLO with Recurrent Neural Networks (RNNs), which could better capture temporal dynamics and context across frames. Additionally, enhancing data preprocessing and employing more sophisticated feature extraction techniques may reduce noise and amplify critical emotional cues. Expanding the dataset with a broader range of expressions and addressing variations in lighting and image quality through robust augmentation strategies are also essential steps. These enhancements could lead to more accurate and consistent performance across all expression types, ultimately making the system more effective for real-world applications in areas like human-computer interaction, security, and mental health monitoring.

Funding
The authors have received financial support from the University "Aleksander Moisiu", Durres, Albania.

Competing Interests

The authors confirm that there are no competing interests related to this work.

REFRENCES

[1] Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial cues.*

[2] Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection, and baseline. *IEEE Transactions on Affective Computing.*

[3] Shreve, M., Godavarthy, S., Goldgof, D. B., & Sarkar, S. (2009). Macro- and micro-expression spotting in long videos using spatio-temporal strain. *IEEE Transactions on Affective Computing.*

[4] Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing, 27(6), 803-816.*

[5] Zhao, G., & Pietikäinen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), 915-928.*

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[7] Zhang, J., Zhao, X., & Liang, Y. (2020). Micro-expression recognition with hybrid deep learning model. *IEEE Transactions on Image Processing, 29, 3547-3559.*

[8] Goodfellow, I., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. *Proceedings of the International Conference on Neural Information Processing Systems, 117-124.*

[9] Lu, W., Liong, S. T., See, J., & Phan, R. C. W. (2018). Micro-expression recognition using a spatiotemporal completed local quantized pattern. *Pattern Recognition Letters, 107, 50-58.*

[10] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767.*

[11] Bakiasi, V., & Muça, M. (2024). Improving facial expression classification through ensemble deep learning models. Nanotechnology Perceptions, 37–47. https://doi.org/10.62441/nano-ntp.vi.387

[12] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934.*

[13] Ma, H., Celik, T., & Li, H. (2021). FER-YOLO: Detection and classification based on facial expressions. *Springer.*

[14] Song, H., Liu, X., Song, M., Bu, J., & Han, G. (2022). Attention-based micro-expression recognition with lightweight convolutional networks. *IEEE Transactions on Neural Networks and Learning Systems, 33(5), 2041-2053.*

[15] Liu, Y., & Zhang, J. (2023). Deep learning-based micro-expression analysis: Recent advances and future trends. *Neurocomputing, 518, 46-62.*

[16] Bushati Hoxha, S., Bushi, F., & Gjecka, A. (2024). Application & challenges of big data technologies in Albanian healthcare system. Interdisciplinary Journal of Research and Development, 11(3), 120. https://doi.org/10.56345/ijrdv11n316.

[17] Pfister, T., Li, X., Zhao, G., & Pietikäinen, M. (2011). Recognizing spontaneous facial micro-expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(6), 1404-1417.*

[18] Verma, A., & Gupta, H. (2020). Bias and fairness in facial emotion recognition algorithms. *ACM Conference on Fairness, Accountability, and Transparency, 423-433.*

[19] Oh, Y., & Jeong, M. (2021). A review on micro-expression databases, features, and classification methods. *Multimedia Tools and Applications, 80(2), 1231-1260.*

[20] Wang, S. J., Li, B. J., Liu, Y. J., & Zhao, G. (2014). Micro-expression recognition using spatiotemporal completed local quantized patterns. *Neurocomputing*, 136, 23-30.

[21] Kim, J., Shin, H., & Kim, J. (2016). Deep learning-based facial expression recognition for facial micro-expressions. *IEEE Transactions on Affective Computing*, 7(3), 274-285.

[22] Liong, S. T., Wong, K., & See, J. (2018). Automatic micro-expression recognition using deep learning. *Journal of Visual Communication and Image Representation*, 55, 185-195.

[23] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788.

[24] Zhao, X., Liu, Y., & Fu, H. (2020). Real-time facial expression recognition using YOLO-based deep learning. *Multimedia Tools and Applications*, 79(7-8), 5167-5180.

[25] Wu, Y., Zhang, H., & Li, P. (2022). Micro-expression recognition via YOLO and RNN-based feature learning. *Pattern Recognition Letters*, 155, 45-52.

[26] Liu, Y., Zhang, J., & Wu, X. (2021). Micro-expression augmentation using GANs for improved recognition. *IEEE Transactions on Affective Computing*, 12(3), 675-689.

[27] Link Dataset: https://www.kaggle.com/datasets/msambare/fer2013.

[28] Demir, S., & Sahin, E. (2023) Random Forest Importance-Based Feature Ranking and Subset Selection for Slope Stability Assessment using the Ranger Implementation. European Journal of Science and Technology. 48, 23-28, https://doi.org.10.31590/ejosat.1254337).

[29] Bushati Hoxha, S., & Bakiasi Shtino, V. (2025). Object detection in real-world scenarios using artificial intelligence and machine learning technologies. Journal of Information Systems Engineering and Management, 10(13s), 223-232. https://doi.org/10.52783/jisem.v10i13s.2025.

[30] Shtino, V. B., Muça, M., & Bushati, S. (2024). Machine learning integration for precise facial micro-expression recognition. Journal of Computer Science, 20(11), 1545–1558. https://doi.org/10.3844/jcssp.2024.1545.1558