Facial Expression Recognition Under Partial Occlusion Using Part-Based Ensemble Learning

Evangelions Felix Yehdeya, Wahyono*
Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia

***Abstract*—Facial expression recognition (FER) under partial occlusion remains a challenging task, especially when key regions of the face, such as the mouth and nose, are covered by medical masks. Such conditions significantly reduce the discriminative features available for accurate emotion recognition, limiting the effectiveness of conventional full-face approaches. To address this issue, this study proposes a part-based learning framework that partitions the face into multiple regions, allowing the model to exploit unoccluded areas for expression recognition. The proposed method employs Support Vector Machine (SVM) classifiers trained on Histogram of Oriented Gradients (HoG) features extracted from 2, 3, 4, and 6 facial partitions. Each part-based model is trained independently, and their outputs are combined through a weighted soft voting ensemble mechanism to generate the final prediction. The experiments were conducted on the MaskedFER2013 dataset, which contains 31,116 grayscale facial images (48×48 pixels) distributed across seven emotion classes. The results demonstrate that the four-part model achieves the best performance, reaching an accuracy of 45%, outperforming both single-part models and full-face baselines under occlusion scenarios. These findings confirm that the proposed part-based ensemble approach enhances the robustness of FER systems by effectively leveraging complementary regional features, thereby providing a promising solution for real-world applications, where facial occlusion is unavoidable.**

***Keywords*—Facial expression recognition; partial occlusion; partial part model; support vector machine; ensemble learning**

I. Introduction

Facial expressions are among the most critical indicators of human emotion in visual communication, serving as an essential channel for conveying affective states such as happiness, sadness, anger, fear, surprise, and disgust [1]. Unlike verbal communication, facial expressions provide immediate and universal cues that enable individuals to understand others' emotional states. FER has become an important research field with applications in human-computer interaction, healthcare, affective computing, surveillance, driver fatigue detection, and entertainment systems [2]. Integrating FER into modern intelligent systems enhances their adaptability, naturalness, and responsiveness, allowing machines to interact more effectively with humans.

Despite its broad potential, FER remains a challenging task, particularly in scenarios where faces are partially occluded. Objects such as sunglasses, hands, or medical face masks often obstruct critical regions of the face, reducing the discriminative power of visual features. This problem has become even more critical during and after the COVID-19 pandemic, when the widespread use of medical masks made occlusion a common condition in everyday environments [3], [4]. Since key regions like the mouth and chin play a central role in expressing emotions, their absence poses significant difficulties for conventional FER models that assume complete face visibility. Occlusion of critical regions such as the mouth and chin significantly degrades recognition accuracy because traditional FER models are typically trained on unobstructed faces and rely on holistic features.

Recent research has focused on developing robust methods under occlusion to address these challenges. Several occlusion-specific datasets, such as MaskedFER2023 [5], have been introduced to reflect real-world scenarios where masks and other obstacles hide facial features. These datasets provide valuable benchmarks for advancing FER systems that must operate reliably in post-pandemic conditions. Moreover, traditional holistic approaches have been increasingly complemented by part-based learning strategies, which divide the face into localized regions (e.g., upper face, lower face, or quadrants) [6]. It allows the system to extract meaningful features from the visible areas even when some parts are hidden.

The main contribution of this study is introducing a Partial Part Model (PPM) combined with ensemble learning to enhance FER under partial occlusion. Unlike previous approaches that treat the face as a whole, the proposed method divides the face into multiple regions, allowing the system to preserve and exploit discriminative features from visible areas while ignoring occluded regions. Feature extraction is performed using HoG, which ensures computational efficiency and robustness against variations in illumination and scale. Each facial part is independently classified using SVM, chosen for their strong performance in small- to medium-sized datasets and their interpretability [7], [8]. However, its performance tends to deteriorate markedly when facial images are affected by partial occlusion. This degradation occurs because SVM typically relies on global feature representations, making it sensitive to missing or distorted regions. When essential facial areas such as the mouth are covered by objects like medical masks, the resulting feature vectors become less discriminative and less representative of the true facial expression. Consequently, a major challenge arises in enhancing the robustness of SVM-based systems to maintain reliable recognition performance under occluded conditions [9]. The outputs of these classifiers are then aggregated through a soft voting ensemble strategy, producing a final decision that leverages complementary information across regions.

This study is organized as follows: Section II presents the theoretical background and related works on FER and occlusion

*Corresponding author.

handling methods. Section III describes the proposed methodology, including the Partial Part Model, feature extraction process, and ensemble learning strategy. Section IV details the experimental setup, dataset description, and evaluation metrics used in this study. It also discusses the experimental results and provides an analysis of the model's performance under partial occlusion. Section V concludes the study.

II. Related Work

FER under partial occlusion, particularly with face masks, has gained increasing attention recently. In [10], masked facial datasets were introduced, and recognition performance was evaluated, reporting an accuracy of 51.9%, which highlights the difficulty of the task. Building on this, [5] proposed two new datasets, MaskedFER2023 and MaskedCK+, demonstrating improved accuracies of 61% and 63%, respectively. These studies emphasize the need for more robust approaches to handle occluded facial regions.

In [5], the authors also discussed three previously proposed approaches: the Region Attention Network (RAN) [11], which achieved an accuracy of 53%; the Attention-CNN (ACNN) [12], which obtained an accuracy of 57%; and the Occlusion Adaptive Deep Network (OADN) proposed [13], which achieved an accuracy of 59%. These results highlight the incremental progress in addressing facial expression recognition under occlusion and the ongoing need for more robust and generalized solutions.

Beyond dataset development, several works have explored different modeling strategies for FER under mask occlusion. In [14], a dual-model framework was proposed: one model detects mask presence, while another predicts emotions. The system achieved 95% accuracy in mask detection, while expression recognition performance varied between 42% and 83%. CNNs were employed for feature extraction, and Haar cascade classifiers for classification. Meanwhile, [15] combined CNN-based FaceNet feature extraction with SVM classification, achieving an accuracy of 98.93%, outperforming conventional methods. These approaches highlight the potential of combining deep feature extraction with classical machine learning classifiers to improve robustness in occluded FER scenarios.

In parallel, traditional FER datasets without occlusion have significantly contributed to the progress of the field, including CK+ [16], JAFFE [17], and FER2013 [18]. FER2013 provides a large-scale collection of 28,709 training, 3,589 validation, and 3,589 test images across seven emotion categories. In this study, the Masked FER2013 dataset is adopted, which extends FER2013 with mask occlusion and consists of approximately 35,900 images with variations in orientation, race, and gender. This dataset serves as the foundation for our experiments.

Despite these advancements, previous research still faces notable limitations. Most methods rely on holistic facial representations, making them vulnerable when critical regions like the mouth or chin are obscured. Deep learning approaches, while powerful, often require large-scale annotated datasets and substantial computational resources, limiting their practicality in real-time or low-resource applications. Moreover, hybrid approaches combining CNN with SVM, although effective, typically treat the face as a single unit rather than considering the independent contribution of visible sub-regions.

To address these gaps, this study proposes PPM integrated with ensemble learning, designed to enhance the robustness of FER under partial occlusion. The proposed method divides the face into multiple regions, extracts local features using HOG, and classifies each region independently using SVM. The outputs are then aggregated through a soft voting strategy, enabling the system to leverage complementary information from visible regions while minimizing the impact of occluded areas.

III. Methodology

This section presents the methodology for developing the proposed FER system under partial occlusion. The approach is based on a PPM combined with ensemble learning using SVM. The methodology addresses the challenges of occluded facial regions, particularly those covered by masks, by partitioning the face into multiple sub-regions and leveraging localized information for robust emotion classification. The following subsections describe the architecture of the ensemble and PPM, as well as the detailed algorithmic workflow of the system.

A. The Design of the Algorithm

The workflow of the proposed system is illustrated in Fig. 1, which outlines the step-by-step process for handling FER partial occlusion. After preprocessing, the images are partitioned into multiple regions (2, 3, 4, or 6), enabling the system to focus on localized areas that remain visible despite occlusion. This step is critical because certain facial regions, such as the eyes or upper face, often retain discriminative features even when the lower face is masked.

In the training phase, HoG features are extracted from each partitioned region to capture essential texture and edge information, which is then fed into SVM) classifiers. Each SVM is trained independently on its corresponding region, ensuring that localized characteristics are learned effectively. The independent models are then aggregated using an ensemble learning strategy with weighted soft voting, where weights are derived from the classification performance of each sub-model. This approach ensures that more reliable regions contribute more strongly to the final decision.

In the testing phase, the same preprocessing, partitioning, and feature extraction procedures are applied to unseen data. The outputs (probability vectors) from all regional SVM classifiers are aggregated through the ensemble mechanism to produce the final emotion prediction. This ensures consistency between training and inference, while also maintaining robustness against varying occlusion levels.

Finally, the performance of the proposed system is quantitatively evaluated using standard metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the system's effectiveness, capturing not only its correctness in emotion classification but also its ability to remain stable and reliable under different occlusion conditions. Together, these evaluations validate the robustness and generalization capability of the proposed PPM

combined with ensemble learning for FER under partial occlusion.

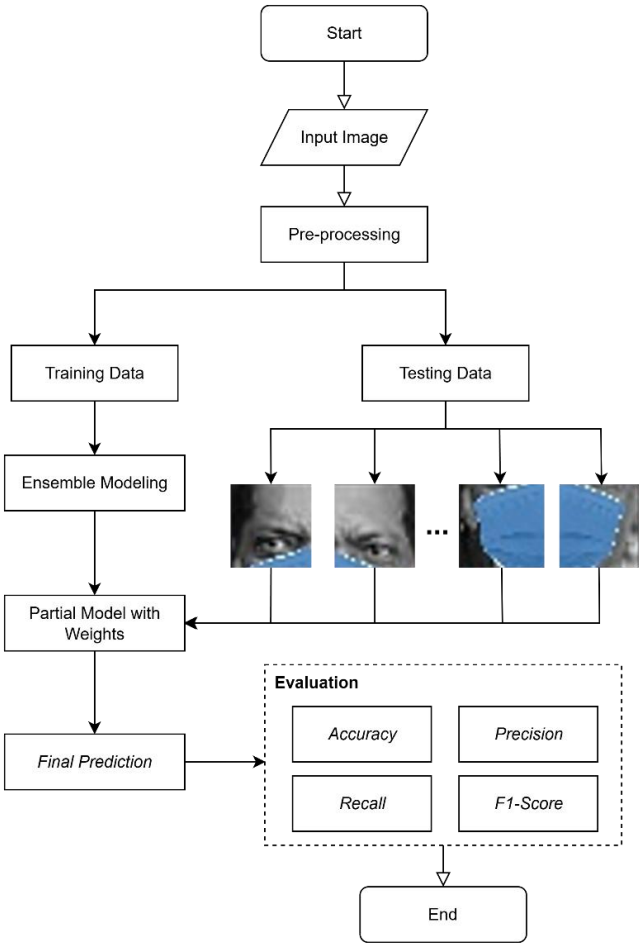

Fig. 1. Workflow of the algorithm.

B. Ensemble and Partial Part Models

Fig. 2 illustrates the architecture of the PPM using an Ensemble SVM approach for FER under partial occlusion [19]. The input facial image is first divided into multiple predefined regions (e.g., top-left, bottom-right), each capturing different parts of the face. These sub-regions are processed independently by separate SVM classifiers, with each model trained to recognize emotional patterns based on its corresponding part.

Each SVM generates a probability distribution over the emotion classes. These outputs are then aggregated using a soft voting strategy, where the final predicted label is determined by averaging the class probabilities across all SVMs and selecting the class with the highest average score. This strategy improves robustness by allowing the model to focus on visible and informative facial areas, compensating for parts affected by occlusion. It also ensures that even if certain regions are occluded or carry less expression-relevant information, other regions can effectively contribute to the final decision. The architecture is shown in Fig. 2.

C. Dataset and Preprocessing

The MaskedFER2013 dataset, containing 31,116 images sized 48×48 pixels, was used. Images were labeled into seven emotion classes: angry, disgust, fear, and many more. Each image underwent grayscale normalization and was partitioned into regions (2, 3, 4, or 6 parts) representing different facial areas [5]. This preprocessing phase is crucial, as it directly influences the quality of features that the model can extract. In this context, dominant features come from the eye and eyebrow regions, which typically display subtle variations across different emotions. Therefore, standardizing image size and format is essential to ensure consistent inputs for the classification model.

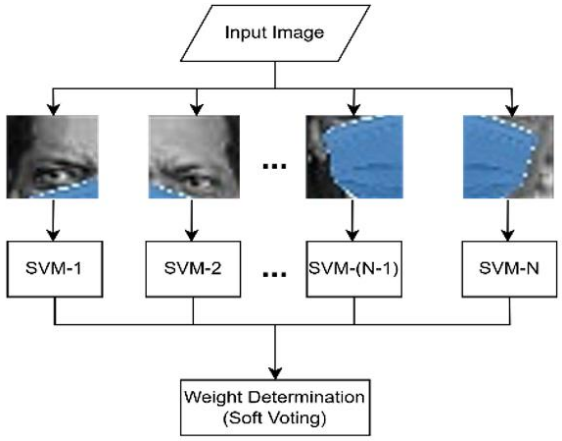

Fig. 2. Architecture of ensemble learning and partial part models.

Facial images were divided into several partitions to address the challenges posed by facial occlusion, particularly due to the use of masks. This partitioning strategy enables extracting localized features from different facial regions, which are combined to improve classification performance.

Fig. 3 illustrates the partitioning strategies applied to facial images with partial occlusion to improve emotion recognition. Each row represents an emotion class (Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise), while the columns show the original preprocessed image and its partitioned forms. The two-partition scheme divides the face into upper and lower regions, separating the eye and mouth areas. The three-partition scheme adds a middle section for finer detail. In contrast, the four-partition scheme splits the face into quadrants (top-left, top-right, bottom-left, bottom-right), enabling localized feature capture across both facial halves. This partitioning helps the model to differentiate features between the left and right sides of the face. The six-partition scheme further subdivides the image into top-left, top-middle, top-right, bottom-left, bottom-middle, and bottom-right regions, providing a highly localized representation at the cost of global context. These strategies enable complementary regional feature extraction, which enhances recognition performance under occlusion when integrated via ensemble learning. Overall, this figure demonstrates how different partitioning strategies will allow the model to capture both local and global discriminative features from unoccluded regions of the face. By leveraging these complementary partitions through ensemble learning, the system becomes more robust in recognizing emotions despite partial occlusion caused by masks. The results of the processed images and partition are shown in Fig. 3.

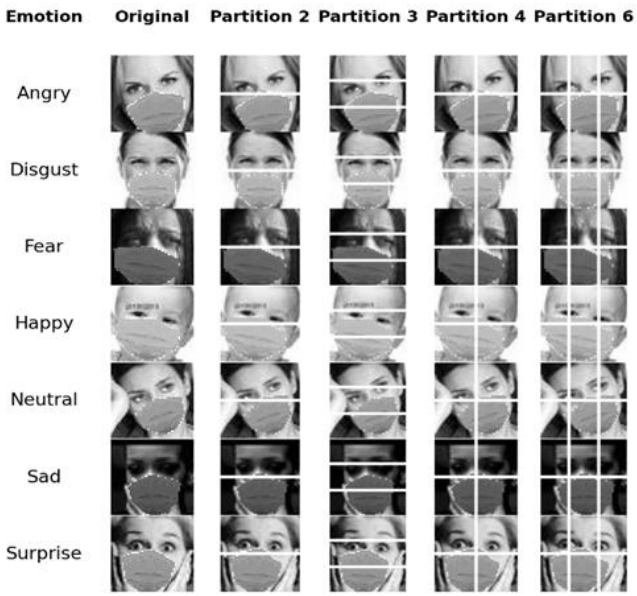

Fig. 3. Preprocessing and partition results.

D. Feature Extraction

In this stage, features were extracted from each partitioned image using the HoG method. HoG is particularly effective for capturing local edge orientations, gradient intensity, and texture patterns essential for distinguishing facial expressions, even under partial occlusion such as masks. By focusing on edge structures rather than pixel intensity, HoG can represent critical facial cues (e.g., eye contour, eyebrow shape, and forehead wrinkles) proper for emotion recognition. Each image partition produced a feature descriptor vector, which was later used as input for the classifiers [20].

E. Ensemble Prediction

Each SVM classifier generated prediction probabilities for the possible emotion classes during the testing phase. Instead of relying on a single classifier, the outputs from all SVMs were aggregated using a weighted soft voting mechanism. In soft voting, the predicted probabilities from individual classifiers were averaged, and the class with the highest average probability was selected as the final decision. This ensemble approach provided robustness, combining complementary information from different facial regions, reducing the impact of occlusion or local noise. Consequently, the ensemble prediction often yielded better overall performance compared to relying on a single model [21].

IV. Experimental Results

In this study, the dataset used is MaskedFER2013, which includes grayscale images with masks applied to simulate real-world occlusions. The dataset is split into two main groups to build and evaluate the model effectively: training data and testing data. From the full dataset, 80% is allocated to the training set and 20% to the testing set. Furthermore, the training set is subdivided into training and validation data, with an 80:20 ratio. The training data is used to train each SVM model on each facial partition. In contrast, the validation data is used during hyperparameter tuning to determine the optimal configuration for each SVM classifier before final training.

A. The Effect of Hyperparameter Tuning

A comprehensive hyperparameter tuning process was conducted to optimize the performance of SVM classifiers across different image partitioning strategies. The key hyperparameters tuned include the kernel type, regularization parameter (C), and kernel coefficient (gamma).

For the partial two-region approach, which divides the image into top and bottom sections, both models achieved the best results using the RBF kernel. The SVM Top model performed best with C = 5 and gamma = 1, reaching an accuracy of 31.19%, while the SVM Bottom model used C = 1 and gamma = 1, achieving 30.12%. This result indicates that both regions still hold significant emotional information despite mask occlusion, with the top region slightly outperforming the bottom. The optimal hyperparameter for the partial 2-Region SVM is shown in Table I.

TABLE I. Optimal Hyperparameter for Partial 2 SVM

Model	Kernel	C	Gamma	Accuracy
SVM Top	Linear	1	0.01	27.57%
	Poly	5	0.1	28.25%
	Rbf	**5**	**1**	**31.19%**
	Sigmoid	5	0.1	26.81%
SVM Bottom	Linear	1	0.1	26.99%
	Poly	1	0.1	28.09%
	Rbf	**1**	**1**	**30.12%**
	Sigmoid	0.1	0.01	26.51%

In the three-region partitioning (Top, Mid, and Bottom), the RBF kernel again yielded the best performance across all sections. The SVM Mid model stood out with C = 1 and gamma = 1, achieving an accuracy of 30.54%, higher than the Top (28.96%) and Bottom (28.13%) regions. This result implies that the middle portion of the face, typically including the eyes and nose bridge, contains richer emotional features under partial occlusion. The optimal hyperparameter for the partial 3-Region SVM is shown in Table II.

TABLE II. Optimal Hyperparameter for Partial 3 SVM

Model	Kernel	C	Gamma	Accuracy
SVM Top	Linear	0.1	0.01	26,47%
	Poly	5	0.1	27.19%
	Rbf	**5**	**1**	**28.96%**
	Sigmoid	0.1	0.01	26.51%
SVM Mid	Linear	1	0.01	27.57%
	Poly	5	0.1	29.52%
	Rbf	**1**	**1**	**30.54%**
	Sigmoid	0.1	0.01	26.51%
SVM Bottom	Linear	0.1	0.01	26.51%
	Poly	1	0.1	26.65%
	Rbf	**1**	**1**	**28.13%**
	Sigmoid	0.1	0.01	26.51%

The four-region partitioning (Top Left, Top Right, Bottom Left, and Bottom Right) showed a consistent trend where all best-performing models used the RBF kernel with C = 5 and gamma = 1. Accuracies in this configuration ranged from 29.64% to 31.05%, indicating symmetrical and localized facial features still provide valuable information for classification when the face is divided into quadrants. The optimal hyperparameters for the partial 4-Region SVM are shown in Table III.

TABLE III. OPTIMAL HYPERPARAMETER FOR PARTIAL 4 SVM

Model	Kernel	C	Gamma	Accuracy
SVM Top Left	Linear	0.1	0.01	28.13%
	Poly	1	0.1	30.14%
	Rbf	**5**	**1**	**30.72%**
	Sigmoid	1	0.1	26.63%
SVM Top Right	Linear	0.1	0.1	28.37%
	Poly	1	0.1	30.36%
	Rbf	**5**	**1**	**31.05%**
	Sigmoid	0.1	0.01	26.51%
SVM Bottom Left	Linear	1	0.1	27.67%
	Poly	1	0.1	28.68%
	Rbf	**5**	**1**	**29.66%**
	Sigmoid	0.1	0.01	26.51%
SVM Bottom Right	Linear	0.1	0.1	27.27%
	Poly	1	0.1	29.36%
	Rbf	**5**	**1**	**29.64%**
	Sigmoid	1	0.1	26.67%

The RBF kernel again dominated in performance in the most detailed six-region partitioning (Top Left, Top Mid, Top Right, Bottom Left, Bottom Mid, and Bottom Right). Most regions achieved optimal results with C = 5 and gamma = 1, except for the Top Mid region, which achieved better accuracy using C = 1. This suggests that the Top Mid region, including the eyes, eyebrows, and forehead, contains more complex features and benefits from lower regularization to avoid overfitting. Accuracies in this configuration ranged from 28.56% to 30.22%. The optimal hyperparameter for the partial 6-Region SVM is shown in Table IV.

Overall, the RBF kernel consistently produced the highest accuracies across all partitioning schemes, highlighting its effectiveness in modeling the non-linear nature of facial expression features. On the contrary, the sigmoid kernel showed the poorest performance in every scenario, indicating it is not suitable for this task. These findings support the hypothesis that local facial analysis, especially under occlusion like face masks, can reveal hidden expression patterns when combined with appropriate kernel functions and parameter settings.

B. Result of the Evaluation Model

To further evaluate the effectiveness of different approaches in handling FER under partial occlusion, a comparative analysis was conducted across several baseline CNN models, a non-partial SVM, and ensemble-based partial SVM. Table V summarizes the performance comparison of each model and partitioning strategy.

TABLE IV. OPTIMAL HYPERPARAMETER FOR PARTIAL 6 SVM

Model	Kernel	C	Gamma	Accuracy
SVM Top Left	Linear	1	1	27.15%
	Poly	1	0.1	29.02%
	Rbf	**5**	**1**	**30.22%**
	Sigmoid	0.1	0.01	26.51%
SVM Top Mid	Linear	0.1	0.1	28.31%
	Poly	**1**	**0.1**	**29.70%**
	Rbf	5	1	29.38%
	Sigmoid	1	0.1	26.99%
SVM Top Right	Linear	0.1	0.01	26.95%
	Poly	1	0.1	28.96%
	Rbf	**5**	**1**	**29.94%**
	Sigmoid	5	0.01	26.55%
SVM Bottom Left	Linear	0.1	0.1	26.81%
	Poly	1	0.1	27.37%
	Rbf	**5**	**1**	**28.66%**
	Sigmoid	5	0.01	26.53%
SVM Bottom Mid	Linear	0.1	0.1	28.11%
	Poly	1	0.1	29.10%
	Rbf	**1**	**1**	**30.08%**
	Sigmoid	5	0.01	27.83%
SVM Bottom Right	Linear	0.1	0.1	26.65%
	Poly	1	0.1	27.41%
	Rbf	**5**	**1**	**28.56%**
	Sigmoid	1	0.1	26.53%

TABLE V. RESULT COMPARISON EVALUATION MODEL

Experiment	Accuracy	Precision	Recall	F1-Score
Non-Partial VGG16	25%	15%	15%	10%
Non-Partial ResNet50	14%	2%	14%	3%
Non-Partial MobileNetV2	21%	14%	15%	13%
Non-Partial SVM	44%	50%	40%	45%
Partial 2 SVM	43%	55%	37%	40%
Partial 3 SVM	40%	59%	33%	34%
Partial 4 SVM	**45%**	61%	39%	42%
Partial 6 SVM	43%	67%	36%	38%

Table V presents the comparative evaluation results of several non-partial deep learning models (VGG16, ResNet50, and MobileNetV2), a non-partial SVM model, and multiple ensemble partial SVM models. These findings suggest that without partial strategies or SVM integration, CNN models

alone are less effective in recognizing facial emotions under occlusion conditions such as masks.

In contrast, the SVM approach delivers significantly more stable results. The Non-Partial SVM model achieves 44% accuracy and 45% F1-score, nearly double the performance of CNN-based models. The relatively higher F1-score indicates that the SVM maintained a better balance between precision and recall, even though the dataset involved masked faces. This finding can be attributed to SVM's ability to work well with small- to medium-sized datasets and handcrafted features (e.g., HoG), which are less data-hungry compared to CNNs. Using non-partitioned images allowed the SVM to capture global facial information (both visible and partially occluded regions), providing a more holistic representation for classification.

Furthermore, ensemble-based partial SVM models reveal distinct performance characteristics. The Partial 2 SVM records 43% accuracy with 55% precision, effectively reducing false positives. Partial 3 SVM shows lower recall (33%) but higher precision (59%), although its F1-score remains relatively modest at 34%. Partial 4 SVM stands out with the highest accuracy (45%) and balanced performance across metrics, achieving 61% precision, 39% recall, and 42% F1-score. Meanwhile, Partial 6 SVM achieves the highest precision (67%), but with lower recall (36%), resulting in an F1-score of 38%.

The F1-score is the harmonic mean of precision and recall, reflecting the balance between these metrics. The Non-Partial SVM achieved the highest F1-score (45%) because it had access to the complete facial representation, enabling it to capture both global and local cues, even when some regions were occluded. In contrast, partition-based models suffered from an imbalance: while precision improved (e.g., 67% for the 6-partition SVM), recall decreased significantly (e.g., 36%), leading to a lower F1 score. This result indicates that partition-based models were more "conservative" in their predictions (fewer false positives), but they missed many true cases (higher false negatives). The holistic non-partial approach helped the SVM achieve a better trade-off, resulting in the highest F1-score overall.

C. The Effect of Ensemble

Ensemble learning was employed in this study to enhance the robustness of FER under partial occlusion. Instead of relying on a single global model, multiple SVM classifiers were trained independently on different facial partitions. Each partition captures localized features from specific regions of the face (e.g., eyes, forehead, or partial cheek areas), which are particularly relevant when masks occlude other regions. During inference, the outputs of these individual classifiers were aggregated using the weighted soft voting technique.

A generalized weighted soft voting scheme was adopted to ensure that the ensemble method remains applicable across different partitioning strategies (two, three, four, and six regions). Let each partition model $i \in \{1, \dots, N\}$ produce a class probability distribution $P_i(c)$ for class $c \in C$. The relative contribution of each model is determined by its validation performance score s_i, which represents accuracy. These scores are normalized into weights as follows:

$$\omega = \frac{s_i}{\sum_{j=1}^{N} S_j}, \quad i = 1, \dots, N \tag{1}$$

The ensemble probability distribution is then obtained through the weighted summation of the outputs of all partition models:

$$P_{ensemble}(c) = \sum_{i=1}^{N} \omega_i P_i(c), \quad \forall c \in C \tag{2}$$

Finally, the predicted class label is assigned based on the maximum ensemble probability:

$$\hat{y} = \arg\max_{c \in C} P_{ensemble}(c) \tag{3}$$

This formulation enables the ensemble strategy to flexibly adapt to different partitioning schemes without modification to the core framework. In the case of equal weighting, setting all $s_i = 1$ yields the traditional unweighted soft voting. By leveraging performance-driven weighting, partitions that demonstrate higher discriminative capability are assigned greater influence in the final decision, thereby improving robustness under partial occlusion scenarios.

Table VI to Table IX summarize the comparative performance between the single partial models and the ensemble approach. The results indicate that the ensemble strategy provides higher overall accuracy and better macro-level metrics compared to individual models. This result shows that weighted soft voting effectively combines complementary strengths from different facial regions. Consequently, the ensemble approach achieves a more balanced and reliable performance across all emotion classes.

TABLE VI. RESULT COMPARISON PARTIAL 2 MODEL

Model	Accuracy	Precision	Recall	F1-Score
SVM Top	0.42	0.50	0.38	0.40
SVM Bottom	0.37	0.44	0.32	0.33
Ensemble Partial 2	0.43	0.55	0.37	0.40

TABLE VII. RESULT COMPARISON PARTIAL 3 MODEL

Model	Accuracy	Precision	Recall	F1-Score
SVM Top	0.38	0.48	0.33	0.35
SVM Mid	0.38	0.45	0.32	0.34
SVM Bottom	0.35	0.48	0.28	0.29
Ensemble Partial 3	**0.40**	**0.59**	**0.33**	**0.34**

TABLE VIII. RESULT COMPARISON PARTIAL 4 MODEL

Model	Accuracy	Precision	Recall	F1-Score
SVM Top Left	0.44	0.52	0.39	0.42
SVM Top Right	0.43	0.51	0.38	0.41
SVM Bottom Left	0.36	0.45	0.30	0.31
SVM Bottom Right	0.37	0.44	0.30	0.31
Ensemble Partial 4	**0.45**	**0.61**	**0.39**	**0.42**

TABLE IX. RESULT COMPARISON PARTIAL 6 MODEL

Model	Accuracy	Precision	Recall	F1-Score
SVM Top Left	0.41	0.52	0.37	0.40
SVM Top Mid	0.36	0.41	0.30	0.31
SVM Top Right	0.41	0.52	0.37	0.39
SVM Bottom Left	0.36	0.47	0.29	0.30
SVM Bottom Mid	0.35	0.46	0.28	0.28
SVM Bottom Right	0.37	0.48	0.30	0.31
Ensemble Partial 6	**0.43**	**0.67**	**0.36**	**0.38**

The evaluation of different partitioning strategies (Table VI to Table IX) reveals varying levels of effectiveness in handling partial occlusion. The two-partition ensemble achieves only moderate performance (accuracy = 0.43), as the division into upper and lower regions often limits feature diversity when masks heavily occlude the lower half. Similarly, the three-partition ensemble slightly improves accuracy to 0.40. However, the finer segmentation does not yield significant performance gains, as critical features may still be missing in one or more regions.

The six-partition scheme demonstrates competitive precision (0.67); however, its overall accuracy (0.43) and F1 score (0.38) remain lower, likely due to the fragmentation of global facial context into overly localized regions. In contrast, the four-partition ensemble strikes the most effective balance by capturing vertical and horizontal facial feature variations. With an accuracy of 0.45 and macro precision of 0.61, it consistently outperforms other partition strategies, confirming that the quadrant-based division provides complementary information while preserving sufficient contextual cues.

Therefore, the four-partition ensemble is identified as the most robust approach for FER under partial occlusion, combining enhanced accuracy with improved precision while maintaining stable recall and F1-score performance.

D. Discussion

This study conducted a series of facial expression classification experiments using the SVM algorithm with different approaches: non-partial, partial two-part, partial three-part, partial four-part, and partial six-part ensembles. The hyperparameter tuning process consistently showed that the best-performing configuration was C=5, gamma=1, with the RBF kernel, regardless of the number of partitions. For example, in the four-part approach (Top Left, Top Right, Bottom Left, Bottom Right), all models converged to the same parameter set.

From the evaluation results, the Non-Partial SVM achieved the highest F1-score (43%), reflecting a balanced trade-off between precision and recall. This result indicates that the model is accurate and sensitive to detecting a wide range of classes, making it suitable for general-purpose applications such as human-computer interaction or assistive technologies. Such performance aligns with the findings of [12], who also reported that traditional CNN and SVM models perform well under full-face visibility but degrade substantially when occlusion increases.

In contrast, the Partial 6 SVM demonstrated the highest precision (67%), making it more reliable for applications requiring high-confidence predictions with minimal false positives. This conservative approach, although sacrificing recall, is valuable in contexts like early emotion disorder detection or security systems. Meanwhile, the Partial 4 SVM achieved the highest accuracy (45%), producing the most significant number of correct classifications overall. Although its F1-score is slightly lower than the non-partial model, this strength makes it suitable for scenarios where maximizing the number of accurate predictions is the main priority. Similar behavior was reported by [13], where models prioritizing confident regions yielded higher precision but lower recall under occluded conditions.

Meanwhile, the Partial 4 SVM obtained the highest accuracy (45%), demonstrating a balanced trade-off between model complexity and region diversity. This configuration benefits from adequate regional representation (Top Left, Top Right, Bottom Left, Bottom Right) without introducing excessive fragmentation that might dilute local feature learning. The performance consistency across partitions suggests that the ensemble learning mechanism, through weighted soft voting, effectively integrates localized cues from multiple facial regions, thus maintaining reliability even under partial occlusion. This finding supports the concept of region-based feature fusion as highlighted in [11], which emphasized the importance of focusing on visible regions to improve recognition accuracy.

To ensure that the model's performance is consistent and not biased toward a specific subset of the data, a 3-fold cross-validation was applied automatically during the hyperparameter tuning process using GridSearchCV from scikit-learn. This mechanism divides the dataset into three parts, where two folds are used for training and one for validation in each iteration. The cross-validation process helps evaluate the model's generalization ability while selecting the best parameters for the Support Vector Machine (SVM).

To further analyze class-wise performance and model consistency, confusion matrices were generated for each regional SVM classifier, as shown in Fig. 4 to Fig. 7. The matrices illustrate that the model performs consistently across multiple emotion categories, with the highest correct classifications observed in neutral and happy classes. This analysis confirms that the model maintains balanced recognition performance across visible facial regions despite partial occlusion.

These matrices clearly show that the "happy" class dominates the correctly classified samples, which aligns with the fact that the "happy" expression has the largest number of samples in the MaskedFER2013 dataset. This class imbalance makes the model more confident in recognizing "happy" features, even under partial occlusion. In contrast, the "disgust" and "fear" classes exhibit noticeably lower recognition rates, often being misclassified as sad or angry. This misclassification trend occurs because these emotions share similar upper-face features (such as eyebrow contraction), which become more dominant when the lower face is covered by a mask.

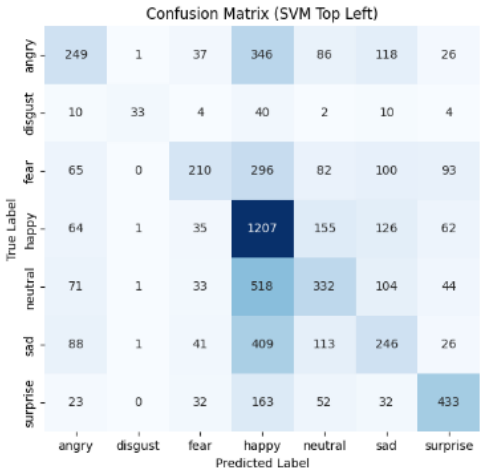

Fig. 4. Confusion matrix (Top Left Partial 4 SVM).

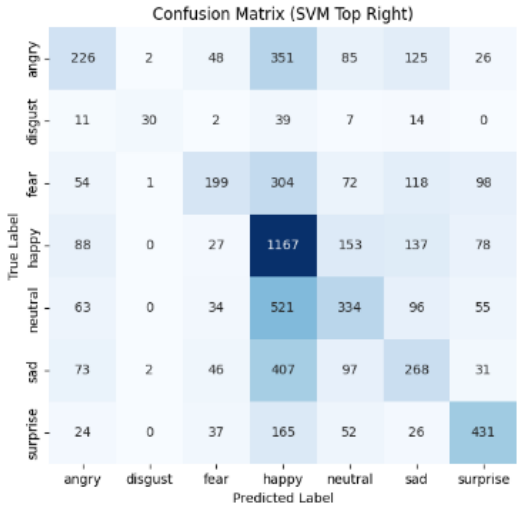

Fig. 5. Confusion matrix (Top Right Partial 4 SVM).

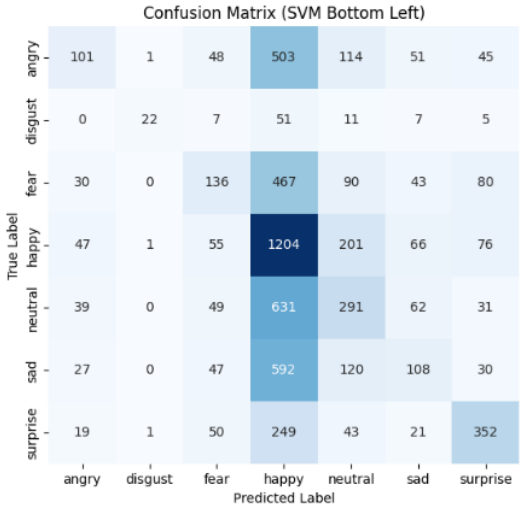

Fig. 6. Confusion matrix (Bottom Left Partial 4 SVM).

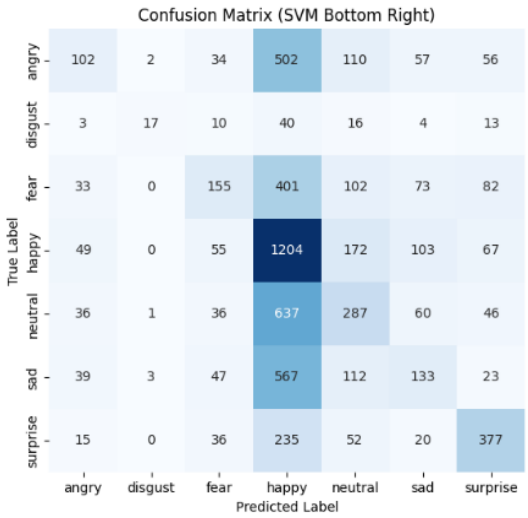

Fig. 7. Confusion matrix (Bottom Right Partial 4 SVM).

Additionally, the "neutral" and "surprise" expressions display moderate accuracy, as these emotions retain distinctive cues around the eye and forehead regions that remain visible even with occlusion. Misclassifications between neutral and happy are also frequent, indicating that when the mouth region is hidden, the model sometimes relies too heavily on the upper-face features, leading to overlapping interpretations.

The relatively consistent diagonal dominance across the four regional SVM (Top Left, Top Right, Bottom Left, and Bottom Right) confirms that the approach maintains robust and regionally stable behavior. This suggests that the ensemble structure successfully mitigates the performance degradation typically caused by partial occlusion, allowing the system to retain a reasonable level of accuracy and interpretability.

Despite these promising results, several limitations remain. First, the proposed method relies heavily on handcrafted features (HOG), which may not fully capture complex spatial and emotional variations compared to deep learning representations. Second, the partitioning strategy is fixed and non-adaptive; dynamic or attention-based segmentation could better capture relevant visible regions under real-world occlusion patterns. Finally, the dataset used (MaskedFER2013) represents a constrained environment with limited occlusion diversity, which may not generalize to unconstrained or in the wild scenarios.

For future research should focus on integrating deep feature extraction with the ensemble SVM architecture, enabling richer and more abstract representations of partially visible faces. Additionally, adaptive region partitioning using attention maps or saliency detection could further enhance robustness by dynamically identifying the most informative facial regions. Expanding the evaluation to include cross-dataset testing (e.g., MaskedFER2023 or MaskedCK+) will also help assess generalization capability. Lastly, incorporating temporal information from video sequences could provide a more comprehensive understanding of emotional dynamics under occlusion.

V. Conclusion

Based on the comparative analysis, no single model can be regarded as the absolute best across all evaluation metrics. Instead, each ensemble configuration demonstrates specific contextual strengths depending on the performance metric prioritized in real-world applications. The Non-Partial SVM model shows a more balanced performance regarding F1-score and recall, making it well-suited for general-purpose use, where both correctness and coverage are equally important. In contrast, the Partial-6 SVM ensemble achieves the highest precision, which indicates its suitability for applications that demand high-confidence decision-making, where minimizing false positives is a priority. Meanwhile, the Partial-4 SVM ensemble yields the highest overall accuracy, suggesting that it is most effective in scenarios where maximizing the total number of correct classifications is the primary objective. These findings highlight that model selection should not rely solely on a single metric, but rather be aligned with the intended application context and performance requirements.

Acknowledgment

The research and publication were funded through the Final Assignment Recognition Program (Program Rekognisi Tugas Akhir) Batch II Year 2022, Universitas Gadjah Mada with Assignment Number 633/UN1.P.III/KPT/HUKOR/2022 and 5722/UN1.P.III/Dit-Lit/PT.01.05/2022.

References

[1] Y. Li, J. Wei, Y. Liu, J. Kauttonen, and G. Zhao, "Deep Learning for Micro-Expression Recognition: A Survey," IEEE Trans. Affect. Comput., vol. 13, no. 4, pp. 2028–2046, 2022, doi: 10.1109/TAFFC.2022.3205170.

[2] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1424–1445, 2012, doi: 10.1109/34.895976.

[3] C. Ramdani, M. Ogier, and A. Coutrot, "Communicating and reading emotion with masked faces in the Covid era : A short review of the literature," Psychiatry Res., vol. 316, no. January, p. 114755, 2022, doi: 10.1016/j.psychres.2022.114755.

[4] D. B. Grahlow M, Rupp CI, "The impact of face masks on emotion recognition performance and perception of threat," pp. 1–16, 2022, doi: 10.1371/journal.pone.0262840.

[5] C. Jiang, R. Hasan, T. Gedeon, and Z. Hossain, "MaskTheFER : Mask-Aware Facial Expression Recognition using Convolutional Neural Network," 2023 Int. Conf. Digit. Image Comput. Tech. Appl., pp. 456–463, 2023, doi: 10.1109/DICTA60407.2023.00069.

[6] M. P. K. Putra and Wahyono, "A Novel Method for Handling Partial Occlusion on Person Re-identification using Partial Siamese Network," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 7, pp. 313–321, 2021, doi: 10.14569/IJACSA.2021.0120735.

[7] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," Proc. - Int. Conf. Image Process. ICIP, vol. 2, no. March 2015, pp. 370–373, 2005, doi: 10.1109/ICIP.2005.1530069.

[8] G. Priyanka and S. Pavithra, "Facial Expression Recognition using SVM With CNN and Handcrafted Features," Int. J. Recent Technol. Eng., vol. 8, no. 4, pp. 3570–3574, 2019, doi: 10.35940/ijrte.d7802.118419.

[9] S. Zhang, X. Zhao, and B. Lei, "Robust facial expression recognition via compressive sensing," Sensors, vol. 12, no. 3, pp. 3747–3761, 2012, doi: 10.3390/s120303747.

[10] S. R. Thavarekere, A. Hebbar, and D. Uma, "A Deep Learning Approach to Facial Expression Recognition in the Presence of Masked Occlusion," INDICON 2022 - 2022 IEEE 19th India Counc. Int. Conf., pp. 1–7, 2022, doi: 10.1109/INDICON56171.2022.10040209.

[11] Y. Wang, Kai and Peng, Xiaojiang and Yang, Jianfei and Meng, Debin and Qiao, "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition," Trans. Img. Proc., vol. 29, pp. 4057–4069, 2020, doi: https://doi.org/10.1109/TIP.2019.2956143.

[12] Y. Li, S. Member, J. Zeng, and S. Shan, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," IEEE Trans. Image Process., vol. 28, no. 5, pp. 2439–2450, 2019, doi: 10.1109/TIP.2018.2886767.

[13] H. Ding, P. Zhou, and R. Chellappa, "Occlusion-Adaptive Deep Network for Robust Facial Expression Recognition," IEEE Int. Jt. Conf. Biometrics, 2020, doi: 10.1109/IJCB48548.2020.9304923.

[14] N. Y. Abdullah and A. M. F. Alkababji, "Masked face with facial expression recognition based on deep learning," Indones. J. Electr. Eng. Comput. Sci., vol. 27, no. 1, pp. 149–155, 2022, doi: 10.11591/ijeecs.v27.i1.pp149-155.

[15] I. J. H. Recto, "Synthetic Occluded Masked Face Recognition using Convolutional Neural Networks," 2022 IEEE Int. Conf. Ind. 4.0, Artif. Intell. Commun. Technol., pp. 124–129, 2022, doi: 10.1109/IAICT55358.2022.9887517.

[16] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010, no. July, pp. 94–101, 2010, doi: 10.1109/CVPRW.2010.5543262.

[17] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," Proc. - 3rd IEEE Int. Conf. Autom. Face Gesture Recognition, FG 1998, pp. 200–205, 1998, doi: 10.1109/AFGR.1998.670949.

[18] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015, doi: 10.1016/j.neunet.2014.09.005.

[19] L. Breiman, "Bagging predictors," Risks, vol. 24, no. 3, pp. 123–140, 1996, doi: https://doi.org/10.1007/BF00058655.

[20] R. Noviyanti and E. Sinduningrum, "Braille Character Recognition with Histogram of Oriented Gradients (HOG) and SVM-Based Image Processing," Syntax Lit. ; J. Ilm. Indones., vol. 9, no. 8, pp. 4411–4419, 2024, doi: 10.36418/syntax-literate.v9i8.16951.

[21] A. Manconi, G. Armano, M. Gnocchi, and L. Milanesi, "A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19," Appl. Sci., vol. 12, no. 15, 2022, doi: 10.3390/app12157554.