Comparative Analysis of MTCNN and Haar Cascades for Face Detection in Images with Variation in Yaw Poses and Facial Occlusions

Omer Abdulhaleem Naser, Sharifah Mumtazah, Khairulmizam Samsudin, Marsyita Hanafi, Siti Mariam Binti Shafie, and Nor Zarina Zamri

***Abstract*—As computer vision and machine learning advance, face detection has become a major focus. Face recognition has several methods and models. Every implementation starts with face detection. Haar Cascades and Multi-task Cascaded Convolutional Networks (MTCNN) are compared for facial pose variation robustness. This research will examine how well these two models detect faces in yaw postures from -90 to +90 degrees. Many studies have contrasted these two models, but the yaw poses of faces were not addressed due to the scarcity of datasets with systematic degrees of face orientation. Thus, the UPM face dataset, created at the UPM embedded systems lab using developed equipment to produce high-resolution photographs and a systematic range of face orientations from -90 to 90 degrees, was used to evaluate the range of degrees these two models can reach. UPM includes 100 students with different yaw angles and occlusions (masks, glasses, or both). The results reveal that MTCNN is the best for detecting faces with yaw poses only, masks, glasses, and both at all degrees (-90 to +90) with 100%, 99.9%, 96.4%, and 80% accuracy. Instead, Haar cascades were 92.5%, 67.3%, 80.4%, and 76.3% accurate.**

***Index terms*—Face Detection, facial occlusions, haar Cascades, MTCNN, occluded faces, UPM dataset, yaw poses.**

I. Introduction

Face detection has progressed with the widespread adoption of deep learning techniques, the ever-increasing security needs, user authentication, and human-computer interaction systems in modern computer vision applications. However, real-world face detection faces challenges such as facial occlusions and variations in yaw poses. In this paper, two recent approaches to face detection — Multi-task Cascaded Convolutional Networks (MTCNN) and Haar Cascades' Classifiers — are compared in terms of performance under different types of facial occlusion and yaw pose variations.

It is well known that MTCNN is widely used in complex and unconstrained environments. Its deep learning methodology and detection performance are optimized through cascaded convolutional networks that perform tasks sequentially. MTCNN generates candidate windows, refines them, and consolidates face landmarks [1, 2]. The model's ability to discern complex facial features, combined with its rigorous training regimen, results in extremely high precision and recall rates, especially in cases with minimal to moderately disruptive facial occlusions [3].

On the other hand, Haar Cascades' classifiers are based on machine learning and use features to detect objects. More specifically, these classifiers were trained using a shallow learning approach with basic Haar features, requiring only a small number of positive and negative instances—making them significantly simpler than the deep features learned by convolutional networks [4]. This is because the model uses a cascading function that sequentially removes non-face regions, resulting in fast computation and suitability for real-time applications. Recent studies have explored the adaptability of Haar Cascades' classifiers to different scenarios, including their compatibility with standard techniques and ability to recognize human faces [5, 6].

The experimental analysis of these two models was carried out on the UPM face dataset, which consists of four subsets. The first subset contains face images with various yaw poses ranging from -90 to 90 degrees. The second subset contains face images covered with masks at various yaw poses within the same range. The third subset contains face images covered with glasses at various yaw poses, and the last subset consists of face images covered with both masks and glasses at different yaw angles. This dataset epitomizes the complexities of real-world scenarios, making it an appropriate benchmark for evaluation.

MTCNN and Haar Cascades have been widely applied and compared in previous studies, as they are two well-known face detection models proven to be robust. However, those studies [7, 8] did not conduct a systematic analysis of face detection across various yaw poses, as they applied datasets that were not specifically designed to measure the robustness of these models in detecting faces at different yaw angles. Therefore, the systematically organized UPM dataset will be used to evaluate the robustness of these two models. Figures 1–3 illustrate examples of non-systematic face orientation datasets

Manuscript received October 18, 2024; revised November 4, 2024. Date of publication March 31, 2025. Date of current version March 31, 2025. The associate editor prof. Vladan Papić has been coordinating the review of this manuscript and approved it for publication.

Authors are with the Universiti Putra Malaysia, Selangor, Malaysia (e-mails: omer.naser@uoitc.edu.iq, {s_mumtazah, khairulmizam, marsyita, mariam}@upm.edu.my, norzarina.zamri@cybersecurity.my).

Digital Object Identifier (DOI): 10.24138/jcomss-2024-0084

used in previous studies to compare the robustness of MTCNN and Haar Cascades.

This selection of MTCNN and Haar Cascades as face detection models for this study was made due to the complementary strengths and weaknesses of these two models. As a deep learning architecture with a multi-stage cascaded convolutional network, MTCNN performs well in scenarios involving complicated pose variations and partial occlusions, where it can detect faces from a wide range of angles. However, Haar Cascades is computationally less demanding, making it a faster and more cost-effective option for simpler scenarios. In face detection, both models have been extensively studied, but their performance under systematic yaw variations—one of the most challenging problems in real-world applications—has not been sufficiently compared.

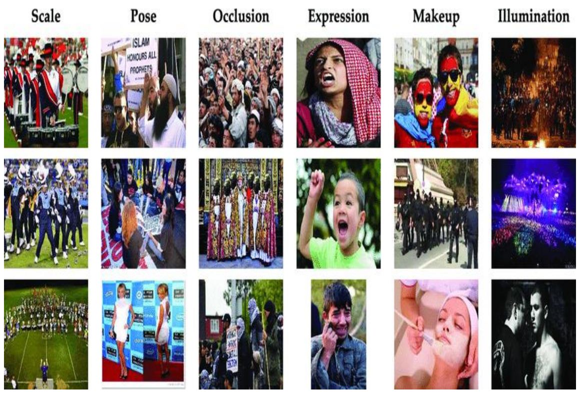

Fig. 1. Wider Face Dataset [7]

Fig. 2. Faces in the Wild Dataset [7]

Fig. 3. Labeled Face in the Wild (LFW) datasets [9]

Figures 1-3 clearly show that face orientations are not being taken in systematic degrees since the degree of each yaw pose of a face is not precisely specified. Therefore, this paper will be analyzing the capability of these two proposed face detection models to check to what degree it can detect faces at different yaw poses. In this study, we prioritize yaw poses (horizontal plane rotations of the face) as they often appear in surveillance and human-computer interaction applications. Unlike tilt and roll, which are often mitigated via camera alignment or preprocessing, yaw poses directly influence the frontal visibility of key facial features and thus are more important for face detection robustness.

The main contributions of this work are:

- Development of a comprehensive dataset with yaw poses and facial occlusions.
- Comparative analysis of MTCNN and Haar Cascades for robust face detection under various conditions.
- Evaluation of face detection performance in occluded and rotated faces, providing insights for real-world applications.

As yaw poses directly affect the visibility of the key facial landmarks (eyes and nose), which are critical for detection, they are prioritized in this study. In contrast to pitch or roll, yaw poses tend not to have been considered in preprocessing or camera alignment.

The focus of this study is on the yaw angle, which frequently occurs in real-world applications, including horizontal face orientation cases for the camera or on the pinpoint input by the user for a surveillance system. Our dataset focuses on yaw variations, as yaw variations are a primary challenge in face detection. Future work could add additional angles (e.g., pitch and roll) to expand the modeling accounting for face detection.

This paper is organized as follows: Section I provides a brief introduction to both face detection models. Section II explores the available face detection models, including MTCNN and Haar Cascades. In Section III, the methodology used in the experiments, including a dataset and face detection models, among others, is described. The experimental results are presented in Section IV, and MTCNN and Haar Cascades are compared under different conditions. In Section V, we discuss the importance of the findings and then conclude the paper in Section VI.

II. Comprehensive Theoretical Basis

One of the basic tasks of computer vision is face detection, which is essential in surveillance, security systems, and social networks. This paper aims to compare the strengths and weaknesses of two widely used face detection models—MTCNN and Haar Cascades Classifiers—along with their real-world implementations.

A. Multi-task Cascaded Convolutional Networks (MTCNN)

The face detection system, which is the Multi-task cascaded convolutional network (MTCNN), has gained the attention of many scholars of this generation due to its effectiveness and flexibility. In [10], the authors described some potential problems that may arise in MTCNN and underlined the fact that these problems should be taken into consideration when using the method in real-world security applications. In [11], the authors provided a detailed study of MTCNN, about the concept and implementation of the model, its applications in computer vision, and its significance. This paper has used MTCNN for face detection using the Wider Face dataset; therefore, the accuracy rate of training this dataset was 85%. Another work presented in [12] introduced an improved MTCNN for face detection in the classrooms that is beneficial for educational systems where the identification of faces is crucial for activities like tracking attendance or monitoring the activity in the classroom.

To enhance small-scale face detection, a deep residual feature generation subnetwork is incorporated into the network architecture. This module has the properties of low-level fine granularity and transforms the initial poor features into higher-level deformation features. Thus, the simplified MTCNN model is arrived at by removing all the components that are linked to the landmarks from the original model. This model is then followed by a deep residual feature generation module to enhance the detection while at the same time enhancing the speed. In [13], MTCNN was not only used in face detection but also for emotion detection and human-computer interaction in real-time, this showed the sophistication of the said model for real-time application. In particular, [14] demonstrated that it is possible to integrate MTCNN with facial expression recognition, which confirms its effectiveness in more complex and multi-faceted tasks, including the identification of humans' feelings and actions.

The author of [15] put forward a system of access control using MTCNN, and the author concentrated on the effectiveness of MTCNN in access control and recognizing systems and also the importance of MTCNN in real-life security systems. The paper [16] also looked into multi-view face detection and landmark localization, where MTCNN was used and has been applied to different tasks, including augmented reality and 3D face models. MTCNN++ [1], a CNN-based face detection algorithm that builds on the MTCNN concept, demonstrates how MTCNN informed the advancement of face detection algorithms and how it defines the advancement of the same. All in all, these papers show that MTCNN is versatile and has applications in the enhancement of face detection across different fields.

B. Haar Cascades Classifiers

Haar Cascades classifiers have made a great contribution to face detection due to their performance, flexibility, and success in a variety of conditions and uses. Other scholars have followed suit and have taken the classifiers' functions to present-day applications such as social media networks, particularly in the context of content curation on Instagram [17]. [18, 19] analyzed the possibilities of the classifier to be combined with other classifiers or methods, for example, the Fisherface algorithm, and the increase in the classifier's efficiency, which makes it promising for application in complex and multifaceted tasks of face analysis. The applicability of the model was investigated by [20, 21], where results revealed the model's effectiveness in different illuminations and the model's ability to be fine-tuned, which is useful for outdoor surveillance and specific applications, respectively.

At the same time, [22] gives a rather detailed and fair appraisal of the Haar Cascades and its effectiveness, which is necessary for understanding the further perspectives of its application in real life. Within the general context of Haar Cascade's recognition as one of the primary methods for face detection, [23] noted that neural networks could be incorporated into it. More evidence of the model's adaptability, [24] used it in multi-face recognition, which is important in crowd and group identification, while [25] focused on its applicability in security-related activities such as biometric identification. Altogether, these works demonstrate the efficiency of Haar Cascades classifiers in face detection, stress their versatility, the possibility of their interaction with other systems, and their significance in the traditional and innovative uses in the constantly developing field of computer vision.

Face detection and facial attribute analysis have also been the focus of a number of recent studies, which explore the problem of pose variations, occlusions, or other real-world challenges. As an example, YOLO-FaceV2 combines YOLOv5 by increasing the receptive field to find small faces and installing attention mechanisms to boost the performance of face detection under occlusions. Additionally, this model employs a novel repulsion loss to minimize false detections resulting from occlusions and outperforms previous YOLO models on subsets of the WiderFace dataset, which is known to be challenging [26]. In particular, it is robust, yet its computational complexity might hinder its application to the real-time case.

Moreover, an IoT-based MTCNN model achieves face detection for such low-resource environments as smart doorbells. This model makes significant gains in detection speed and accuracy under occlusions and unconstrained pose variations through depthwise separable convolution blocks. Though designed for IoT deployment, performance requires computational optimizations that enable it to operate well under noisy lighting and multiple face detections [27].

The Pose Invariant Face Recognition (PIFR) is addressed in another novel approach using a self-supervised Random Mask Attention GAN (RMAGAN) to fill in the gaps that occur due to pose variations present in face images [28]. Unlike traditional GAN-based PIFR methods, Mask Rotate GAN does not require paired frontal view data and is therefore more scalable. The method presented here is computationally expensive, especially for real-time applications, and although the model exhibits good geometry-preserving properties while dealing with extreme poses, there is room for future improvement.

These recent advancements reinforce the existing difficulties of pose distortions and occlusions in face detection and highlight the necessity to balance model precision against affordance for computation. In this study, we compare MTCNN to Haar Cascades to gain more insights into these trade-offs and, in particular, under real-world surveillance scenarios so common.

Therefore, MTCNN and Haar Cascades Classifiers mainstream research confirms that framework flexibility, reliability, and efficiency when performing real-life projects. These models have not only developed face detection but have also expanded ideas about them and indicated a hint at the development of new models and solutions. The balance of this paper is based on this literature review where a comparative analysis of MTCNN and Haar Cascades models under different face detection environments is done.

Lately, there have been some big improvements in real-time face detection, such as YOLOFace and RMAGAN, which perform better than the conventional methods in real time but lack detailed robustness analysis on systematic yaw variations. This study focuses on these specific scenarios to understand Haar cascades and MTCNN practical performance under controlled conditions.

III. Method

This study aims to evaluate and compare the performance of two prominent face detection models, Multi-task Cascaded Convolutional Networks (MTCNN) and Haar Cascades classifiers, using a custom dataset known as UPM. The UPM dataset is composed of systematically created images of various yaw pose degrees and occlusions such as masks and glasses, which provides a comprehensive ground for analysis. The main aim of this newly created dataset is to testify to which degree each model can detect faces because face detection can be more challenging once the pose of a face gets larger. This section outlines the methods used for model evaluation, dataset generation, and performance comparison. Here is the workflow of the proposed methodology of the two models represented as a flowchart in Figure 4.

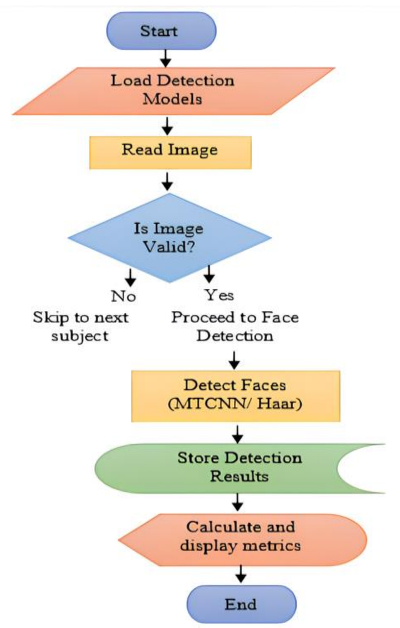

Fig. 4. Flowchart of the proposed models

The face detection process flow chart is shown above. The initial step involves loading two models: MTCNN and Haar Cascades. Images are read, validated, and for each subject, the MTCNN and Haar Cascades models are used to detect faces and extract their distinct features. The face coordinates are stored in a data frame resulting from the results. The detection performance of each subject is determined by returning accuracy, ROC curve, and confusion matrix after models are executed in batch mode.

A. Aim

The main objective of this research is to analyze the effectiveness of MTCNN and Haar Cascades classifiers on the UPM dataset. The emphasis is made on the fact that facial landmarks' localization is tested on yaw poses and occlusions of faces.

B. Dataset Description

The UPM dataset is unique, featuring 100 individuals comprising 54 males and 46 females from diverse ethnic backgrounds, capturing undergraduate and postgraduate students. The dataset was collected under a controlled environment where the lightning of face images was stable. Also, the dataset of face images does not include facial expressions. The images were captured by the embedded system's laboratory camera. The camera type was Canon, and the resolution of the captured images was 72 dpi. The images' dimensions are (3456*4608) pixels. UPM face dataset is composed of images with various yaw degrees, and those degrees were accurately set and measured using the engineering protractor under the supervision of the lab's instructors. Therefore, the UPM dataset consists of four subsets, each designed to test different face detection conditions. Images for the training and testing sets were selected randomly within each subset of the UPM dataset. The training and testing sets in this study were carefully constructed to evaluate the models under controlled and systematic variations. For each subset of the UPM dataset (e.g., yaw poses, masks, glasses, both), 13 images representing different yaw angles (-90° to +90°, at 15° intervals) were selected. From these, 10 images per subject were randomly chosen for training, ensuring a representative sample for each condition, and the remaining 3 images were reserved for testing. This approach balances training sufficiency with evaluation integrity.

- Set 1: Faces with varied yaw degrees (ranging from -90 to 90) without any masks or glasses as illustrated in Figure 5.

Fig. 5. Sample images of yaw poses without face accessories

- Set 2: Faces with the same yaw variation, covered with masks only.

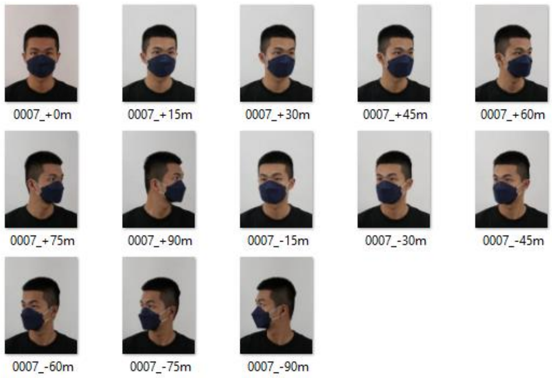

Fig. 6. Yaw poses with masks only

- Set 3: Faces with yaw variations, wearing glasses only.

Fig. 7. Yaw poses with glasses only

- Set 4: Faces subjected to yaw changes, obscured with both masks and glasses as represented in Figure 8.

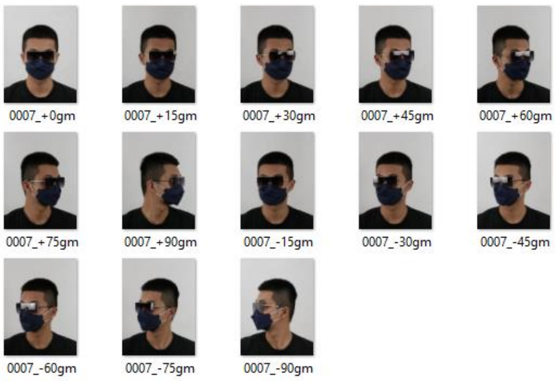

Fig. 8. Yaw poses with masks and glasses

Each subset is designed to challenge the detection capabilities of the models very thoroughly, especially with partial face visibility and varying orientations.

C. Face Detection Models

In this paper, face detection was applied via MTCNN and Haar Cascades classifiers. The deep learning mechanism behind MTCNN is used for accurate face detections, while traditional machine learning-based Haar Cascades classifiers are known to be fast and efficient.

C.1 MTCNN Model

The Multi-Task Cascaded Convolutional Network (MTCNN) is an instance of a deep learning model designed to do face detection, having the main feature of the structure of convolutional networks that are cascaded and give output results from the 1st to the final stage. The MTCNN operates through three primary stages: the proposal network (P-Net), the refine network (R-Net), and the output network (O-Net). Window and bounding box proposals are produced by the P-Net; the refiner is R-Net and filters false positives and produces final refined windows and landmarks as O-Net [31].

We thus discuss the factors that make MTCNN feasible for face detection with pose and occlusion variability. MTCNN is first trained on large amounts of face images with different direction angles as well as faces hidden in part. This full training of the network helps the network learn and then draw the complex facial features, making it more effective. Second, the MTCNN has a multiple-stage model so that the objects can be better detected based on the output of its previous stage. The whole face should be learned at every level of the proposed model, and hence, partial occlusion and angle change. Also, the facial landmarks and the face-bounding box are available, provided with the face-bounding box together by MTCNN. Since MTCNN recognizes facial landmarks, it can then align and flip faces, which is crucially important when facing up to the task of detecting faces because they are looking at the face and some parts of the face may be hidden [31] [32] [33]. Figure 9 clearly illustrates the network architecture of MTCNN.

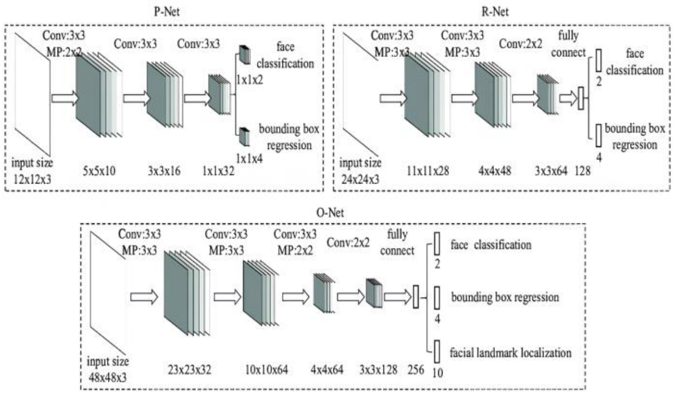

Fig. 9. MTCNN network architecture [34]

C.2 Haar Cascades Model

Haar Cascades, proposed by Viola and Jones, are widely used techniques for object detection, particularly for face detection in images and videos. The method operates based on machine learning to train a cascade of AdaBoost classifiers using Haar-like features, which are digital image features used for object detection. The process involves a cascade of stages, each consisting of a classifier trained with a set of positive and negative samples to determine whether a specific region of an image contains a face. At each stage, the classifier receives the image, evaluates the region, and either rejects it as non-face or passes it to the next stage [35].

One key advantage of Haar Cascades, compared to other techniques, is its ability to detect faces in real-time, which is particularly useful when time is critical. The approach utilizes the Integral Image to speed up the extraction of Haar-like features and AdaBoost for feature selection and classifier learning. This helps optimize resource usage and improves the efficiency of the face detection process within the cascade structure.

Haar cascades are very effective when the conditions defined are met but are a complete failure when the subject is at a different angle, partially obscured, or has different lighting. Nevertheless, for the faces at different poses or occluded, the method will not be effective since the Haar features defined will not comprise all the aspects of such faces. The structure of the classifiers is determined by the training data set, and there is no certainty that it will be effective for the new and quite different samples of variation in the direction of the face or occlusion. Hence, Haar cascades [35-37] can be applied efficiently for the simple function of face detection, while for the more complicated task with different poses and occlusions of the face, other algorithms or some preprocessing steps may be required.

D. Evaluation Procedure

The performance of these models is evaluated by first processing all images in the UPM dataset using both models, extracting detection results, and then comparing based on some parameters. The Python programming language is used to facilitate this implementation with CV2 and MTCNN libraries.

- Image Processing: This data set of images is preprocessed by resizing them to handle the model requirements. As the images are already in RGB format, there is no need for the MTCNN to perform conversion.
- Face Detection: The main evaluation step is the application of both face detection models to each image. Face detection using the different challenges of the UPM dataset is tested against the models and yields the coordinates of bounding boxes for the detected faces.
- Data Extraction and Compilation: Once detected, the bounding box coordinates are extracted from the data and organized as a structured format like a data frame for further analysis.

Finally, each model is compared by the accuracy and detection performance, and confusion matrices and ROC curves are utilized to visualize the results.

E. Performance Comparison

Following the data compilation, the study conducts a rigorous comparison analysis based on detection accuracy across all scenarios, where each scenario represents a set of UPM datasets.

- Scenario-based Analysis: This level of analysis involves evaluating model performance, which will be decided based on the accuracy rates of each specific scenario (set) within the UPM dataset. This step determines how each model fares in conditions of varying complexities introduced by different yaw angles and obstructions.

The UPM dataset is composed of four sets of data, where each set contains 13 face images of various yaw degrees. The first set contains face images with various yaw angle degrees without any occlusion accessories. The second set is also composed of 13 face images, and those images are covered with glasses only. The third set also contains 13 face images, but those images are covered with masks only, while the fourth set contains the same number of face images, but they are covered with both masks and glasses. Each set has been divided into two sets: the training set and the testing set. The training set contains 10 face images, and the testing set contains 3 face images. The images of the testing set were randomly selected from the training set.

F. Computational Environment

Experiments were performed on an Intel Core i7-1165G7 processor (11th generation, 2.80GHz) and 8 GB of RAM. With these setups, we have sufficient processing power for real-time analysis of face detection models. While we have enough memory available with this dataset, improving the performance by increasing available memory would also be useful when using larger datasets or more complex models.

IV. Experiments and Results

The comprehensive evaluation of MTCNN and Haar Cascades models applied to the UPM dataset, which incorporates various challenges in the form of yaw poses and facial occlusions, provides valuable insights into their performance. The results are categorized by tables and highlight the contrasting abilities of these models, offering guidance for their application in different real-world contexts. One subject will be randomly selected from each set of UPM datasets and then compared to check the two models' robustness in detecting faces at all yaw pose degrees. Then the overall accuracy of the 100 subjects will also be classified in a table. After that, the ROC curve and confusion matrix of that randomly selected subject of each set will also be illustrated.

Tables I–IV show that MTCNN has high detection accuracy as it is robust to occlusions and yaw pose variations. The observed extremes in accuracy (100% or 0%) for individual subjects and angles originate from the binary nature of detection. These results imply that MTCNN exhibits strong performance in face detection with all tested angles and conditions, while Haar Cascades fails for some excessive yaw poses and occlusions, resulting in poorer overall performance.

It focuses on systematic variations in yaw poses for which granular insights into each model's detection capability are provided, closing the gap of prior studies. This setup guarantees reproducibility and appraisal precision, hence the methodological rigor of the study.

- Set 1: Faces with varied yaw degrees (ranging from -90 to 90) without any masks or glasses.

Table I represents the accuracy results of subject eighty-nine using the two models. The (1s) mentioned in the table mean that faces were correctly and accurately detected, while the (0s) mean that faces were not detected. In Table I, the MTCNN model has precisely detected subject eighty-nine at all angles, while the Haar Cascades model did not detect subject eighty-nine at angle (90).

TABLE I
SUB. 89 ACCURACIES OF MTCNN AND HAAR (NO MASKS & NO GLASSES)

Degrees	MTCNN Detection	MTCNN Accuracy %	Degrees	Haar Detection	Haar Accuracy %
0	1	100	0	1	100
15	1	100	15	1	100
-15	1	100	-15	1	100
30	1	100	30	1	100
-30	1	100	-30	1	100
45	1	100	45	1	100
-45	1	100	-45	1	100
60	1	100	60	1	100
-60	1	100	-60	1	100
75	1	100	75	1	100
-75	1	100	-75	1	100
90	1	100	90	0	0
-90	1	100	-90	1	100

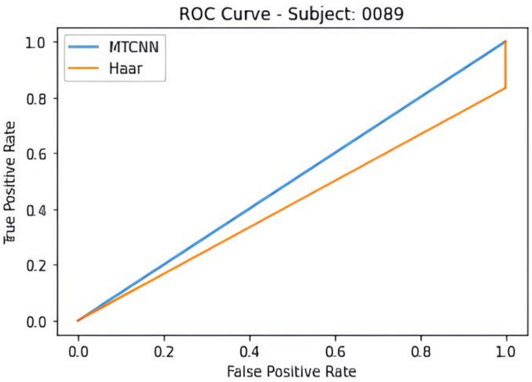

Fig. 10. ROC of subject 89 data from Table 1

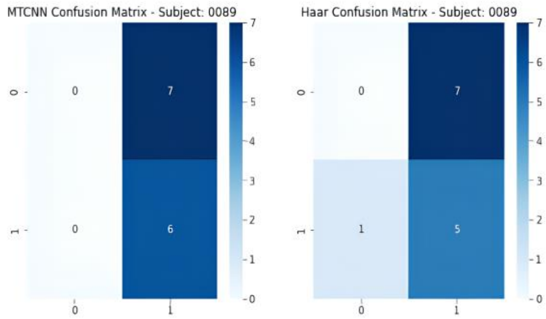

Fig. 11. Confusion matrix of both MTCNN and Haar Cascades from Table I

The results from Set 1 confirm our hypothesis that yaw poses have a large effect on detection accuracy. For example, as the yaw angle increases, decided features like eyes and mouth disappear to some extent, and performance degrades, especially in cases of less complex models, such as Haar cascades.

- Set 2: Faces occluded with glasses only at various yaw degrees (ranging from -90 to 90).

TABLE II
SUB. 49 ACCURACIES OF MTCNN AND HAAR (GLASSES ONLY)

Degrees	MTCNN Detection	MTCNN Accuracy %	Degrees	Haar Detection	Haar Accuracy %
0	1	100	0	1	100
15	1	100	15	1	100
-15	1	100	-15	1	100
30	1	100	30	1	100
-30	1	100	-30	1	100
45	1	100	45	1	100
-45	1	100	-45	1	100
60	1	100	60	1	100
-60	1	100	-60	1	100
75	1	100	75	0	0
-75	1	100	-75	0	0
90	1	100	90	1	100
-90	1	100	-90	0	0

Subject 49's accuracy scores when using the two models are shown in Table II. The table shows that the (1s) mean that faces covered with glasses were correctly and accurately spotted, while the (0)s mean that faces were not detected. Table II shows that the MTCNN model correctly identified subject 49 from all angles, but the Haar Cascades model failed to do so at angles (75, -75, -90).

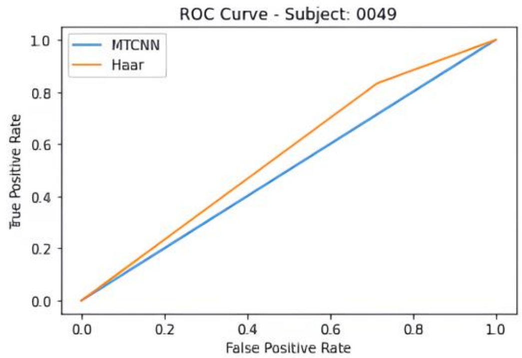

Fig. 12. ROC of subject 49 data from Table II

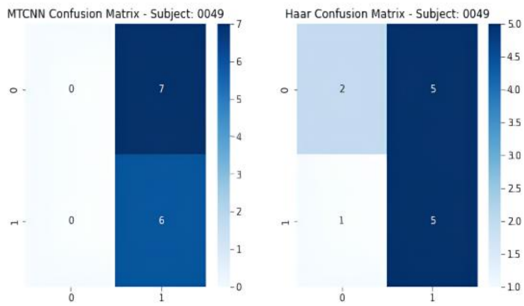

Fig. 13. Confusion matrix of both MTCNN and Haar Cascades from Table II

- Set 3: Faces occluded with masks only at various yaw degrees (ranging from -90 to 90).

With the introduction of masks as obstructions in this scenario, MTCNN's performance remains robust in face detection. Haar Cascades, while still exhibiting lower accuracy compared to

MTCNN, shows some adaptability to the presence of masks. Hence, Table III shows that MTCNN was able to detect faces of subject 98 at all yaw angle degrees, while Haar cascades could not handle detecting faces covered with masks at angles (60, 75, 90, and - 90).

TABLE III
SUB. 49 ACCURACIES OF MTCNN AND HAAR (MASKS ONLY)

Degrees	MTCNN Detection	MTCNN Accuracy %	Degrees	Haar Detection	Haar Accuracy %
0	1	100	0	1	100
15	1	100	15	1	100
-15	1	100	-15	1	100
30	1	100	30	1	100
-30	1	100	-30	1	100
45	1	100	45	1	100
-45	1	100	-45	1	100
60	1	100	60	0	0
-60	1	100	-60	1	100
75	1	100	75	0	0
-75	1	100	-75	1	100
90	1	100	90	0	0
-90	1	100	-90	0	0

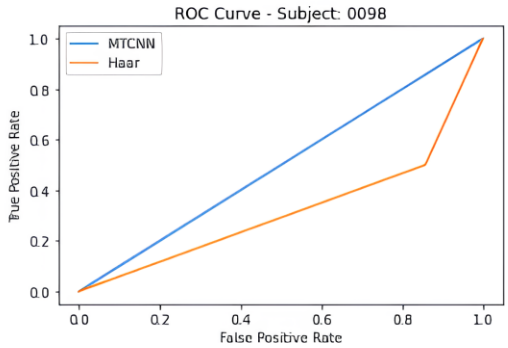

Fig. 14. ROC of subject 98 data from Table III

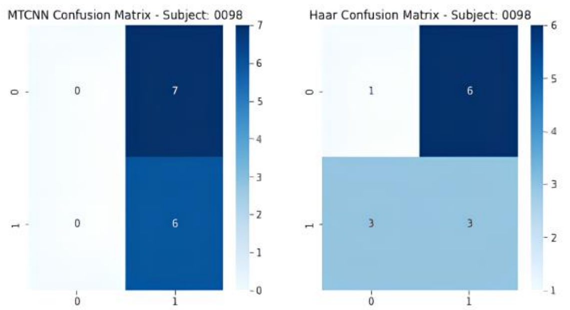

Fig. 15. Confusion matrix of both MTCNN and Haar Cascades from Table III

- Set 4: Faces occluded with both masks and glasses at different yaw degrees (ranging from -90 to 90).

In this challenging scenario with both masks and glasses as occlusions, MTCNN displays superior performance compared to Haar cascades. Haar cascades' accuracy decreases significantly, while MTCNN thrives in these complex conditions. The accuracy results can be seen in Table IV, where MTCNN was not detecting a yaw face image of subject 19 at angle (-30), while Haar cascades could not detect faces at angles (-45, 60, and 75). The ROC and confusion matrix of Table IV are also presented in Figures 15 and 16.

TABLE IV
SUB. 19 ACCURACIES OF MTCNN AND HAAR (MASKS AND GLASSES)

Degrees	MTCNN Detection	MTCNN Accuracy %	Degrees	Haar Detection	Haar Accuracy %
0	1	100	0	1	100
15	1	100	15	1	100
-15	1	100	-15	1	100
30	1	100	30	1	100
-30	0	0	-30	1	100
45	1	100	45	1	100
-45	1	100	-45	0	0
60	1	100	60	0	0
-60	1	100	-60	1	100
75	1	100	75	0	0
-75	1	100	-75	1	100
90	1	100	90	1	100
-90	1	100	-90	1	100

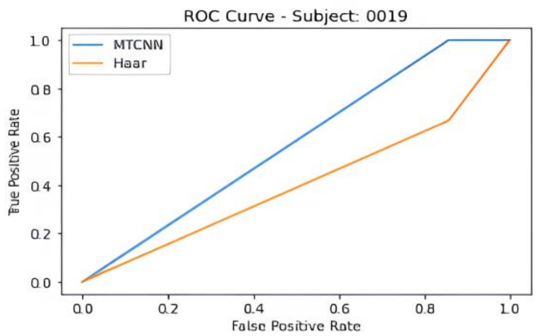

Fig. 16. ROC of subject 19 data from Table IV

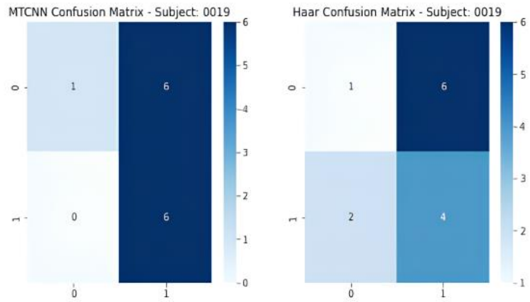

Fig. 17. Confusion matrix of both MTCNN and Haar Cascades from Table IV

The final overview of total accuracy displayed in Table V shows the two models, MTCNN and Haar Cascades, with regard to face detection in the four subsets of the UPM dataset. Each of the four subsets is designed in order to evaluate the abilities of the face detection models in various conditions concerning different angles and forms of occlusion. The presented table shows the accuracy of both models, which can show the user how well the model can solve the problem and where it fails as it has been compared with the other model on the similar probability that arises out of the data set.

TABLE V
THE OVERALL ACCURACY OF MTCNN AND HAAR OF EACH SUBSET OF UPM

Models	UPM Datasets	Overall Accuracy%
MTCNN	Set 1(yaw poses only)	100%
	Set 2 (yaw poses with masks)	99.9%
	Set 3 (yaw poses with glasses only)	96.4%
	Set 4 (yaw poses with masks and glasses)	80%
Haar Cascades	Set 1(yaw poses only)	92.5%
	Set 2 (yaw poses with masks)	67.3%
	Set 3 (yaw poses with glasses only)	80.4%
	Set 4 (yaw poses with masks and glasses)	76.3%

A. Computational Complexity and Speed Analysis

Table VI presents the average processing time per subject of 20 subjects for MTCNN and Haar Cascades models to show the speed trade-offs of each approach. In contrast, Haar Cascades tends to process subjects much faster than MTCNN, as Haar is simpler and has a deep learning structure.

TABLE VI
AVERAGE PROCESSING TIME PER SUBJECT FOR MTCNN AND HAAR CASCADES

Subject	MTCNN Avg Time (s)	Haar Avg Time (s)
0001	7.359769	1.155869
0002	7.507557	0.982754
0003	7.470676	0.947931
0004	7.761935	1.258538
0005	7.623581	1.293862
0006	8.270880	1.207304
0007	8.438117	1.176684
0008	7.881345	1.394028
0009	8.338813	1.187892
0010	8.282511	1.212107
0011	8.471939	1.143962
0012	8.200059	1.096631
0013	8.019938	1.412541
0014	8.527875	1.395329
0015	8.199158	1.028784
0016	7.825409	1.352500
0017	7.403527	1.171280
0018	7.415334	1.063909
0019	7.382514	0.911707
0020	7.783880	1.623783

This computational efficiency is shown in Table VI in comparison with MTCNN, which is significantly faster. Nowadays, although modern methods like YOLOV5 make decisions faster than before, they consume more computational resources and thus can't be used in environments that lack resources.

V. DISCUSSION

Therefore, the findings of this paper, which utilize both MTCNN and Haar Cascades for face detection on the new UPM dataset—containing systematic variations in yaw poses and occlusion—provide valuable insights into their performance and applicability. From the results tables, it is evident that MTCNN outperforms Haar Cascades [38] in detecting faces under all tested conditions. More precisely, MTCNN demonstrated near-perfect accuracy in detecting faces across yaw poses ranging from -90 to 90 degrees, including those partially occluded by a mask, glasses, or both. Due to this high reliability, MTCNN emerges as the preferable model for applications demanding greater accuracy in complex and dynamic environments.

Haar Cascades offer the advantage of speed and are quick to implement; however, under challenging conditions, they exhibit significantly lower performance than MTCNN [29], [30]. Therefore, the choice of model for a given problem involves a trade-off between speed and the need for high accuracy. Further studies might explore the possibility of combining the advantages of both methods, such as integrating Haar Cascades—which learn faster than MTCNN and other deep learning algorithms—into more precise detection techniques.

When it comes to surveillance and security, MTCNN provides the ability to identify faces even in scenarios with extreme occlusion and angles when integrated into real-time applications such as monitoring. This paper has successfully developed a model that enhances human-computer interaction systems, particularly by enabling users to interact with the system in various settings that may impose facial constraints, such as when wearing personal protective gear.

Haar Cascades and MTCNN demonstrate reliability primarily in systematic, controlled scenarios, whereas YOLO-Face and RMAGAN excel in real-time, dynamic environments. In applications with constrained computational resources or specialized requirements, such insights are particularly valuable.

Although Haar Cascades and MTCNN are considered traditional, they remain highly useful for applications requiring low computational cost or those involving systematic pose variation. MTCNN delivers superior performance but at the expense of speed; for real-world applications with increasing complexity, Haar Cascades offer the fastest detection but are best suited for simpler tasks.

VI. LIMITATIONS OF THE STUDY

However, while this study yields useful measurements of the performance of MTCNN and Haar Cascade models in different yaw angles and occlusion conditions, several do deserve to be mentioned. The analysis was then performed on the UPM face dataset alone, a dataset that is robust concerning systematic angle variations and occlusions but might not cover the possible diversity of real-world cases, e.g., for different lighting conditions, backgrounds, and dynamic expressions. Thus, testing further on other datasets with different environmental factors would allow a more comprehensive examination of model performance as observed in different environments than the current one.

Second, while the MTCNN model worked well, it is computationally resource-intensive and has reached a limit on speed, making it less suitable for real-time scenarios in resource-constrained environments. Haar Cascades are fast in nature but may not be the best option where high accuracy is required in challenging conditions. The overall approach could be extended to future research by considering hybrid approaches that combine the strengths of both models in a computationally efficient way.

Lastly, this paper mostly centers on variations in yaw angle, with occlusions taken to include only masks and glasses. Further validation of the model's robustness in different conditions could be achieved by this analysis being expanded to cover other types of occlusions (e.g., hats, scarves, facial hair) and other facial poses (pitch and roll).

VII. Conclusion and Future Work

A. *Summary of Findings and Model Performance*

This paper assesses the performance of two face detection algorithms, MTCNN and Haar Cascades, on the UPM dataset and concludes that MTCNN outperforms Haar Cascades in all tested conditions. In the first UPM scenario (faces with different yaw poses and no occlusion), MTCNN achieved an accuracy of 100%, while Haar Cascades came in second with 92%. With masked faces, MTCNN reached 99.9% accuracy, whereas Haar Cascades dropped to 67.3%. For faces with glasses, MTCNN achieved 96%, compared to Haar Cascades at 80.4%. In the final scenario (faces with both masks and glasses), MTCNN showed an accuracy of 80%, with Haar Cascades at 76.3%.

B. *Theoretical and Practical Implications*

In face detection, as with other methods, this study shows that MTCNN is more precise than other methods, regardless if the subject’s face is rotated or not fully occluded. This is due to the fact that MTCNN has greater accuracy when compared to Haar Cascades and is well suited to surveillance applications where precise face identification under tough conditions is required.

C. *Research Contributions and Practical Advantages*

The main contribution of this study lies in enhancing human-computer interaction systems, particularly in environments where users may wear personal protective equipment. Additionally, the results underscore the relationship between detection speed and accuracy, a key consideration in selecting face detection models. While Haar Cascades provides faster processing, MTCNN’s superior accuracy makes it preferable for applications that prioritize precision.

D. *Research Limitations and Future Directions*

More occlusions and lighting conditions could be included in future research to investigate further the model's stability. Furthermore, MTCNN can be combined with other deep learning techniques for higher detection efficiency and effectiveness. In order to build on this study, datasets should incorporate yaw, pitch, and roll poses in order to more fully evaluate face detection models.

References

[1] S. S. Khan, D. Sengupta, A. Ghosh, and A. Chaudhuri, "MTCNN++: A CNN-based face detection algorithm inspired by MTCNN," The Visual Computer, pp. 1-19, 2023, doi: 10.1007/s00371-023-02822-0.

[2] L. Zhou, H. Zhao, and J. Leng, "MTCNet: Multi-task collaboration network for rotation-invariance face detection," Pattern Recognition, vol. 124, p. 108425, 2022, doi: 10.1016/j.patcog.2021.108425.

[3] S. Bhattacharya, M. Ghosh, and A. Dey, "Face detection in unconstrained environments using modified multitask cascade convolutional neural network," Proceedings of the International Conference on Industrial Instrumentation and Control (ICI2C 2021), pp. 287-295, 2022, doi: 10.1007/978-981-16-7011-4_29.

[4] C. H. Choi, J. Kim, J. Hyun, Y. Kim, and B. Moon, "Face detection using Haar cascade classifiers based on vertical component calibration," Human-centric Computing and Information Sciences, vol. 12, no. 11, pp. 1-17, 2022.

[5] J. M. Shamrat, F. M., A. Majumder, P. R. Antu, S. K. Barmon, I. Nowrin, and R. Ranjan, "Human face recognition applying Haar cascade classifier," Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021, pp. 143-157, Springer Singapore, 2022, doi: 10.1007/978-981-16-5640-8_12.

[6] V. Dores, "Identifikasi masker pada face detection dengan menggunakan metode Haar Cascade dan CNN," Jurnal Sistim Informasi dan Teknologi, vol. 4, no. 4, pp. 149-154, 2022, doi: 10.37034/jsisfotek.v4i4.154.

[7] K. A. Majeed, Z. Abbas, M. Bakhtyar, Z. Durrani, J. Baber, and I. Ullah, "Face detectors evaluation to select the fastest among DLIB, HAAR Cascade, and MTCNN," Pakistan Journal of Emerging Science and Technologies, vol. 2, no. 1, pp. 1-13, 2021.

[8] M. H. Robin, M. M. U. Rahman, A. M. Taief, and Q. N. Eity, "Improvement of face and eye detection performance by using multi-task cascaded convolutional networks," 2020 IEEE Region 10 Symposium (TENSYMP), pp. 977-980, June 2020, doi: 10.1109/TENSYMP50017.2020.9230681.

[9] M. Gao and Y. Tang, "Rotation face detection based on three-window convolutional neural networks," Journal of Physics: Conference Series, vol. 1229, no. 1, p. 012013, May 2019, doi: 10.1088/1742-6596/1229/1/012013.

[10] E. Kaziakhmedov, K. Kireev, G. Melnikov, M. Pautov, and A. Petiushko, "Real-world attack on MTCNN face detection system," 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 422-427, Oct. 2019, doi: 10.1109/SIBIRCON48586.2019.8958122.

[11] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on MTCNN," 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), pp. 154-158, Sept. 2020, doi: 10.1109/ICCNEA50255.2020.00040.

[12] M. Gu, X. Liu, and J. Feng, "Classroom face detection algorithm based on improved MTCNN," Signal, Image and Video Processing, vol. 16, no. 5, pp. 1355-1362, 2022, doi: 10.1007/s11760-021-02087-x.

[13] A. Ghofrani, R. M. Toroghi, and S. Ghanbari, "Realtime face-detection and emotion recognition using MTCNN and MiniShuffleNet V2," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), pp. 817-821, Feb. 2019, doi: 10.1109/KBEI.2019.8734924.

[14] J. Xiang and G. Zhu, "Joint face detection and facial expression recognition with MTCNN," 2017 4th International Conference on Information Science and Control Engineering (ICISCE), pp. 424-427, July 2017, doi: 10.1109/ICISCE.2017.95.

[15] C. Wu and Y. Zhang, "MTCNN and FaceNet based access control system for face detection and recognition," Automatic Control and Computer Sciences, vol. 55, pp. 102-112, 2021, doi: 10.3103/S0146411621010090.

[16] M. Ma and J. Wang, "Multi-view face detection and landmark localization based on MTCNN," 2018 Chinese Automation Congress (CAC), Xi'an, China, Nov. 2018, pp. 4200-4205, doi: 10.1109/CAC.2018.8623535.

[17] A. Priadana and M. Habibi, "Face detection using Haar cascades to filter selfie face images on Instagram," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, Mar. 2019, pp. 6-9, doi: 10.1109/ICAIIT.2019.8834526.

[18] C. L., Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," 2017 13th IEEE International Conference on Electronic

Measurement & Instruments (ICEMI), Yangzhou, China, Oct. 2017, pp. 483-487, doi: 10.1109/ICEMI.2017.8265863.

[19] I. Gangopadhyay, A. Chatterjee, and I. Das, "Face detection and expression recognition using Haar cascade classifier and Fisherface algorithm," Recent Trends in Signal and Image Processing: Proceedings of ISSIP 2018, Springer Singapore, 2019, pp. 1-11, doi: 10.1007/978-981-13-6783-0_1.

[20] D. T. P. Hapsari, C. G. Berliana, P. Winda, and M. A. Soeleman, "Face detection using Haar cascade in different illumination," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, Indonesia, Sep. 2018, pp. 555-559.

[21] A. Singh, H. Herunde, and F. Furtado, "Modified Haar-cascade model for face detection issues," International Journal of Research in Industrial Engineering, vol. 9, no. 2, pp. 143-171, 2020.

[22] R. Padilla, C. F. F. Costa Filho, and M. G. F. Costa, "Evaluation of Haar cascade classifiers designed for face detection," World Academy of Science, Engineering and Technology, vol. 64, pp. 362-365, 2012.

[23] A. Sharifara, M. S. Mohd Rahim, and Y. Anisi, "A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection," 2014 International Symposium on Biometrics and Security Technologies (ISBAST), Kuala Lumpur, Malaysia, Aug. 2014, pp. 73-78, doi: 10.1109/ISBAST. 2014.7013097.

[24] T. Mantoro and M. A. Ayu, "Multi-faces recognition process using Haar cascades and eigenface methods," 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, Morocco, May 2018, pp. 1-5, doi: 10.1109/ICMCS.2018.8525935.

[25] Z. Han, W. Song, X. Yang, and Z. Ou, "Face pose estimation with ensemble multi-scale representations," Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition, Beijing, China, Aug. 2019, pp. 97-101, doi: 10.1145/3357254.3357278.

[26] Z. Yu et al., "Yolo-facev2: A scale and occlusion aware face detector," Pattern Recognition, vol. 155, p. 110714, 2024, doi: 10.1016/j.patcog.2024.110714.

[27] V. Gaikwad et al., "Design and implementation of IoT-based face detection and recognition," Computing & Intelligent Systems, 2024, pp. 923-933, doi: 10.56155/978-81-955020-2-8-78.

[28] J. Liao, T. Guha, and V. Sanchez, "Self-supervised random mask attention GAN in tackling pose-invariant face recognition," Pattern Recognition, vol. 155, p. 111112, 2024, doi: 10.1016/j.patcog.2024.111112.

[29] N. Singh and R. M. Brisilla, "Comparison analysis of different face detecting techniques," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Chennai, India, Nov. 2021, pp. 1-6, doi: 10.1109/i-PACT52855.2021.9696583.

[30] S. A. Hashmi, "Face detection in extreme conditions: A machine-learning approach," arXiv preprint arXiv:2201.06220, 2022.

[31] L. Zhou, H. Zhao, and J. Leng, "MTCNet: Multi-task collaboration network for rotation-invariance face detection," Pattern Recognition, vol. 124, p. 108425, 2022, doi: 10.1016/j.patcog.2021.108425.

[32] M. T. Islam et al., "Convolutional neural network-based partial face detection," 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), Pune, India, Apr. 2022, pp. 1-6, doi: 10.1109/I2CT54291.2022.9825259.

[33] O. Naser, S. Ahmad, K. Samsudin, S. Shafie, and N. Zamri, "Facial recognition for partially occluded faces," Indonesian Journal of Electrical Engineering and Computer Science, vol. 30, no. 3, pp. 1846-1855, 2023, doi: 10.11591/ijeecs.v30.i3.pp1846-1855.

[34] S. Li, Y. Dou, J. Xu, K. Yang, and R. Li, "GBCNN: A full GPU-based batch multi-task cascaded convolutional networks," IEEE Access, vol. 7, pp. 20225-20234, 2019, doi: 10.1109/ACCESS.2019.2894589.

[35] C. Rahmad et al., "Comparison of Viola-Jones Haar cascade classifier and histogram of oriented gradients (HOG) for face detection," IOP Conference Series: Materials Science and Engineering, vol. 732, no. 1, p. 012038, 2020, doi: 10.1088/1757-899X/732/1/012038.

[36] C. H. Choi et al., "Face detection using Haar cascade classifiers based on vertical component calibration," Human-centric Computing and Information Sciences, vol. 12, no. 11, pp. 1-17, 2022.

[37] A. Singh, H. Herunde, and F. Furtado, "Modified Haar-cascade model for face detection issues," International Journal of Research in Industrial Engineering, vol. 9, no. 2, pp. 143-171, 2020.

[38] O. A. Naser, S. M. Syed Ahmad, K. Samsudin, and M. Hanafi, "Enhancing 2D face recognition systems: Addressing yaw poses and occlusions with masks, glasses, and both," Advances in Artificial Intelligence and Machine Learning, vol. 4, no. 3, pp. 2545-2574, 2024.

Omer Abdulhaleem Naser is a final-year Ph.D. student at the University Putra Malaysia, Faculty of Engineering. Majored in computational methods in engineering. The specified research field is biometrics, particularly, facial recognition for occluded faces. Currently, Omer is working as a lecturer at the University of Information Technology and Communications, Department of Electronic Computer Center. Omer can be contacted via these emails: omar.abdulhalem592@gmail.com & omer.naser@uoitc.edu.iq.

Sharifah Mumtazah Syed Ahmad (Assoc. Prof. Dr.) is an associate professor who currently works as a senior lecturer and supervisor at the University Putra Malaysia, Faculty of Engineering. Dr. Sharifah is specialized in many different areas, including biometrics, security, image and signal processing, and machine learning. She earned her PhD in electronics in 2004 from the University of Kent, United Kingdom. Currently, she has over 50 cited publications with a Scopus index of 12. s_mumtazah@upm.edu.my.

Khairulmizam b. Samsudin (Dr.) is a Dr. who currently works as a senior lecturer and supervisor at the University Putra Malaysia, Faculty of Engineering. Dr. Khairulmizam's areas of specialties include embedded systems, robotics, and intelligent systems. Dr. Khairulmizam can be contacted via this email: khairulmizam@upm.edu.my.

Marsyita Hanafi is a Dr. who currently works as a senior lecturer and supervisor at the University Putra Malaysia, Faculty of Engineering. Dr. Marsyita specializes in many different areas, including precision agriculture, smart monitoring systems, biometrics, security, image processing, and machine learning. Dr. Marsyita can be contacted via this email: marsyita@upm.edu.my.

Siti Mariam Shafie received her M.Sc. and B.Eng. degrees in the Faculty of Engineering from Universiti Putra Malaysia in 2003 and 1999, respectively. She is currently working at the same university in the Department of Computer and Communication Systems Engineering. Her research interests include embedded systems, IoT, and machine learning. She can be contacted at email: mariam@upm.edu.my.

Ts. Nor Zarina Zamri was appointed as Research Lead for the Malaysian Security Evaluation Facility Department under CyberSecurity Malaysia (CSM MySEF). Background in Bachelor (BCs Hons) in Computer Science (Software Engineering) and Master of Science (MSc) in Information Technology. She specializes in smart card, biometric, and hardware security and leads teams in cyber security research exploration, explores new technologies, and produces in-house test tools, test methods, guidelines, training, etc. Nor Zarina can be contacted via this email: norzarina.zamri@cybersecurity.my.