Facial Emotion Recognition under Occlusion:

A Systematic Literature Review

Ari Apriansyah, Kusprasapta Mutijarsa, Fadhil Hidayat

School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia

Abstract

Facial Emotion Recognition (FER) faces significant challenges in real-world environments, particularly due to face occlusion and pose variations, which can obscure important features for expression recognition. This study offers a Systematic Literature Review (SLR) that includes 59 articles indexed in Scopus, Google Scholar, and Web of Science from Q1 to Q3, published between 2020 and 2025. The research utilises the Kitchenham methodology for the evaluation and examination of pertinent literature. The findings indicate that occlusion in the eye and mouth regions has a significant impact on the reduction of FER accuracy. Various approaches have been developed, such as the utilization of attention mechanisms in deep learning architectures and face reconstruction techniques, including frontalization, inpainting, and optical flow. While effective in specific contexts, numerous of these methodologies continue to face challenges in achieving optimal generalisation within the intricate conditions of the real world. These results underscore the necessity for further advancement and refinement of both methodologies and more representative and balanced datasets, in addition to the investigation of multimodal and real-time approaches to bolster the resilience of facial expression recognition systems against occlusion and fluctuating facial dynamics.

Keywords

facial emotion, facial expression recognition, facial occlusion, occlusion, occluded face.

# Introduction

F

ACIAL Expression Recognition (FER) is a rapidly advancing domain within image processing and computer vision. This technology has many important uses, such as making human-computer interaction more natural, creating security systems that use facial recognition, and analysing emotions in psychology and marketing [1]. Systems that can accurately read facial expressions can tell how someone is feeling in real time, which is very useful in many social and technological situations.

Facial expression recognition, however, encounters numerous intricate challenges in practice. Occlusion is a big problem for FER applications. It happens when other things, like glasses, masks, hair, hands, or other things that block the view of facial features, cover or block part of the face [2]. This occlusion is prevalent in real-world scenarios, especially in the post-pandemic context characterised by the extensive use of masks, rendering it a significant challenge that FER systems must tackle.

When the face is covered, important information is lost from the covered areas, which makes it harder for the system to get accurate expression features [3]. The recognition model must be able to work with incomplete data to still make accurate classifications when facial features that show certain emotions aren't there. This necessitates the creation of more resilient and flexible approaches for managing occlusion.

Different strategies have been devised to tackle the occlusion problem in FER. Image pre-processing techniques, such as detection and restoration of occluded areas [4], the use of 3D models to reconstruct the face [5], and machine learning algorithms that are resistant to occluded facial data have been the main focus of research in recent years. These methods try to lessen the effects of occlusion and make it easier to recognise expressions in real life.

[6] recently came up with the Multi-Angle Feature Extraction (MAFE) method. This method improves the system's ability to deal with occlusion by combining global features at different scales, fine local features, and important features for specific regions. The Occlusion-RAFDB dataset had an accuracy of 89.42%, and the Occlusion-FERPlus dataset had an accuracy of 86.94%. This was better than what had been done before. This method also worked well on datasets that didn't have occlusion, showing that it is both flexible and reliable.

Moreover, a prevalent challenge associated with occlusion is considerable variation in facial pose. Prior research has introduced methodologies such as the Region Attention Network (RAN) [2] and the Patch Attention Convolutional Vision Transformer (PACVT) [7], which adeptly ascertain the significance of information from various facial regions. These methods effectively tackle both occlusion and pose variation, thereby improving expression recognition performance in intricate and dynamic environmental contexts.

The progress of deep learning and transformer technologies creates new possibilities for facial expression recognition that is resistant to occlusion. These models can learn more complex and contextual feature representations, which makes it easier to deal with missing data caused by occlusion [8]. Nevertheless, challenges such as the necessity for extensive datasets and computational resources persist as significant concerns.

Despite notable advancements in various methodologies, occlusion continues to be the principal challenge impacting FER system performance, especially in real-world scenarios characterised by variability and disturbances [9]. Consequently, additional research and development are imperative to establish more resilient, adaptable, and precise systems capable of effectively managing occlusion and intricate environmental conditions.

This article aims to thoroughly examine the challenge of occlusion in facial expression recognition and to discuss various methods and recent advancements that seek to enhance the system's resilience to such conditions. This review is intended to be a significant reference for the advancement of more dependable and functional FER technology in the future. The research direction will be based on the following research questions (RQ):

RQ1: What types of facial occlusion have a significant impact on the accuracy of FER systems?

RQ2: What composition of data is needed to test the resilience of FER models against occluded face conditions?

RQ3: What approaches have been implemented to enhance FER performance on occluded facial data?

RQ4: How effective are existing approaches in improving FER accuracy on occluded facial data?

# Methodology

This part gives an overview of the steps taken to do this systematic review, such as the sources of information, the search strategy, the criteria for article eligibility, the filtering methods, the coding, the data extraction, and the analysis procedures. We worked together throughout the process to make sure that each step was done correctly and without bias.

## Search Strategy

We got the literature for this study by searching three big databases: Scopus, Google Scholar, and Web of Science (WoS). We used institutional accounts to get to these databases. To ensure the completeness of the search results, the article search was limited to publications from 2020 to 2025, with the aim of ensuring that the reviewed articles are recent and relevant research.

The search process utilized a search string designed with consideration of various synonymous terms and equivalents with similar meanings. This approach aimed to increase the coverage and relevance of the articles found. The search string used was as follows: ("occlusion" OR "occluded face" OR "face occlusion" OR "facial occlusion") AND ("facial expression recognition" OR "facial emotion recognition" OR "face emotion recognition"). A summary of the total number of papers retrieved can be seen in Table I.

TABLE I. Number of Search Results

|  |  |
| --- | --- |
| Publisher | Result |
| Scopus Google Scholar Web of Science  Total | 446 199 64  **709** |

## Inclusion and Exclusion Criteria

Inclusion and exclusion criteria are necessary to maintain the quality of the articles to be reviewed, prevent duplication of search results, and ensure the discussion remains focused on the predefined topic [10]. The criteria for article selection in this review are as follows:

* Articles must be from journals indexed in Q1, Q2, and Q3. Articles from journals indexed below Q3 will be excluded.
* Articles must meet two discussion requirements: first, they must include a discussion on face occlusion, and second, they must be applied to facial expression recognition or facial emotion recognition. Articles that meet only one of these two criteria will be excluded.

## Study Selection

Table I shows the results of the search: Scopus found 446 articles, Google Scholar Direct found 199, and WoS found 64. This is a total of 709 articles from all three sources. The article selection process is illustrated using a PRISMA diagram in Figure 1. In a systematic review, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is a diagram that visually represents the study selection process, which includes identification, screening, eligibility, and inclusion/exclusion of studies [11].

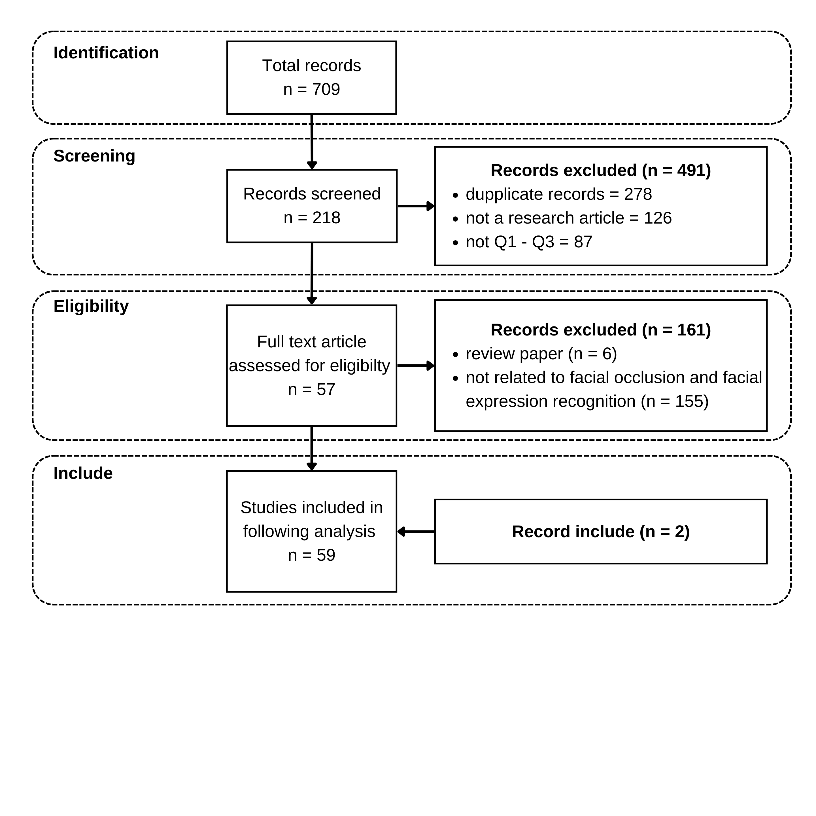


Fig. 1. Review Protocol.

## Selection Result

A total of 59 articles were included in this review after undergoing a rigorous selection process. All selected articles are publications indexed in the databases listed in Table I, with quartile categories ranging from Q1 to Q3, ensuring the quality and relevance of the analyzed research. A complete list of these articles, including their publishers and quartile rankings, is presented in Table II.

TABLE II. Final article result

|  |  |  |
| --- | --- | --- |
| Quartile | Papers | Sum |
| Q1 | [5], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [7], [40], [41], [42], [43], [2], [4], [44], [45], [46], [47], [48], [49], [50] | 43 |
| Q2 | [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65] | 15 |
| Q3 | [66] | 1 |

# Result

## RQ1: What types of facial occlusion have a significant impact on the accuracy of FER systems?

To address Research Question 1 (RQ1) concerning the types of facial occlusion that substantially affect the accuracy of FER systems, a review was performed on the various types of occlusions examined in the literature. However, based on the analysis of the articles that met the selection criteria, most studies did not explicitly mention the specific types of occlusion addressed.

Generally, occlusion is mentioned broadly as a disturbance to facial expression detection, without classifying its form or location. Nevertheless, some studies focus on specific types of occlusion, such as those induced by variations in head pose, glasses, medical masks, or hair, and strive to create methods suited to these circumstances. Table III shows the different types of occlusion that this study looked at. Based on Table III, the types of occlusion used by researchers to test their proposed methods can be categorized into two types: natural occlusion and synthetic occlusion.

TABLE III. Final article result

|  |  |  |
| --- | --- | --- |
| Occlusion | Type of Occlusion Data | Paper |
| Head Pose Variation | Natural | [5], [18], [22], [54], [57], [61], [62] |
| Illumination Variation | Natural | [18] |
| Occlusion in NFPA Context | Natural | [42] |
| Medical Mask | Synthetic | [14], [17], [66] |
| Area Around The Eyes and Hair | Synthetic | [16], [21] |
| Block Masking | Synthetic | [4], [43] |
| Hair and Glasses | Natural | [18], [55], [59] |

*Natural Occlusion*: Occlusion that occurs naturally in real-world environments and is generally uncontrollable by the system, such as head pose variation, illumination variation, and other attributes like hair and glasses. Studies [5], [18], [22], [54], [57], [61], and [62] show that changes in head pose can significantly obscure important areas on the face, thereby reducing the accuracy of FER systems. A similar finding was observed in illumination variation [18], which can lead to inconsistencies in visual features, as well as personal attributes like hair and glasses [55], [18], which cover key facial areas such as the eyes and eyebrows.

*Synthetic Occlusion:* Occlusion that is artificially simulated in experiments to assess the resilience of FER systems under extreme conditions. Common examples include medical masks [14], [17], [66], which cover the mouth and chin area, block masking [43], [4], which involves creating black block objects to randomly cover facial areas, and occlusion of specific areas like the eyes and hair [21], [16]. This type of occlusion is designed to isolate the impact of certain areas on FER accuracy, helping to analyze the system's sensitivity to the loss of visual information.

From this synthesis, it can be concluded that occlusions covering key expression areas such as the eyes, mouth, and eyebrows have a significant impact on the accuracy of FER systems. Both natural and synthetic occlusions have been shown to reduce performance, depending on which part of the face is covered and the overall visual context. Therefore, mapping the types of occlusion is a crucial first step in developing FER systems that are robust to real-world conditions.

## RQ2: What composition of data is needed to test the resilience of FER models against occluded face conditions?

To answer RQ2, this study finds datasets that have facial conditions that are blocked, both naturally and artificially. These datasets are essential for assessing the resilience of Facial Expression Recognition (FER) models to the degradation of visual information in particular facial regions. Several datasets have been explicitly designed or modified to reflect various occlusion scenarios, including the use of medical masks, partial face covering, and visual disturbances caused by viewpoint angles or facial attributes such as hair and glasses. This section will talk about different datasets that are often used in research to see how well FER models do when they have to deal with occlusion problems. It will also talk about how each dataset is different in terms of how well it represents different types of occlusion.

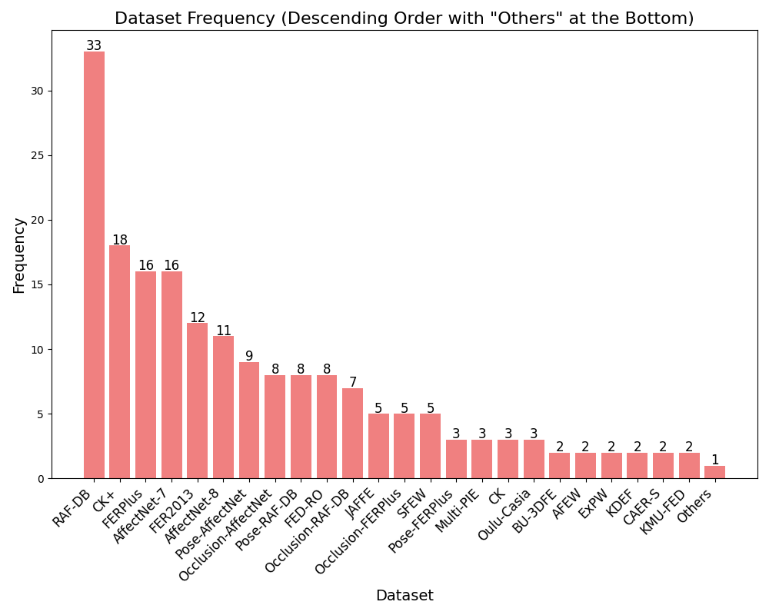


Fig. 2. *Dataset freq used*.

The graph shows the frequency distribution of dataset usage, which shows that many popular and specialised datasets are often used for this purpose in research. Datasets like RAF-DB [67], CK+ [68], AffectNet-7 [69], FERPlus [70], and FER2013 [71] are the most used.

* CK+ (Extended Cohn-Kanade) is a dataset that has 593 sequences of facial expressions from 123 people. Each sequence has an emotion label for the last frame. This dataset is widely used for dynamic expression analysis because it focusses on controlled expressions that change from neutral to peak.
* AffectNet-7 is a part of AffectNet and has more than 450,000 facial images with labels for 7 basic emotions. The images were collected through internet searches utilising emotion-related keywords and were manually annotated by human annotators, resulting in one of the largest and most intricate datasets in emotion recognition.
* FER2013 is a dataset introduced in the Kaggle competition, which is a dataset of 35,887 greyscale facial images that are 48x48 pixels in size and have labels for seven basic emotions. The pictures were automatically taken from the web and became an early standard for recognising facial expressions in a simple visual domain.
* FERPlus is an improved version of FER2013. This dataset also contains 35,887 facial images of size 48x48 pixels, but the emotion labels were updated through re-annotation by eight raters per image, covering 8 emotions: happy, sad, surprise, fear, disgust, anger, neutral, and contempt.

Additionally, there are datasets specifically designed to test models under occlusion conditions, such as Occlusion-AffectNet, Occlusion-RAF-DB, Occlusion-FERPlus, Pose-AffectNet, Pose-RAF-DB, and Pose-FERPlus. All these datasets were created by Wang, K. et al. (2020) [2], with the goal of testing the resilience of models against facial data variations in pose and various forms of occlusion.

* Occlusion-AffectNet is a version of AffectNet that has been modified by adding artificial occlusion (such as hands, masks, or other objects covering part of the face) to evaluate the resilience of models against partially occluded facial expressions.
* Occlusion-RAF-DB is a variant of RAF-DB that introduces occlusion to facial images, either synthetically or from original images with parts of the face covered (e.g., by hands or glasses). This version is used to test the robustness of models under real-world, non-ideal conditions.
* Occlusion-FERPlus is a modification of FERPlus where facial images are occluded in areas such as the eyes, nose, or mouth, aiming to assess the model’s ability to recognize emotions when important facial features are obstructed.
* Pose-AffectNet is a subset of AffectNet focused on facial pose variations, including images with sideways, downward, upward, or tilted poses. This dataset is designed to evaluate the performance of models in recognizing expressions from non-frontal face viewpoints.
* Pose-RAF-DB is a version of RAF-DB that focuses on facial pose variations, including both frontal and non-frontal poses. Pose labels are typically obtained from landmarks or face pose estimation to analyze how facial angle influences emotion classification accuracy.
* Pose-FERPlus is a dataset resulting from the augmentation or subset of FERPlus that retains facial pose variations, used to test model reliability in recognizing expressions when the face is not directly facing the camera.

Information about these datasets is crucial for answering RQ2, as this study provides facial data with real occlusion disturbances, allowing the evaluation of FER model performance in more challenging and realistic situations.

Meanwhile, datasets with lower frequencies such as JAFFE [72], SFEW [73], and Oulu-Casia [74] still provide valuable contributions to model testing, especially for specific conditions or expressions. In Figure 2, there is a category labeled “Others,” which represents datasets in this study with a frequency of one, and these will still be presented in Table IV as alternative datasets that can be used to test FER models under occlusion conditions.

TABLE IV. The "Others" category dataset as shown in **Figure 2**

|  |  |
| --- | --- |
| Dataset | |
| SAMM [75] | LFW [76] |
| RafD [77] | GENKI [78] |
| RM-FRD [79] | DFEW [80] |
| **OLFER [36]** | CelEBA [81] |
| OLFED-HO [44] | Casia-WebFace [82] |
| NFPA [42] | CFEE [83] |
| MYRAF-3\_KN95 [66] | CEFE [30] |
| MYLFW\_KN95 [66] | CASME II [84] |
| MMI [85] | Aff-Wild2 [86] |

The frequency data of dataset usage not only mirrors the inclinations of the research community but also offers a definitive summary of suitable and illustrative dataset selections for evaluating the robustness of FER models in the presence of occluded data. Researchers can utilise this information as a reference for selecting datasets that correspond with their testing requirements, whether they are general or specifically focused on occlusion, thus improving the validity and robustness of the models under development.

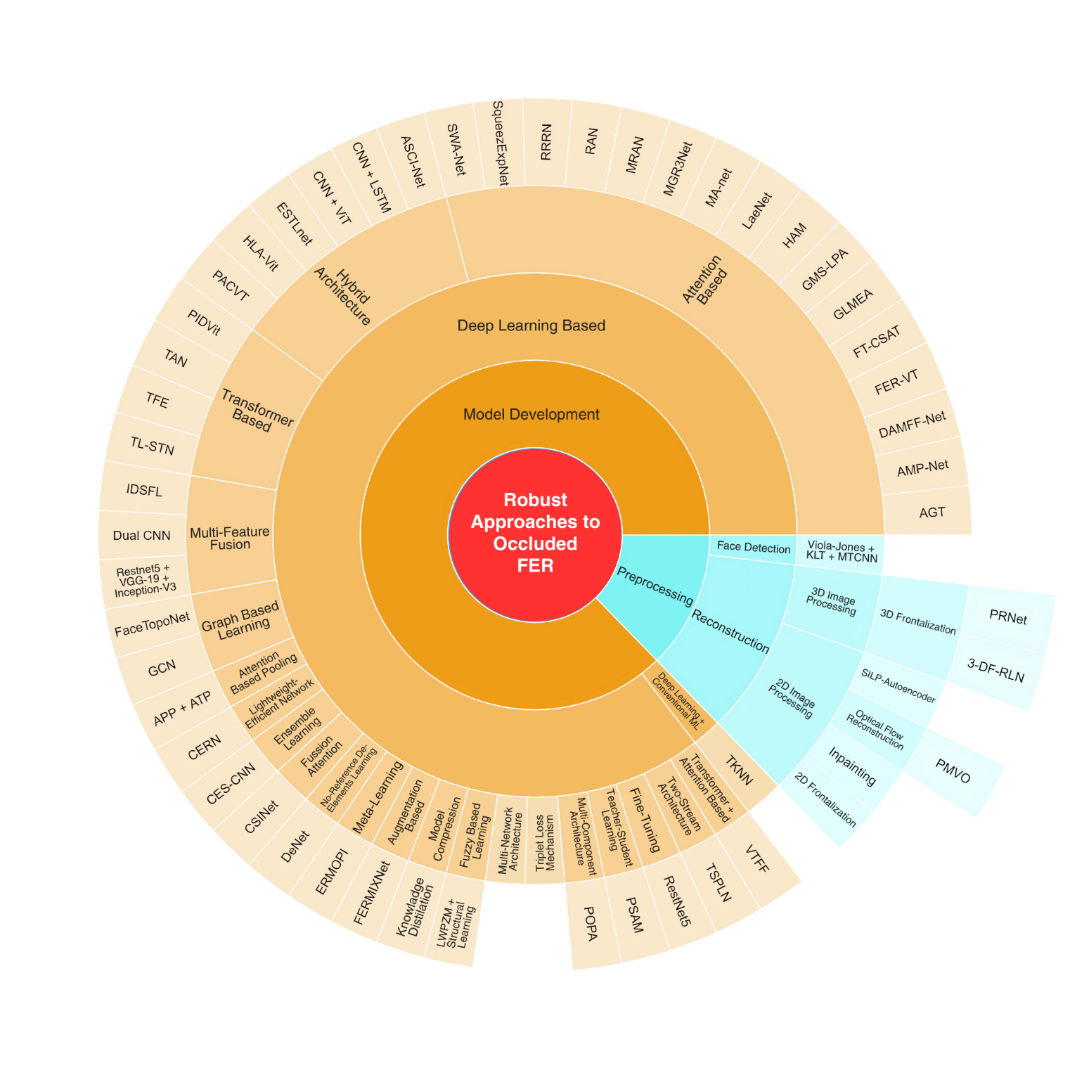


Fig. 3. **Categories of Approaches to Address Occlusion Issues in FER**.

## RQ3: What approaches have been implemented to enhance FER performance on occluded facial data?

All of the chosen papers that talk about occlusion in FER suggest different ways to solve this problem. These differences show that researchers use different strategies to make FER work better on occluded facial data. In this section, we try to group these methods into groups based on their main features, like how they preprocess data, how they build models, or how they extract features. To provide a clearer picture, we present a visualization (Figure 3) that shows the distribution of the approach categories used in these studies.

The research articles from the period 2020-2025 presented in Figure 3 indicate that model development approaches based on deep learning dominate the methods used to tackle occlusion in FER. A more detailed look at the different ways to develop deep learning-based models shows that the attention mechanism is the most popular one. The attention mechanism is a technique that lets the model dynamically focus on the most informative or nonoccluded parts of the face when it is trying to recognise facial expressions. In the realm of FER, attention is employed to ignore occluded or extraneous facial components (e.g., obscured by hands, masks, or other objects) [34], or to highlight essential expressive characteristics such as the eyes, eyebrows, or mouth [32].

In addition to the model development-focused approach, there are also results indicating other approaches such as preprocessing techniques or feature extraction processes. In the preprocessing stage, techniques such as 3D [5] and 2D [43] face reconstruction dominate. Face reconstruction with frontalization is also used to manipulate head poses that are not facing forward to make them face the front, either by using half-face data [62] or by combining two rotated images [61]. This study also highlights a feature extraction approach, specifically through multi-channel feature fusion [34].

## RQ4: How effective are existing approaches in improving FER accuracy on occluded facial data?

To answer RQ4 regarding the improvement in accuracy of existing methods in occluded FER, this study applied a selection rule where the included papers are those that perform tests comparing the baseline condition and the results of the proposed method to assess its impact on FER system accuracy. Unfortunately, not all papers included in the study provided comparisons between the baseline condition and the results of their proposed method. The papers that meet the inclusion criteria are presented in Table V.

A total of 30 studies specifically present the results of ablation experiments, comparing the performance evaluation of models under baseline conditions (without additional components or modifications) and the proposed method (with the suggested approach or architecture). Each study reports evaluation metrics such as accuracy (and, in some cases, weighted average recall) to demonstrate the performance improvement of the model across various benchmark datasets such as RAF-DB, AffectNet, FER2013, CK+, FED-RO, and others.

In the papers presented in Table V, the experiments are not always conducted on just one dataset but may involve two or more datasets. However, in this study, we focus on ablation experiment data from datasets that are representative of occluded facial data and are commonly used to validate the proposed methods, such as RAF-DB, AffectNet, and FED-RO, which contain extensive facial occlusion data

The results of ablation experiments are crucial in measuring the effectiveness of the contributions from the approaches taken to address the occlusion problem in FER. These results also serve as important insights to answer RQ4 regarding the effectiveness of the approaches implemented.

TABLE V**. Comparison of Performance Between Baseline Condition (A) and Proposed Method Result (B)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Evaluation Metrics | Performance % | | Dataset |
| A | B |
| Resnet50 [14] | Accuracy | 72.00 | 78.00 | M-LFW-FER |
| CERN [15] | Accuracy | 63.25 | 65.25 | FED-RO |
| CSINet [53] | Accuracy | 76.43 | 89.67 | RAF-DB |
| GMS-LPA [54] | Accuracy | 84.14 | 87.02 | Multi-PIE |
| Improved Efficient MAS [19] | Accuracy | 84.63 | 86.51 | RAF-DB |
| ESTLnet [20] | Accuracy | 78.54 | 89.38 | Oulu-Casia |
| FT-CSAT [22] | Accuracy | 87.29 | 88.61 | RAF-DB |
| HLA-Vit [57] | Accuracy | 87.22 | 90.45 | RAF-DB |
| TL-STN [24] | Accuracy | 96.87 | 99.41 | CK+ |
| SWA-Net [26] | Accuracy | 87.13 | 90.03 | RAF-DB |
| FER-VT [27] | Accuracy | 76.83 | 84.31 | RAF-DB |
| VTFF [28] | Accuracy | 86.37 | 88.19 | RAF-DB |
| ASCI-Net [60] | Accuracy | 86.29 | 86.86 | RAF-DB |
| FERMIXNet [29] | Accuracy | 82.99 | 86.67 | Occlusion-RAF-DB |
| Face Frontalization [61] | Accuracy | 70.95 | 73.17 | FER2013 |
| GLMEA [30] | Accuracy | 88.04 | 90.09 | RAF-DB |
| HAM [31] | Accuracy | 57.80 | 61.04 | Pose-AffectNet |
| MA-net [33] | Accuracy | 53.55 | 55.30 | Pose-AffectNet |
| IDFSL [34] | Accuracy | 52.69 | 64.02 | Occlusion-AffectNet |
| Knowladge Distilation [36] | Accuracy | 77.50 | 79.50 | RAF-DB |
| MGR3Net [37] | Accuracy | 85.23 | 88.82 | RAF-DB |
| MRAN [38] | Accuracy | 87.84 | 90.03 | RAF-DB |
| DAMFF-Net [64] | Accuracy | 80.15 | 85.94 | RAF-DB |
| PACVT [7] | Accuracy | 84.96 | 88.21 | RAF-DB |
| RAN [2] | Accuracy | 73.33 | 83.63 | Occlusion-FERPlus & Pose-FERPlus |
| RAN [44] | Accuracy | 84.80 | 89.00 | AffectNet |
| TAN [46] | Accuracy | 86.62 | 89.12 | RAF-DB |
| DeNet [49] | Accuracy | 69.75 | 71.50 | FED-RO |
| APP + ATP [50] | Accuracy | 85.02 | 86.03 | Aff-Wild2 |
| RRRN [39] | Weight Avarage Recall | 0.487 | 0.557 | SAMM + CASME II |

# General Discussion

## Significant Impact of Occlusion on FER

Based on the findings in this study, the types of occlusion identified in Table III show a significant contribution to the decline in the accuracy of FER systems. These types of occlusions can be categorized into three main groups based on the areas of the face that are covered: the eye area, the nose area, and the mouth area. Occlusion in the eye area tends to disrupt the identification of expressions that rely on the dynamics of eyebrow and eyelid movement, while occlusion around the nose, although having a relatively smaller impact, still affects the overall facial feature structure. Meanwhile, occlusion in the mouth area significantly affects emotion recognition, particularly expressions associated with mouth shape changes, such as happiness and sadness. This classification provides an important foundation for developing more robust FER methods that are resilient to various types of facial occlusions.

Occlusion in the eye area in real-world environments is commonly caused by the use of glasses, both for medical or protective purposes, as well as by hair covering parts of the face [59], such as bangs or naturally falling strands. This condition is frequently encountered in various surveillance system applications, especially in the context of monitoring vehicle drivers or Advanced Driver Assistance Systems (ADAS) [63], where cameras are directed at the driver's face to detect alertness or emotions. Such occlusion can hinder the accuracy of systems in recognizing facial expressions, as the eye area plays a crucial role in expressing emotions like anger, fear, and surprise. Therefore, the presence of occlusion in the eye area presents a unique challenge in developing vision-based FER systems that are implemented in real-world scenarios.

Occlusion in the mouth area also significantly impacts the accuracy of FER systems. Usually, occlusion in the mouth area can take the form of facial accessories such as masks, which are widely used by people during illness or as part of regular habits. During the COVID-19 pandemic, people were required to wear masks, causing FER systems to become less accurate [14]. Another form of occlusion is when the hand covers the mouth or face, which often occurs during online learning sessions [44]. This situation arises when students feel bored or due to health factors such as a stiff neck or other medical issues [87].

Occlusion in the nose area also influences the recognition of several basic emotions, although its contribution is not as significant as occlusion in the eye or mouth areas. Previous studies have shown that approaches based on modifying muscle movements around the nose area can improve confidence levels in detecting basic emotions like happiness, sadness, and anger [62]. This suggests that, although the nose area is not the primary center of facial expression, its role is still significant in supporting the overall representation of emotional expressions. Therefore, considering the muscle characteristics in this area could be an important aspect of developing more accurate and robust FER systems against partial occlusion.

In addition to occlusion, another factor contributing to the decrease in FER system accuracy is head pose variation [5], [18], [22], [54], [57], [61], [62], which occurs when the orientation of the face in the image is not directly facing the camera due to changes in head direction or tilt. This variation can cause spatial distortion in facial features that are key to expression recognition, making it difficult for the model to consistently extract emotional information. Pose changes such as yaw, pitch, or roll rotations can obscure or even hide essential areas of the face, such as the eyes, eyebrows, and mouth, which significantly impact expression classification. Therefore, handling head pose variation becomes an important challenge in developing robust and reliable FER systems in real-world environments.

## Datasets For Testing The Resilience Of FER Models Against Occluded Facial Data

In efforts to build robust models against occluded facial data, a collection of facial datasets representing real-world conditions, both with and without scenario modifications, is essential. Ideally, these datasets should include variations of occlusion caused by objects such as hands, hair, masks, or other environmental elements, as well as other realistic conditions such as poor lighting, extreme poses, and complex facial expressions. The availability of data with these characteristics is crucial for both training the model to improve its generalization ability and for validating the model's performance in real-world situations more accurately.

Facial data can be gathered from several public datasets such as FED-RO, RAF-DB, AffectNet, FER2013, FERPlus, and others that include a variety of occlusion types. Occlusion types can also be synthetically generated, such as by positioning random black blocks that cover portions of the face [4], [43]. Datasets specifically collected with certain types of occlusion include Occlusion-AffectNet, Occlusion-RAF-DB, and Occlusion-FERPlus for datasets with diverse occlusion types, and Pose-AffectNet, Pose-RAF-DB, and Pose-FERPlus for facial data with varying head poses and angles, which are subsets of AffectNet, RAF-DB, and FERPlus.

The class distribution balance in a dataset is a critical factor in developing reliable and accurate machine learning models. Data imbalance, such as the dominance of certain classes, can lead to models being biased towards the majority class while neglecting the representation of minority classes. Therefore, datasets with balanced label distributions are essential to ensure that the built model can recognize all categories proportionally and maintain generalization performance on unseen data.

* FERPlus is an extension of the FER2013 dataset that includes 28,709 training images and 3,589 test images. This dataset was re-annotated by ten annotators per image, resulting in higher-quality and more reliable labels compared to the original FER2013 dataset [2].
* CK+ (Extended Cohn-Kanade) is a commonly used dataset for FER tasks in controlled laboratory conditions. Although it was not explicitly designed with class balance in mind, its smaller size and controlled data collection environment make it highly consistent and clean [38].
* RAF-DB is an in-the-wild dataset containing both basic and compound expressions. The basic expression subset consists of 15,339 images split into training and testing data. Since the data was collected from real-world environments, RAF-DB offers high variability in terms of pose, lighting, and occlusion [38].
* FED-RO is a facial expression dataset containing real occlusion data collected through internet searches. This dataset is focused on testing and designed to avoid overlap with other datasets like RAF-DB or AffectNet, making it useful for evaluating model robustness against occlusion in real-world environments [38].
* AffectNet-7 is a subset of the large AffectNet dataset, with over 1 million facial images collected from the internet. Approximately 400,000 images were manually annotated into seven basic emotion categories. Its large size and diverse image collection conditions make AffectNet-7 one of the most representative datasets for real-world FER applications [38].
* FER2013 was developed for the ICML 2013 competition and consists of 48×48 pixel grayscale images. Compared to AffectNet and RAF-DB, FER2013 has a smaller scope and more limited expression variety, but it is still widely used as a benchmark for FER model testing [2].
* Occlusion-FERPlus, Occlusion-AffectNet, and Occlusion-RAF-DB are modified subsets of the main datasets that explicitly add occlusion elements (both natural and artificial) to the facial images. These datasets are intended to provide focused evaluation on model robustness in recognizing emotions from occluded faces [37] [28].
* Pose-AffectNet, Pose-RAF-DB, and Pose-FERPlus are subsets designed to evaluate model performance under facial pose variations. Additional annotations are made to identify images with rotations greater than 30° or 45°, allowing for in-depth study of model robustness against changes in face orientation [46] [28].

In terms of data balance, AffectNet-7 is considered one of the most comprehensive and diverse datasets, both in terms of data quantity and annotation variations. This makes it highly suitable for developing FER models that can generalize effectively in real-world environments. For validation of models on occluded facial data with varying head poses, datasets like Occlusion-FERPlus, Occlusion-AffectNet, Occlusion-RAF-DB, or Pose-AffectNet, Pose-RAF-DB, and Pose-FERPlus can be used.

## Technical Approaches to Address Occlusion Issues In FER

In this study, the approach taken by researchers is predominantly focused on model development aimed at building robust models against occluded data based on deep learning (Fig. 3). When broken down in more detail, the dominant approach is based on attention mechanisms, which assign different weights to occluded parts of the face and give more importance to non-occluded areas.

Attention mechanisms have played a significant role in facial expression recognition, both in static and dynamic contexts. At the spatial level, both global and local attention are used to identify key areas of the face that are closely related to expressions by designing spatial attention modules that highlight relevant information from the entire face as well as specific local regions [38]. In dynamic expression recognition, attention mechanisms are used to enhance the discriminative power of spatial-temporal features by utilizing transformer architectures, enabling the model to understand expression sequences more effectively [20].

Additionally, attention mechanisms can also be applied within Convolutional Neural Network (CNN) architectures to develop FER systems that are robust against occlusion, with gating units directing the focus from occluded areas to exposed sections of the face [18]. One innovative approach is the Patch-Gated Convolutional Neural Network (PG-CNN), which is designed to maintain focus on non-occluded, discriminative parts of the face while processing information from occluded areas, thus improving feature extraction effectiveness in complex visual conditions [17].

In addition to the model development-focused approach, another focus area in the machine learning/deep learning pipeline is preprocessing. The preprocessing techniques in this study focus on face reconstruction, both using optical [16], [63], or techniques for frontalizing the ) [61], [62].

Frontalization techniques can be performed in both 2D and 3D image processing. In 2D image processing, frontalization is carried out by augmenting a sideways-facing image of the face into a frame facing the opposite direction, then blending the original and augmented images together using blending techniques to create a frame that appears to face directly forward [61]. In 3D image processing, frontalization involves converting the 2D image into 3D, adjusting the face's position to face forward, and then converting it back into 2D. This approach can be implemented using PRNet [62].

Another form of face reconstruction can be achieved using Generative Adversarial Networks (GANs). GANs can restore occluded parts of the face by detecting and removing the occluding object and filling in the missing section with a representative portion from the available database, commonly referred to as inpainting [43].

The approach to addressing occlusion in FER with a focus on preprocessing is considered crucial in achieving significant results, especially with frontalization techniques. This is based on the idea that frontal faces have better visibility of key emotional determinants, particularly the eyes, facial muscles, and mouth [88]. This ensures that the model is not biased towards the data being tested, especially when working with models trained on datasets that lack sufficient data on occluded faces or head pose variations. This approach helps overcome model bias caused by the limited variation of occluded faces in the training dataset.

# Challenges and Future Research Directions

Although various technical approaches have been discussed in this study to address the challenges of occlusion in facial expression recognition, there are several limitations that need to be considered and serve as avenues for future research development, including:

## Challenges in Dataset Representation

There are datasets that include occluded faces and head pose variations, but most of the data still comes from real-world situations, which don't always show systematic occlusion. For instance, hands can block parts of the face in certain situations, like when someone is learning online. Also, the sample sizes are not evenly spread out among expression classes and types of occlusion, which can have a big effect on how well the model works. We need projects to make new datasets that clearly show different types of occlusion, with more detailed notes about things like the type of object that is blocking the view, the facial areas that are blocked, and the situation (like what activities are going on or how someone is feeling).

## Reliance Only on Attention Mechanism and CNN

Most current methods rely heavily on attention mechanisms in CNN or transformer architectures. These methods work well, but they can have problems when there is a lot of occlusion or visual noise (like bad lighting or low resolution). Using a combination of audio, text (like captions or transcripts), and physiological signals (like EEG or EDA) can help people understand their emotions better when they can't see things clearly because of occlusion.

## Model Generalization to Real-World Data

A lot of models do well on test data from the same dataset (intra-dataset testing), but do poorly on test data from different datasets (cross-dataset testing). This shows that models have a problem with generalisation. We need to develop domain adaptation or domain generalisation methods to make sure that models can handle differences in data distribution between domains, such as occlusion, pose, and expression conditions.

## Challenges in Preprocessing Techniques

Frontalization and inpainting are examples of reconstruction techniques that are often used in preprocessing. These techniques often make fake results that aren't perfect. Optical flow and other techniques need sequential data (like videos), which isn't always easy to find. We need to come up with more realistic and situation-based reconstruction methods, like hybrid generative models that use both GANs and diffusion models, and methods that use more accurate 3D face priors.

## Limited Evaluation in Real-World Contextual Conditions

Most evaluation studies remain confined to controlled or semi-controlled environments, inadequately reflecting the intricacies of real-world applications such as e-learning, emotional monitoring in smart classrooms, or Advanced Driver Assistance Systems (ADAS). Future research must assess FER systems in authentic deployment contexts, including educational, occupational, or digital social interactions, while taking into account environmental variables and user behavior.

# Conclusion

This study presents a systematic literature review (SLR) on Facial Emotion Recognition (FER) under occluded facial conditions, covering literature from 2020 to 2025. The review includes 59 articles from reputable journals (Q1–Q3), examining technical strategies, dataset challenges, and the effectiveness of proposed approaches.

The study found several key insights regarding occlusion in FER. Occlusion in the eye and mouth areas has the most significant impact on the accuracy of FER systems. Both natural occlusions, such as glasses, hair, and head pose, and synthetic occlusions, such as random masking, significantly affect FER performance, especially when they cover key facial expression areas. The review also highlighted popular datasets used to test model robustness against occlusion, such as Occlusion-AffectNet, Occlusion-RAF-DB, and Pose-FERPlus. However, many of these datasets do not fully represent real-world conditions or explicitly include occlusion-type annotations, and class imbalance and lack of contextual representation remain critical challenges.

The review also examined the technical approaches used to tackle occlusion in FER, with a dominant focus on deep learning models integrated with attention mechanisms, both spatial and temporal. Additionally, preprocessing techniques, such as face reconstruction, frontalization, optical flow, and inpainting with GANs, were emphasized to improve the model's ability to focus on non-occluded areas of the face and improve accuracy despite occlusion. The majority of the 30 studies that compared baseline and proposed methods demonstrated a significant improvement in accuracy, with gains of 5–15%, depending on the dataset used, which demonstrates that approaches specifically designed to address occlusion have a positive effect on system performance.

In conclusion, the study emphasizes that, although many technical approaches have been developed to overcome the challenges of occlusion in FER, there are limitations in model generalization, dataset representation, and real-world application scenarios. Future research should focus on the development and optimization of both technical strategies and datasets with detailed occlusion annotations, integrating multimodal inputs such as visual, audio, and physiological data to support FER in extreme occlusion conditions, and creating end-to-end systems capable of handling occlusion across different domains. This SLR is expected to be an important reference for researchers and practitioners aiming to develop FER systems that are not only technically accurate but also robust against visual disturbances and applicable in real-world scenarios like e-learning, smart vehicles, and human-computer interaction systems.

Declaration of Conflicts of Interest

No conflict of interest exists, as the authors affirm that there are no circumstances that could have influenced the content, interpretation, or conclusions of this article, and that all aspects of the research were conducted with full academic integrity and transparency.

Acknowledgment

This research is supported by the resources provided by Ministry of Communication and Digital Affairs of the Republic of Indonesia and Institut Teknologi Bandung

References

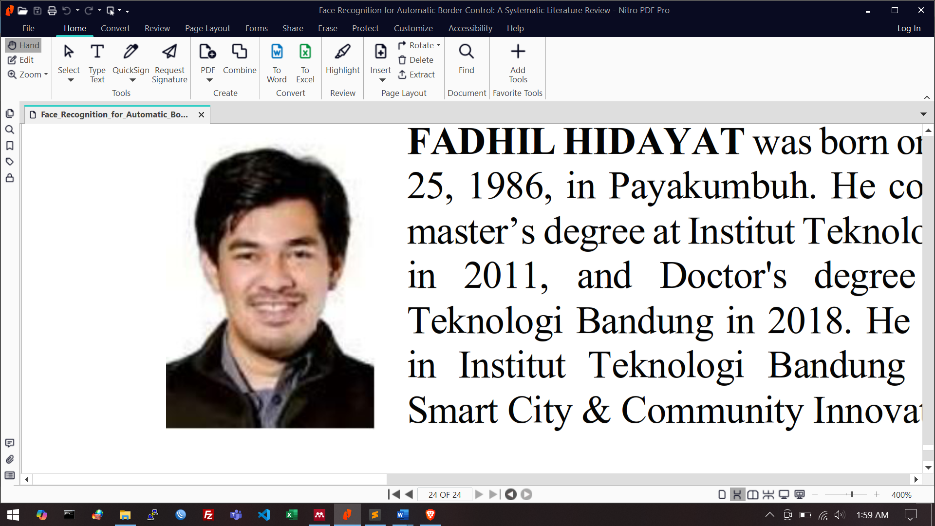
1. O. S. Ekundayo and S. Viriri, “Facial Expression Recognition: A Review of Trends and Techniques,” *IEEE Access*, vol. 9, pp. 136944–136973, 2021, doi: 10.1109/ACCESS.2021.3113464.
2. K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, “Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition,” *IEEE Trans. Image Process.*, vol. 29, pp. 4057–4069, 2020, doi: 10.1109/TIP.2019.2956143.
3. A. F. Abate, L. Cimmino, B. C. Mocanu, F. Narducci, and F. Pop, “The limitations for expression recognition in computer vision introduced by facial masks,” *Multimed. Tools Appl.*, vol. 82, no. 8, pp. 11305–11319, Mar. 2023, doi: 10.1007/s11042-022-13559-8.
4. D. Sun, W. Xie, Z. Ding, and J. Tang, “SILP-autoencoder for face de-occlusion,” *Neurocomputing*, vol. 485, pp. 47–56, 2022, doi: 10.1016/j.neucom.2022.02.035.
5. N. Sun, J. Tao, J. Liu, H. Sun, and G. Han, “3-D Facial Feature Reconstruction and Learning Network for Facial Expression Recognition in the Wild,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 15, no. 1, pp. 298–309, 2023, doi: 10.1109/TCDS.2022.3157772.
6. Y. Li, H. Liu, J. Liang, and D. Jiang, “Occlusion-Robust Facial Expression Recognition Based on Multi-Angle Feature Extraction,” *Appl. Sci.*, vol. 15, no. 9, p. 5139, May 2025, doi: 10.3390/app15095139.
7. C. Liu, K. Hirota, and Y. Dai, “Patch attention convolutional vision transformer for facial expression recognition with occlusion,” *Inf. Sci. (Ny).*, vol. 619, no. 5, pp. 781–794, 2023, doi: 10.1016/j.ins.2022.11.068.
8. T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, “Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets,” *Inf.*, vol. 15, no. 3, 2024, doi: 10.3390/info15030135.
9. D. Zeng, R. Veldhuis, and L. Spreeuwers, “A survey of face recognition techniques under occlusion,” *IET Biometrics*, vol. 10, no. 6, pp. 581–606, 2021, doi: 10.1049/bme2.12029.
10. A. Falavigna and M. Blauth, “Critical review of a scientific manuscript: a practical guide for reviewers,” *J. Neurosurg.*, vol. 128, no. January, pp. 312–321, 2018, doi: 10.3171/2017.5.JNS17809.312.
11. M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.
12. H. Liu, H. Cai, Q. Lin, X. Li, and H. Xiao, “Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6253–6266, 2022, doi: 10.1109/TCSVT.2022.3165321.
13. N. Sun, Y. Song, J. Liu, L. Chai, and H. Sun, “Appearance and geometry transformer for facial expression recognition in the wild,” *Comput. Electr. Eng.*, vol. 107, 2023, doi: 10.1016/j.compeleceng.2023.108583.
14. G. Castellano, B. De Carolis, and N. Macchiarulo, “Automatic facial emotion recognition at the COVID-19 pandemic time,” *Multimed. Tools Appl.*, vol. 82, no. 9, pp. 12751–12769, 2023, doi: 10.1007/s11042-022-14050-0.
15. D. Gera, S. Balasubramanian, and A. Jami, “CERN: Compact facial expression recognition net,” *Pattern Recognit. Lett.*, vol. 155, pp. 9–18, 2022, doi: 10.1016/j.patrec.2022.01.013.
16. D. Poux, B. Allaert, N. Ihaddadene, I. M. Bilasco, C. Djeraba, and M. Bennamoun, “Dynamic Facial Expression Recognition under Partial Occlusion with Optical Flow Reconstruction,” *IEEE Trans. Image Process.*, vol. 31, pp. 446–457, 2022, doi: 10.1109/TIP.2021.3129120.
17. R. Khoeun, P. Chophuk, and K. Chinnasarn, “Emotion Recognition for Partial Faces Using a Feature Vector Technique,” *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124633.
18. S. Kuruvayil and S. Palaniswamy, “Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7271–7282, 2022, doi: 10.1016/j.jksuci.2021.06.012.
19. N. Li, Y. Huang, Z. Wang, Z. Fan, X. Li, and Z. Xiao, “Enhanced Hybrid Vision Transformer with Multi-Scale Feature Integration and Patch Dropping for Facial Expression Recognition,” *Sensors*, vol. 24, no. 13, p. 4153, Jun. 2024, doi: 10.3390/s24134153.
20. W. Gong, Y. Qian, W. Zhou, and H. Leng, “Enhanced spatial-temporal learning network for dynamic facial expression recognition,” *Biomed. Signal Process. Control*, vol. 88, no. PC, p. 105316, 2024, doi: 10.1016/j.bspc.2023.105316.
21. M. Kolahdouzi, A. Sepas-Moghaddam, and A. Etemad, “FaceTopoNet: Facial Expression Recognition Using Face Topology Learning,” *IEEE Trans. Artif. Intell.*, vol. 4, no. 6, pp. 1526–1539, Dec. 2023, doi: 10.1109/TAI.2022.3207450.
22. H. Yao, X. Yang, D. Chen, Z. Wang, and Y. Tian, “Facial Expression Recognition Based on Fine-Tuned Channel–Spatial Attention Transformer,” *Sensors*, vol. 23, no. 15, p. 6799, Jul. 2023, doi: 10.3390/s23156799.
23. H. Kim, J. H. Lee, and B. C. Ko, “Facial Expression Recognition in the Wild Using Face Graph and Attention,” *IEEE Access*, vol. 11, no. June, pp. 59774–59787, 2023, doi: 10.1109/ACCESS.2023.3286547.
24. J. Kim and D. Lee, “Facial Expression Recognition Robust to Occlusion and to Intra-Similarity Problem Using Relevant Subsampling,” *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052619.
25. M. Ahmady, S. S. Mirkamali, B. Pahlevanzadeh, E. Pashaei, A. A. R. Hosseinabadi, and A. Slowik, “Facial expression recognition using fuzzified Pseudo Zernike Moments and structural features,” *Fuzzy Sets Syst.*, vol. 443, pp. 155–172, 2022, doi: 10.1016/j.fss.2022.03.013.
26. S. Qiu, G. Zhao, X. Li, and X. Wang, “Facial Expression Recognition Using Local Sliding Window Attention,” *Sensors*, vol. 23, no. 7, 2023, doi: 10.3390/s23073424.
27. Q. Huang, C. Huang, X. Wang, and F. Jiang, “Facial expression recognition with grid-wise attention and visual transformer,” *Inf. Sci. (Ny).*, vol. 580, pp. 35–54, 2021, doi: 10.1016/j.ins.2021.08.043.
28. F. Ma, B. Sun, and S. Li, “Facial Expression Recognition With Visual Transformers and Attentional Selective Fusion,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1236–1248, Apr. 2023, doi: 10.1109/TAFFC.2021.3122146.
29. Y. Huang *et al.*, “FERMixNet: An Occlusion Robust Facial Expression Recognition Model with Facial Mixing Augmentation and Mid-Level Representation Learning,” *IEEE Trans. Affect. Comput.*, 2024, doi: 10.1109/TAFFC.2024.3454102.
30. Z. Fei, B. Zhang, W. Zhou, X. Li, Y. Zhang, and M. Fei, “Global multi-scale extraction and local mixed multi-head attention for facial expression recognition in the wild,” *Neurocomputing*, 2025, doi: 10.1016/j.neucom.2024.129323.
31. H. Tao and Q. Duan, “Hierarchical attention network with progressive feature fusion for facial expression recognition,” *Neural Networks*, vol. 170, pp. 337–348, 2024, doi: 10.1016/j.neunet.2023.11.033.
32. D. Gera and S. Balasubramanian, “Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition,” *Pattern Recognit. Lett.*, vol. 145, pp. 58–66, 2021, doi: 10.1016/j.patrec.2021.01.029.
33. Z. Zhao, Q. Liu, and S. Wang, “Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild,” *IEEE Trans. Image Process.*, vol. 30, pp. 6544–6556, 2021, doi: 10.1109/TIP.2021.3093397.
34. Y. Tan, H. Xia, and S. Song, “Learning informative and discriminative semantic features for robust facial expression recognition,” *J. Vis. Commun. Image Represent.*, vol. 98, 2024, doi: 10.1016/j.jvcir.2024.104062.
35. C. Wang, J. Xue, K. Lu, and Y. Yan, “Light Attention Embedding for Facial Expression Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1834–1847, 2022, doi: 10.1109/TCSVT.2021.3083326.
36. Y. Chen, K. Li, F. Tian, G. Wei, and M. Seberi, “Lightweight expression recognition combined attention fusion network with hybrid knowledge distillation for occluded e-learner facial images,” *Neurocomputing*, vol. 628, 2025, doi: 10.1016/j.neucom.2025.129656.
37. Y. Wang *et al.*, “MGR3Net: Multigranularity Region Relation Representation Network for Facial Expression Recognition in Affective Robots,” *IEEE Trans. Ind. Informatics*, vol. 20, no. 5, pp. 7216–7226, May 2024, doi: 10.1109/TII.2024.3353912.
38. D. Chen, G. Wen, H. Li, R. Chen, and C. Li, “Multi-Relations Aware Network for In-the-Wild Facial Expression Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 3848–3859, 2023, doi: 10.1109/TCSVT.2023.3234312.
39. Q. Mao, L. Zhou, W. Zheng, X. Shao, and X. Huang, “Objective Class-Based Micro-Expression Recognition Under Partial Occlusion Via Region-Inspired Relation Reasoning Network,” *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 1998–2016, 2022, doi: 10.1109/TAFFC.2022.3197785.
40. Y. F. Huang and C. H. Tsai, “PIDViT: Pose-Invariant Distilled Vision Transformer for Facial Expression Recognition in the Wild,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 3281–3293, 2023, doi: 10.1109/TAFFC.2022.3220972.
41. P. Liu *et al.*, “Point Adversarial Self-Mining: A Simple Method for Facial Expression Recognition,” *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12649–12660, 2022, doi: 10.1109/TCYB.2021.3085744.
42. Y. Zhao *et al.*, “Pose-invariant and occlusion-robust neonatal facial pain assessment,” *Comput. Biol. Med.*, vol. 165, 2023, doi: 10.1016/j.compbiomed.2023.107462.
43. K. Hu, G. Huang, Y. Yang, C. M. Pun, W. K. Ling, and L. Cheng, “Rapid facial expression recognition under part occlusion based on symmetric SURF and heterogeneous soft partition network,” *Multimed. Tools Appl.*, vol. 79, no. 41–42, pp. 30861–30881, 2020, doi: 10.1007/s11042-020-09566-2.
44. L. Lyu *et al.*, “Spontaneous facial expression database of learners’ academic emotions in online learning with hand occlusion,” *Comput. Electr. Eng.*, vol. 97, 2022, doi: 10.1016/j.compeleceng.2021.107667.
45. A. R. Shahid and H. Yan, “SqueezExpNet: Dual-stage convolutional neural network for accurate facial expression recognition with attention mechanism,” *Knowledge-Based Syst.*, vol. 269, 2023, doi: 10.1016/j.knosys.2023.110451.
46. F. Ma, B. Sun, and S. Li, “Transformer-Augmented Network With Online Label Correction for Facial Expression Recognition,” *IEEE Trans. Affect. Comput.*, vol. 15, no. 2, pp. 593–605, Apr. 2024, doi: 10.1109/TAFFC.2023.3285231.
47. W. Xie, H. Wu, Y. Tian, M. Bai, and L. Shen, “Triplet Loss with Multistage Outlier Suppression and Class-Pair Margins for Facial Expression Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 690–703, Feb. 2022, doi: 10.1109/TCSVT.2021.3063052.
48. X. Yang, M. Han, Y. Luo, H. Hu, and Y. Wen, “Two-Stream Prototype Learning Network for Few-Shot Face Recognition Under Occlusions,” *IEEE Trans. Multimed.*, vol. 25, pp. 1555–1563, 2023, doi: 10.1109/TMM.2023.3253054.
49. H. Li, N. Wang, X. Yang, X. Wang, and X. Gao, “Unconstrained Facial Expression Recognition With No-Reference De-Elements Learning,” *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 173–185, 2024, doi: 10.1109/TAFFC.2023.3263886.
50. F. Xue, Q. Wang, Z. Tan, Z. Ma, and G. Guo, “Vision Transformer With Attentive Pooling for Robust Facial Expression Recognition,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 3244–3256, 2023, doi: 10.1109/TAFFC.2022.3226473.
51. A. B. Ahadit and R. K. Jatoth, “A novel dual CNN architecture with LogicMax for facial expression recognition,” *J. Inf. Sci. Eng.*, vol. 37, no. 1, pp. 15–39, 2021, doi: 10.6688/JISE.202101\_37(1).0002.
52. L. Xiong, J. Zhang, X. Zheng, and Y. Wang, “Context Transformer and Adaptive Method with Visual Transformer for Robust Facial Expression Recognition,” *Appl. Sci.*, vol. 14, no. 4, 2024, doi: 10.3390/app14041535.
53. Y. Cheng and D. Kong, “CSINet: Channel–Spatial Fusion Networks for Asymmetric Facial Expression Recognition,” *Symmetry (Basel).*, vol. 16, no. 4, p. 471, Apr. 2024, doi: 10.3390/sym16040471.
54. C. Liu, X. Liu, C. Chen, and K. Zhou, “Deep Global Multiple-Scale and Local Patches Attention Dual-Branch Network for Pose-Invariant Facial Expression Recognition,” *C. - Comput. Model. Eng. Sci.*, vol. 139, no. 1, pp. 405–440, 2023, doi: 10.32604/cmes.2023.031040.
55. S. B. Sukhavasi, S. B. Sukhavasi, K. Elleithy, A. El-Sayed, and A. Elleithy, “Deep Neural Network Approach for Pose, Illumination, and Occlusion Invariant Driver Emotion Detection,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 4, 2022, doi: 10.3390/ijerph19042352.
56. Z. Ullah *et al.*, “Emotion Recognition from Occluded Facial Images Using Deep Ensemble Model,” *Comput. Mater. Contin.*, vol. 73, no. 3, pp. 4465–4487, 2022, doi: 10.32604/cmc.2022.029101.
57. Y. Tian, J. Zhu, H. Yao, and D. Chen, “Facial Expression Recognition Based on Vision Transformer with Hybrid Local Attention,” *Appl. Sci.*, vol. 14, no. 15, 2024, doi: 10.3390/app14156471.
58. L. Ruan, Y. Han, J. Sun, Q. Chen, and J. Li, “Facial expression recognition in facial occlusion scenarios: A path selection multi-network,” *Displays*, vol. 74, 2022, doi: 10.1016/j.displa.2022.102245.
59. S. Bellamkonda, N. P. Gopalan, C. Mala, and L. Settipalli, “Facial expression recognition on partially occluded faces using component based ensemble stacked CNN,” *Cogn. Neurodyn.*, vol. 17, no. 4, pp. 985–1008, Aug. 2023, doi: 10.1007/s11571-022-09879-y.
60. X. Li, C. Zhu, and F. Zhou, “Facial Expression Recognition: One Attention-Modulated Contextual Spatial Information Network,” *Entropy*, vol. 24, no. 7, p. 882, Jun. 2022, doi: 10.3390/e24070882.
61. K. Y. Tsai, Y. W. Tsai, Y. C. Lee, J. J. Ding, and R. Y. Chang, “Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system,” *Signal Process. Image Commun.*, vol. 96, no. April, p. 116321, 2021, doi: 10.1016/j.image.2021.116321.
62. T. Cao, C. Liu, and J. Chen, “Nonfrontal Expression Recognition in the Wild Based on PRNet Frontalization and Muscle Feature Strengthening,” *Math. Probl. Eng.*, vol. 2021, pp. 1–21, Jul. 2021, doi: 10.1155/2021/6620752.
63. S. S. Sudha and S. S. Suganya, “On-road driver facial expression emotion recognition with parallel multi-verse optimizer (PMVO) and optical flow reconstruction for partial occlusion in internet of things (IoT),” *Meas. Sensors*, vol. 26, 2023, doi: 10.1016/j.measen.2023.100711.
64. C. Ge, “Overcoming occlusions in complex environments to achieve robust perception of human emotions,” *Eng. Res. Express*, vol. 6, no. 4, 2024, doi: 10.1088/2631-8695/ad9fd6.
65. J. Gao and Y. Zhao, “TFE: A Transformer Architecture for Occlusion Aware Facial Expression Recognition,” *Front. Neurorobot.*, vol. 15, 2021, doi: 10.3389/fnbot.2021.763100.
66. K. Zheng, L. Tian, Z. Li, H. Li, and J. Zhang, “Incorporating eyebrow and eye state information for facial expression recognition in mask-obscured scenes,” *Electron. Res. Arch.*, vol. 32, no. 4, pp. 2745–2771, 2024, doi: 10.3934/ERA.2024124.
67. S. Li, W. Deng, and J. P. Du, “Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2584–2593, 2017, doi: 10.1109/CVPR.2017.277.
68. 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.
69. A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.
70. E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” *ICMI 2016 - Proc. 18th ACM Int. Conf. Multimodal Interact.*, pp. 279–283, 2016, doi: 10.1145/2993148.2993165.
71. 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.
72. M. Kamachi, M. Lyons, and J. Gyoba, “The japanese female facial expression (jaffe) database,” *Proc. 3rd Int. Con- ference Autom. Face Gesture Recognit.*, vol. 21, no. January, p. 32, 1998, [Online]. Available: https://zenodo.org/records/3451524
73. A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, “Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2106–2112, 2011, doi: 10.1109/ICCVW.2011.6130508.
74. G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, “Facial expression recognition from near-infrared videos,” *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, 2011, doi: 10.1016/j.imavis.2011.07.002.
75. A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, “SAMM: A Spontaneous Micro-Facial Movement Dataset,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 116–129, 2018, doi: 10.1109/TAFFC.2016.2573832.
76. G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments,” in *Workshop on Faces in “Real-Life” Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 2008. [Online]. Available: https://inria.hal.science/inria-00321923
77. O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, “Presentation and validation of the Radboud Faces Database,” *Cogn. Emot.*, vol. 24, no. 8, pp. 1377–1388, 2010, doi: 10.1080/02699930903485076.
78. V. Jain, J. Crowley, and others, “Smile detection using multi-scale gaussian derivatives,” *12th WSEAS Int. Conf. Signal Process. Robot. Autom.*, pp. 149–154, 2013.
79. Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, “Masked Face Recognition Dataset and Application,” *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 5, no. 2, pp. 298–304, 2023, doi: 10.1109/TBIOM.2023.3242085.
80. X. Jiang *et al.*, “DFEW: A Large-Scale Database for Recognizing Dynamic Facial Expressions in the Wild,” in *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2881–2889. doi: 10.1145/3394171.3413620.
81. Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 3730–3738, 2015, doi: 10.1109/ICCV.2015.425.
82. D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Learning Face Representation from Scratch,” 2014, [Online]. Available: http://arxiv.org/abs/1411.7923
83. S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 15, 2014, doi: 10.1073/pnas.1322355111.
84. W. J. Yan *et al.*, “CASME II: An improved spontaneous micro-expression database and the baseline evaluation,” *PLoS One*, vol. 9, no. 1, pp. 1–8, 2014, doi: 10.1371/journal.pone.0086041.
85. M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” *IEEE Int. Conf. Multimed. Expo, ICME 2005*, vol. 2005, pp. 317–321, 2005, doi: 10.1109/ICME.2005.1521424.
86. D. Kollias and S. Zafeiriou, “Analysing Affective Behavior in the second ABAW2 Competition,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-Octob, pp. 3645–3653, 2021, doi: 10.1109/ICCVW54120.2021.00408.
87. R. Beekman, C. C. Tijssen, L. H. Visser, and R. L. L. A. Schellens, “Dropped head as the presenting symptom of primary hyperparathyroidism [3],” *J. Neurol.*, vol. 249, no. 12, pp. 1738–1739, 2002, doi: 10.1007/s00415-002-0898-7.
88. T. Cao, C. Liu, J. Chen, and L. Gao, “Nonfrontal and Asymmetrical Facial Expression Recognition through Half-Face Frontalization and Pyramid Fourier Frequency Conversion,” *IEEE Access*, vol. 9, pp. 17127–17138, 2021, doi: 10.1109/ACCESS.2021.3052500.

**Ari Apriansyah**

Obtained his bachelor degree in Computer System from Universitas Tanjungpura in Pontianak in 2016. Currently, he is a Master's student in Smart-X at STEI Bandung Institute of Technology, a recipient of the Smart X scholarship from the Ministry of Communication and Informatics, His research interests primarily focus on embedded systems and machine learning.

**Kusprasapta Mutijarsa**

He is currently an Assistant Professor with the School of Electrical Engineering and Informatics, Bandung Institute of Technology, Indonesia. He is also the Head of the Autonomous Vehicle Research Group. His current research interests include autonomous and intelligent systems, artificial intelligence and machine learning, robotics, and autonomous vehicles.

**Fadhil Hidayat**

He was born on September 25, 1986, in Payakumbuh. He completed his master’s degree at Institut Teknologi Bandung in 2011, and Doctor's degree at Institut Teknologi Bandung in 2018. He is a lecturer in Institut Teknologi Bandung and Joined Smart City & Community Innovation Center.