

# Enhancing touchless smart locker systems through advanced facial recognition technology: a convolutional neural network model approach

Abdul Haris Rangkuti, Evawaty Tanuar, Febriant Yapson, Felix Octavio Sijoatmodjo,  
Varyl Hasbi Athala

Department of Computer Science, School of Computers Science, Bina Nusantara University, Bandung, Indonesia

## Article Info

### Article history:

Received May 7, 2024

Revised Mar 29, 2025

Accepted Jun 8, 2025

### Keywords:

Facial

Locker

Multi-task cascaded  
convolutional networks

RetinaFace

VGG-face

## ABSTRACT

As the world recovers from COVID-19, demand for contactless systems is increasing, promising safety and convenience. Touchless technology, particularly public locker security systems that use facial recognition and hand detection, is advancing rapidly. The system minimizes physical contact, increasing user safety. It uses advanced models such as multi-task cascaded convolutional networks (MTCNN) and RetinaFace, FaceNet512, ArcFace, and visual geometry group (VGG)-Face for face detection and recognition, with a combination of RetinaFace, ArcFace, and L2 norm Euclidean or cosine as the most effective distance metric method, where the accuracy reaches 96 and 90%. 'Yourvault', an application demonstrating this efficient security feature, provides notifications for mask detection, facial authenticity and locker status, offering a solution to the problem of convenience and security of public spaces. Future research could investigate the impact of photo age on facial recognition accuracy, potentially making touchless systems more efficient. In general, the application of this technology is an important step towards a safer and more comfortable world after the pandemic. This model approach can be followed up with more optimal facial recognition.

This is an open access article under the [CC BY-SA](#) license.



## Corresponding Author:

Abdul Haris Rangkuti

Department of Computer Science, School of Computers Science, Bina Nusantara University

Pasirkaliki St. No. 25-27, Paskal Hyper Square Bandung 40181, West Java, Indonesia

Email: rangku2000@binus.ac.id

## 1. INTRODUCTION

The adopting artificial intelligence as a system for controlling internet of things (IoT) devices has been a less favored alternative due to its comparatively time-intensive nature when juxtaposed with biometric-based security measures like fingerprint recognition [1]. However, since the onset of the COVID-19 pandemic, a paradigm shift has prompted a significant transformation of systems that traditionally required physical contact into operations devoid of such physical interaction [2]. Voice and image-based control and security systems have experienced rapid proliferation during this period, driven by heightened demand. Consequently, the imperative for developing integrated security systems within the IoT framework, bolstered by artificial intelligence, has become paramount, aiming to supersede technologies still reliant on physical touch [3]. One system that still necessitates physical touch is the locking mechanism employed in public lockers. An alternative to replace physical keys involves using passcodes for locker access [3]. However, entering a passcode still requires physical touch and remains susceptible to breaches, rendering it suboptimal as an alternative. Another proposed option is using radio frequency identification

(RFID) cards, such as the electronic identity card (E-KTP), as a substitute for physical keys [4]. While this alternative efficiently reduces physical contact, the security of RFID cards is still susceptible to duplication, thus failing to achieve optimal security levels. Another alternative, which obviates the need for physical touch while maintaining robust security, is quick response (QR)-based security [5]. QR codes minimize physical contact to the utmost extent, ensuring maintained security, albeit necessitating additional steps for operation. Consequently, ongoing development in public locker locking systems is imperative to ensure secure, convenient usage with minimal physical touchpoints. In this era, technological progress is swiftly advancing, propelled by the pervasive integration of artificial intelligence across multifarious domains, most notably on the IoT realm. The infusion of artificial intelligence into IoT systems begets a paradigm shift, facilitating the seamless and efficient processing of real-time data [6]. Beyond this, incorporating artificial intelligence in the IoT milieu augments data security within the system and orchestrates a judicious optimization of power utilization [7]. Consequently, the orchestration of IoT systems is poised for refinement, centering its focus on data acquisition by utilizing integrated, energy-efficient platforms such as Arduino, espressif systems (ESP), and Raspberry [8].

The proposed methodological approach in this research involves developing a face recognition-based locking system with hand detection as its navigational system. In a previous study on creating a face recognition-based locker system [9], the navigation system and user interface (UI) still necessitated physical contact. As an additional security measure, liveliness detection research [10] has been explored to prevent the use of photographs to access lockers. However, this study relies solely on blink detection, leaving it vulnerable to exploitation through video-based approaches. Incorporating additional security features is crucial, especially considering reports wherein an elementary school student breached a face recognition-based package locker using only a photograph [11]. Therefore, the methodological approach under development in this research is imperative to enhance locker security and construct a system that eliminates the need for physical touch in its locking mechanism. The locker system developed in this research comprises two main components: the locker system and the Yourvault booth for locker access. The Yourvault booth is constructed using face recognition from the LightFace framework, Mask and Fake face detection utilizing you only look once version 8 (YOLOv8), and hand detection using Mediapipe. The locker system is built with ESP8266, serving as the central control unit for locker keys. The booth system can connect with multiple lockers, provided the lockers and booths are in the exact same registered location. Firebase firestore and firebase real-time database govern the ecosystem of Yourvault.

## 2. LITERATURE REVIEW

A comprehensive investigation into image processing, characterized by precision and relevance, becomes imperative to implement a touchless paradigm within locker security systems. The touchless navigation framework under examination in this research is grounded in the nuanced realm of hand detection. This selection stems from its inherent attributes of efficiency, user intuitiveness, and a remarkable ability to cater to individuals facing communicative challenges [12]. To augment hand detection capabilities, this study integrates Mediapipe technology, renowned for its commendable average accuracy of up to 99%, thereby fortifying the foundation of reliability [13]. Furthermore, this research endeavors to elevate the touchless locker experience by introducing a sophisticated face recognition system as the linchpin for securing these storage units. The framework of choice for this purpose is the illustrious LightFace, meticulously curated for its nimble character, versatile support for face recognition models, proficient metric distance computations, and adept face detection prowess [14].

As an additional layer of security, this research integrates a mask detection and facial authenticity system using YOLOv8. The utilization of YOLOv8 in discerning mask usage exhibits the capability to accurately classify individuals correctly wearing masks, those with improper mask placement, and those not wearing masks, achieving high precision and accuracy [15]. Mask usage detection enhances face recognition performance by comprehensively removing the user's mask when necessary. Employing YOLOv8 for facial authenticity detection aims to mitigate the risk of unauthorized data usage through media such as photos and videos. Mask usage detection enhances face recognition performance by comprehensively removing the user's mask when necessary. Employing YOLOv8 for facial authenticity detection aims to mitigate the risk of unauthorized data usage through media such as photos and videos. An overview of research related to facial recognition techniques, which is a security key technique for doing something can be seen in Table 1.

The models utilized in this research encompass FaceNet512, ArcFace, and VGG-Face, coupled with various distance metrics available in the LightFace framework, such as cosine, Euclidean, and Euclidean L2 norm. Drawing on previous research findings, FaceNet512 exhibits an accuracy rate of up to 97%, whereas ArcFace demonstrates accuracy levels reaching 88% [16]. However, a report from the Singapore defense science and technology agency regarding the resilience of face recognition methods against noise suggests that VGG-Face boasts the highest accuracy rate, reaching 82% compared to FaceNet and ArcFace [17]. Hence, this

study strategically incorporates all three models to ascertain the methodology that attains high accuracy and demonstrates resilience against noise, aligning with the characteristics of the camera devices in use.

Table 1. Comparison analysis of previous methods

References	Condition	Methods	Accuracy
[16]	Normal	FaceNet512	97.4%
		ArcFace	87.8%
[17]	Noise 0%, 25%, 50%, 75% and 100%	FaceNet512	< 60%
		ArcFace	80.4%
		VGG-Face	81.6%
[18]	Labeled faces in the wild (LFW) and celebrities in frontal-profile in the wild (CFP-FP)	MTCNN+ArcFace	LFW: 99.83%
			CFP-FP: 98.37%
		RetinaFace+ArcFace	LFW: 99.86%
			CFP-FP: 99.49%
[19]	LFW	MTCNN	Easy: 85.1%
			Hard: 60.7%
		RetinaFace	Easy: 87.8%
			Hard: 47.3%
[20]	Testing using 5 different datasets	ArcFace	Lowest value:
		FaceNet512	97.4%
		VGG-Face	85.45%
			93.95%
[21]	Testing on 200 face samples	MTCNN+ArcFace	70-90%
		RetinaFace+ArcFace	80-100%

Other studies describe the position of the classroom surveillance camera as not fixed. So that the angle of the face recorded through the surveillance video is different. Thus, the deep learning-based facial verification model has achieved sufficient and controlled results [22]. In general, there are about 30-40% of genetic disorders associated with certain facial characteristics called dysmorphic features. In another study by analyzing the performance of classifiers based on deep learning facial recognition models through dysmorphic feature detection [23]. The face detection alignment methods utilized in this study are multi-task cascaded convolutional networks (MTCNN) and RetinaFace. These methods differ in the face detection stage, with MTCNN utilizing a three-stage face detection approach, while RetinaFace adopts a single-stage approach employing a single convolutional network [24]. This dissimilarity has implications for the results, as RetinaFace demonstrates superior speed and detection capabilities, particularly in low-light conditions, compared to MTCNN [18]. Additionally, RetinaFace excels in detecting faces in various poses, including tilted poses, and exhibits robust performance across different scales [25]. However, it is noteworthy that MTCNN also performs admirably, offering more accurate and stable face detection than RetinaFace [19]. Consequently, this research incorporates MTCNN and RetinaFace for face detection alignment to determine the method that achieves high accuracy while operating efficiently and swiftly.

### 3. PROPOSED METHOD

This research This research centers around the application of facial data hailing from a database encompassing 50 students. The assembly of this facial data was a pivotal element of the research, realized through the use of diverse devices throughout the study's progression. The Samsung Galaxy J5 Prime smartphone camera, tethered to a laptop via the DroidCam application, functioned as the principal instrument for data gathering. The experimental procedure was bifurcated into two distinct segments. The initial segment entailed the recognition of 30 pre-registered faces. Subsequently, the second phase zeroed in on the detection of nine unique faces, each exhibited in five diverse facial positions. The system's performance was primarily evaluated by measuring the false negative rate. This metric reflects instances where individuals were correctly discerned, yet the quantified distance value surpassed the established minimum threshold. Threshold is derived by pinpointing the smallest distance between erroneously detected individuals. In conducting multiple experiments with various faces in numerous positions, we hope to facilitate a comprehensive analysis of the system's robustness. Concurrently, the system's effectiveness will be reinforced based on its proficiency in accurately identifying individuals, accounting for variations in distinct facial poses. An additional aim is to verify the reliability of the distance accuracy metrics employed throughout the evaluation process. Therefore, to bolster accuracy performance, we implement several similarity methods and a convolutional neural network (CNN) model system. These tools are going to aid in attaining a clear understanding of optimal performance in the detection of human faces. For this reason, an overview of the stages of the training data retrieval process in the locker lending system is provided.

Figure 1 provides a depiction of the operational mechanism of the locker lending system. It demonstrates the asynchronous synchronization and training process of several models dedicated to the detection and recognition of facial images stored in the database. The system only initiates data synchronization when modifications are made at the time of the last synchronization of the facial database. Changes to the synchronization time can transpire due to various reasons, such as the addition or input of a new user's face, the process of deleting a user, or alterations to a user's image that is considered unclear. If a modification is executed at the time of the last synchronization, the system will extract photos from the database and train the model using these images. Generally, the LightFace framework conducts the training process independently and performs it exclusively during the synchronization process. Following the model's formation through this training process, the system will initiate a transition process to the primary display menu. In the face detection system carried out by the locker, the system will continue to carry out the process optimally as the number of users involved in storing goods or securities in the locker increases.

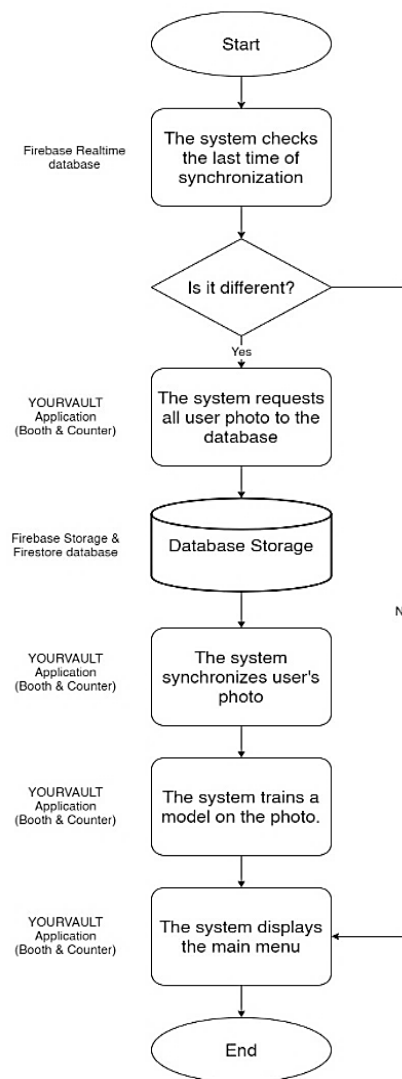


Figure 1. The synchronization mechanism of the locker borrowing system

Figure 2 visually depicts the operational interface of the Yourvault locker rental application. The system is ingeniously engineered to automatically synchronize data whenever modifications occur, always providing users with the most current information. The application's navigation is seamlessly facilitated using hand detection technology, offering a user-friendly experience. As part of the process, users can access different menu options by simply raising their hand. When a transaction is initiated at a Yourvault kiosk, a facial verification process is put into action to validate the user's identity. Provided the user's face is

accurately recognized and they possess adequate funds for locker rental, the system proceeds to reveal the locker that is either currently borrowed or in use. After this, the system unlocks the designated locker while appropriately deducting the necessary funds from the user's account. However, should the system fail to confirm the user's identity during the facial verification stage, the transaction is instantly annulled. The interface then reverts to its main menu. This procedure is implemented to bolster transactional security and prevent unauthorized locker access.

Figure 3 depicts the functionality of a navigation system employing hand detection technology for facial recognition. Within the main menu, users are prompted to raise either their right or left hand, with distinct functions attributed to each hand. Specifically, the left hand signifies the process of borrowing a locker, while the right hand denotes accessing the locker. Upon reaching the locker opening menu, users are presented with two clear options: to either proceed with the locker rental process or to terminate the locker rental. Users can make their choice by aligning their hand selection (right or left) with the corresponding instructions displayed in the application. This user-friendly option system simplifies the decision-making process for users, whether they wish to continue renting a locker or cease the rental process. A visual representation of this process is provided in the accompanying chart.

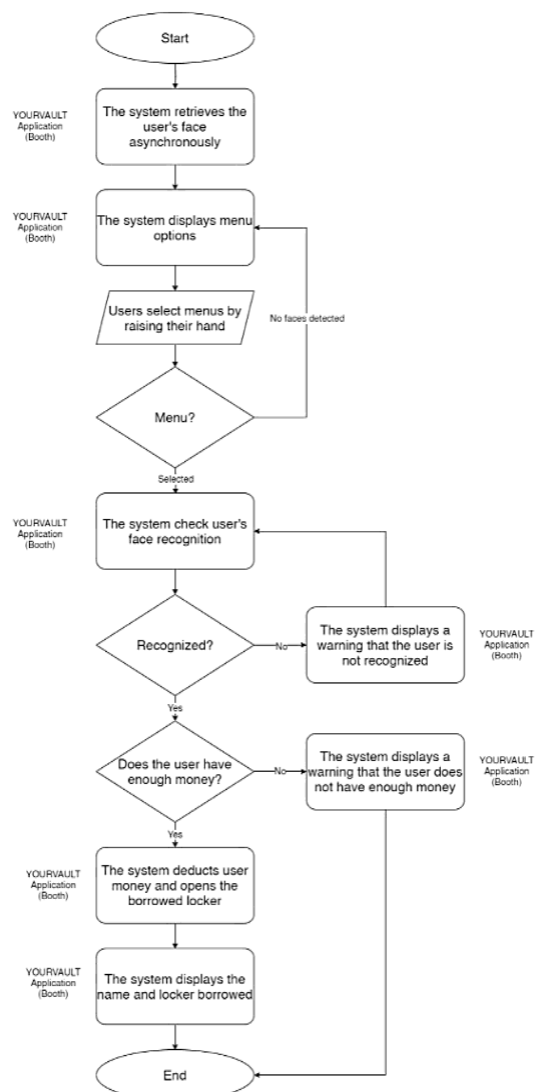


Figure 2. The operational interface of the Yourvault booth application

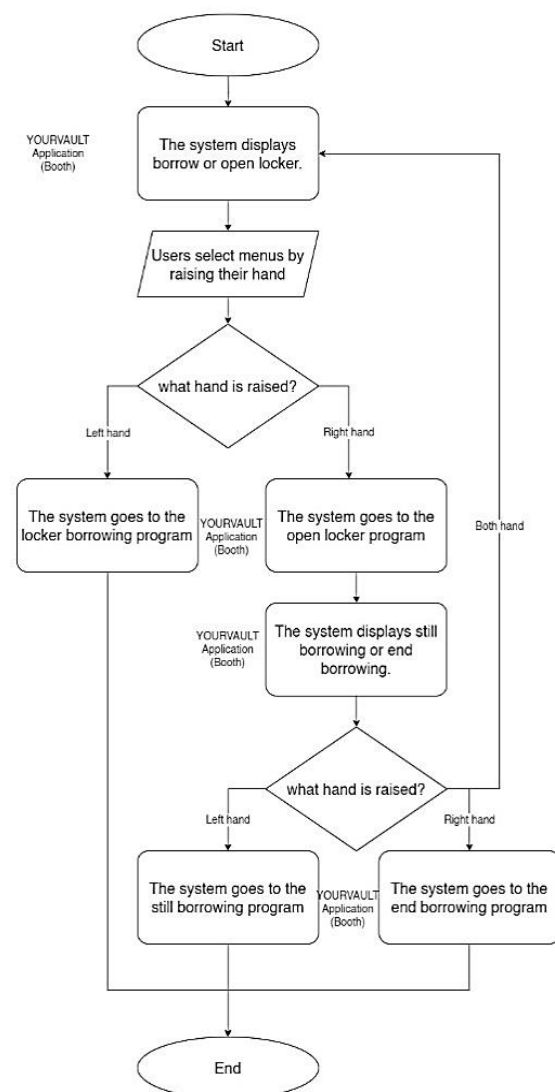


Figure 3. The functionality of the navigation system utilizing hand detection

Figure 4 showcases the pre-face recognition process, which is meticulously designed to ensure that only one face is detected within the camera frame. Following this, the system proceeds to verify that the

user's face is not obstructed by any form of covering or mask. This crucial step serves to enhance security measures, effectively preventing the use of spoofing methods. Furthermore, the process includes a validation mechanism to authenticate the genuineness of the detected face, distinguishing it from printed or displayed images. This stringent validation ensures that the facial image captured by the camera is authentic and not a reproduction from a different medium. In the final stage, the system validates the user's face, ensuring it is accurately detected within a specified distance threshold. Upon meeting all these stringent criteria, the system successfully detects the user's face and proceeds to return the user ID for further processing, enabling recognition of the tenant's face stored within the user face database.

Figure 5 provides an overview of the operation of the smart locker system, designed for user and customer rental use. Typically, smart lockers display a green light when available and a red light when in use. However, if the locker light is off, it signifies that the locker is unlocked and has been opened by the user to deposit or retrieve items. Subsequently, the lights will re-illuminate once the locker is closed, automatically securing the locker. This seamless process of opening and closing the locker furnishes users with pertinent information if they wish to avail themselves of a smart locker. The intelligent locker system adeptly recognizes the faces of its tenants with precision, swiftly relaying information to the locker system to facilitate the opening and closing processes. An overview of the dataset used in the experiment can be found at: <https://www.kaggle.com/datasets/yohanes07/yourvault-face-recognition-datasets>. In essence, this research presents an innovative approach to utilizing lockers based on facial recognition and associated protocols.

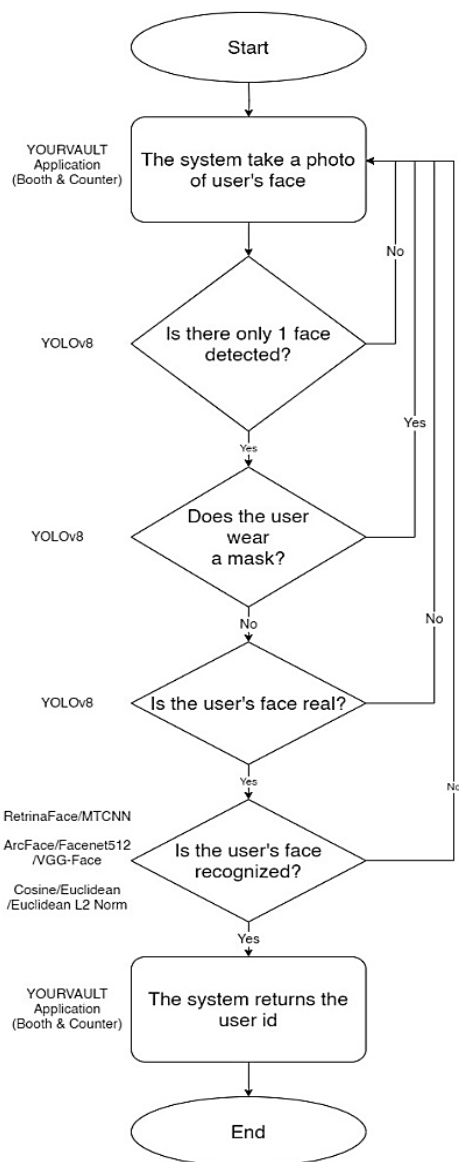


Figure 4. The pre-facial recognition process

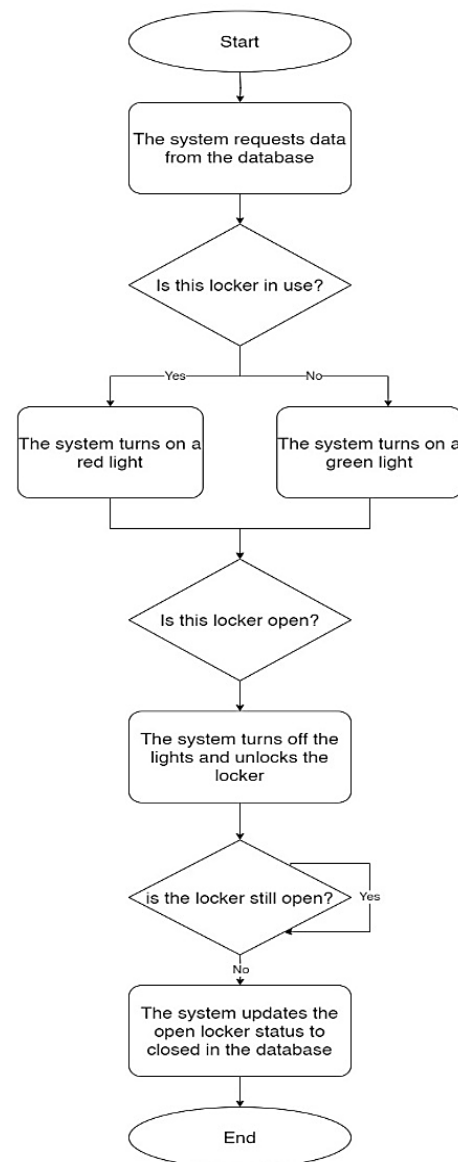


Figure 5. The operation of the Yourvault locker system

#### 4. RESULTS AND DISCUSSION

The research encompassed two distinct experiments aimed at identifying the method with the highest and most consistent accuracy. The first experiment entailed the recognition of 30 out of 50 pre-registered faces, all positioned facing the front. Subsequently, the second experiment focused on discerning the facial features of 9 individuals across five different facial orientations: front, top, bottom, and slanting left and right at 45 degrees. Following the completion of both experiments, a thorough evaluation process was initiated, during which the top 5 methods, distinguished by the highest number of correct recognitions, underwent a meticulous thresholding elimination procedure. This process meticulously defined the parameters for differentiating between true positive and false positive detections, effectively eliminating any instances of false positives. Consequently, a comprehensive analysis was conducted to determine the false negative rate and accuracy metrics for these aforementioned top 5 methods.

Based on the findings presented in Table 2, the experiment's primary aim was to recognize 30 out of 50 registered faces with a frontal orientation, utilizing the MTCNN as the face detection algorithm. The results indicate that the ArcFace model demonstrates optimal performance, particularly when employing cosine and Euclidean L2 norm, achieving an accuracy of more than 96%. Conversely, when utilizing FaceNet512 performance, the most effective performance is achieved with Euclidean and Euclidean L2 norms, yielding an accuracy of 70%. Notably, the performance of VGG-Face is inferior to both ArcFace and FaceNet512. Furthermore, it is evident that the Euclidean L2 norm consistently delivers a more stable performance compared to both cosine and Euclidean L2 norms when evaluating the identified facial features.

Table 2. Facial recognition experiments for 30 people using MTCNN

Model	Distance metric	Percentage		
		Correct (%)	Wrong (%)	Invalid (%)
ArcFace	Cosine	96.7	3.3	0
	Euclidean	63.3	6.7	30
	Euclidean L2	96.7	3.3	0
FaceNet512	Cosine	66.7	13.3	20
	Euclidean	70	30	0
	Euclidean L2	70	30	0
VGG-Face	Cosine	56.7	43.3	0
	Euclidean	53.3	46.7	0
	Euclidean L2	56.7	43.3	0

Experiment findings based on Table 3 aimed to recognize 30 faces out of 50 registered faces with a frontal orientation, employing RetinaFace as the face detection algorithm. The table illustrates that the ArcFace model yields optimal performance when utilizing both cosine and Euclidean L2 norm, achieving an optimal accuracy of 90%, albeit with fewer correct recognitions than when MTCNN is used. In the case of FaceNet512 performance, the best result is achieved with Euclidean and Euclidean L2 norm, with an accuracy of more than 76%, accompanied by a higher number of correct recognitions than MTCNN. Notably, VGG-Face exhibits inferior performance compared to both ArcFace and FaceNet512. Stability of Euclidean L2 norm with RetinaFace, Furthermore, it is evident that the Euclidean L2 Norm with RetinaFace consistently provides a more stable performance compared to both cosine and Euclidean L2 norms when assessing the identified facial features.

Table 3. The result of facial recognition experiments for 30 people using RetinaFace

Model	Distance metric	Percentage		
		Correct (%)	Wrong (%)	Invalid (%)
ArcFace	Cosine	90	10	0
	Euclidean	60	3.3	36.7
	Euclidean L2	90	10	0
FaceNet512	Cosine	73.3	16.7	10
	Euclidean	76.7	23.3	0
	Euclidean L2	76.7	23.3	0
VGG-Face	Cosine	56.7	43.3	0
	Euclidean	56.7	43.3	0
	Euclidean L2	56.7	43.3	0

Based on the data presented in Table 4, the experiment focused on recognizing the faces of 9 individuals across five different facial positions, utilizing MTCNN as the face detection algorithm. The

results indicate that the ArcFace model exhibits superior performance when employing both cosine and Euclidean L2 norm with optimal accuracy of more than 73%. Conversely, for FaceNet512, the best performance is achieved with Euclidean, while the least effective performance, relative to other methods, is observed when using cosine and Euclidean L2 norm. Surprisingly, in this experiment, VGG-Face demonstrates a performance level equivalent to FaceNet512. These findings underscore the superior performance of ArcFace, particularly when employing cosine and Euclidean L2 norm in the context of recognizing faces across varied positions, while also highlighting the competitive performance of VGG-Face compared to FaceNet512.

Table 4. The result of facial recognition experiments for 5 different pose using MTCNN

Model	Distance metric	Percentage		
		Correct (%)	Wrong (%)	Invalid (%)
ArcFace	Cosine	73.3	20	6.7
	Euclidean	33.3	22.2	44.4
	Euclidean L2	73.3	17.8	8.9
FaceNet512	Cosine	40	26.7	33.3
	Euclidean	57.8	42.2	0
	Euclidean L2	48.9	48.9	2.2
VGG-Face	Cosine	57.8	37.8	4.4
	Euclidean	55.6	44.4	0
	Euclidean L2	57.8	37.8	4.4

Based on the data presented in Table 5, the experiment aimed to recognize the faces of 9 individuals across five different facial positions, utilizing RetinaFace as the face detection algorithm. The results indicate that the ArcFace model demonstrates optimal performance when utilizing both cosine and Euclidean L2 norm with accuracy more than 66%, albeit with fewer correct recognitions than when MTCNN is used. For FaceNet512, the best performance is achieved with Euclidean L2 Norm, accompanied by a higher number of correct recognitions than MTCNN. However, it is noteworthy that VGG-Face exhibits the poorest performance compared to ArcFace and FaceNet512 and falls short of MTCNN. These results emphasize the nuanced performance trade-offs across different algorithms, with ArcFace excelling in specific metrics but potentially exhibiting lower absolute recognition numbers. Meanwhile, FaceNet512 displays competitive metrics, particularly concerning cosine and Euclidean, and VGG-Face lags behind its counterparts.

Table 5. The result of facial recognition experiments for 5 different pose using RetinaFace

Model	Distance metric	Percentage		
		Correct (%)	Wrong (%)	Invalid (%)
ArcFace	Cosine	68.9	28.9	2.2
	Euclidean	37.8	20	42.2
	Euclidean L2	66.7	24.4	8.9
FaceNet512	Cosine	57.8	26.7	15.6
	Euclidean	57.8	42.2	0
	Euclidean L2	66.7	33.3	0
VGG-Face	Cosine	51.1	48.9	0
	Euclidean	53.3	46.7	0
	Euclidean L2	51.1	48.9	0

Based on Table 6, the 5 models with the highest accuracy are determined based on the total number of correct facial image recognitions from several experiments. An invalid value, considered a failure in capturing and recognizing a face, is excluded from the recognition error calculation. Despite the process of recognizing facial images based on its accuracy in several experiments, errors in facial recognition persist. Therefore, the elimination of recognition errors necessitates the establishment of a threshold for the resulting distance value. This process is crucial to prevent errors in facial recognition. The method utilizing the Euclidean L2 norm and the cosine method for image measurements demonstrates that as the distance approaches zero, the accuracy of face recognition increases. Consequently, the threshold determination is based on the smallest value among the false recognition distances, rounded to two decimal places, ensuring a zero number of false positives. Subsequently, true negatives are measured by determining whether the distance value for correct recognition exceeds a set threshold. A value is considered true positive if the correct recognition distance is less than the specified threshold. Following this, the false negative rate and accuracy are calculated, aiding in identifying the most effective method. These findings highlight the intricate



trade-offs in performance across different algorithms, with each method exhibiting distinct strengths and weaknesses in the context of facial recognition accuracy.

Table 6. Top 5 methods with the highest number of correct recognitions

Alignment	Model	Distance metric	Correct	Wrong
MTCNN	ArcFace	Euclidean L2	62	9
MTCNN	ArcFace	Cosine	62	10
RetinaFace	ArcFace	Cosine	58	16
RetinaFace	ArcFace	Euclidean L2	57	14
RetinaFace	Facenet512	Euclidean L2	49	26

Table 7 shows that RetinaFace, utilizing the ArcFace model with the cosine distance metric, achieves a significantly higher level of accuracy with a relatively fast average recognition time. As an alternative, with a slight variance in accuracy, RetinaFace, employing the ArcFace model with the Euclidean L2 norm distance metric, succeeds in reducing the number of incorrect recognitions, while maintaining a comparable accuracy level and a relatively similar average recognition time. In conclusion, when the Euclidean L2 norm is used as the distance metric, this experiment demonstrates better stability due to a lower number of false negatives compared to using cosine. Consequently, this experiment determines that RetinaFace with the ArcFace model utilizing the Euclidean L2 norm distance metric is the optimal choice. Not only does it provide sufficient accuracy, but it also results in lower false negatives, contributing to enhanced stability and consistency in facial recognition.

Table 7. Top 5 methods with the lowest false negative rate values and highest accuracy after limiting on the distance until no false positives

Alignment	Model	Distance metric	Distance limiter	True positive	True negative	False negative	False negative rate	Accuracy	Time (seconds)
RetinaFace	ArcFace	Cosine	0.40	34	24	16	0.41379	0.676	1.473
RetinaFace	ArcFace	Euclidean L2	0.89	33	24	14	0.42105	0.662	1.482
MTCNN	ArcFace	Euclidean L2	0.91	30	32	9	0.51613	0.549	2.494
MTCNN	ArcFace	Cosine	0.41	29	33	10	0.53226	0.542	2.518
RetinaFace	FaceNet512	Euclidean L2	0.56	5	44	26	0.89796	0.413	1.430

An overview of the UI display for the box borrowing using the recognition method in this research can be seen in Figure 6. The UI dashboard in Figure 6 presents the primary interface for locker borrowing within the Yourvault application. In the first menu, users are prompted to select by raising their left hand to borrow a locker or their right hand to open a borrowed locker. As users raise their hands, a countdown animation confirms the menu selection. Upon choosing to open a locker, users are directed to the second menu, which offers the choice to either open a currently borrowed locker or end the locker borrowing session. The navigation system in the second menu also involves hand gestures. Once users have selected either borrowing or opening a locker, the application displays a UI guiding users to position their faces for effective recognition. At the bottom of the interface, notifications inform users about using masks, non-authentic facial features, detection of more than one face, and guidance to move closer or farther from the camera. Upon successful facial detection, the system captures an image and identifies the user's face ID. If the user is recognized and possesses sufficient tokens, the application displays the user's name and the number of the borrowed locker. This interface automatically changes to the first menu once users securely close the borrowed locker door.

The UI display of the Yourvault application in Figure 7 represents notification messages to the user. If the user lacks tokens, the successfully recognized user's name and a notification prompting them to replenish their tokens are shown. In case of an error in facial recognition, the system issues a notification advising the user to recheck their facial positioning. Users are informed if they have yet to borrow a locker when attempting to open one or if they have already borrowed a locker when attempting to borrow another, as each user is limited to borrowing only one locker at a given location. Additionally, the system alerts the user if the lockers are occupied. During database synchronization, the system is temporarily unavailable, and a notification is displayed, informing users that the system is undergoing data synchronization. This notification manages user expectations and provides transparency regarding the application's status during synchronization processes.



Figure 6. UI display of Yourvault booth application



Figure 7. Alert or notification display UI of Yourvault booth application

## 5. CONCLUSION

Based on the findings of this research, the implementation of face recognition with hand detection as its navigation system in the locker borrowing system has proven to be successful in providing efficient and rapid security while minimizing physical touch during the locker locking process. The experimental results demonstrate that the utilization of RetinaFace with the ArcFace model and applying the Euclidean L2 norm distance metric result in high and stable accuracy levels. The Yourvault application exhibits commendable facial recognition capabilities, even amidst variations in facial poses. Furthermore, incorporating mask detection and facial authenticity verification using YOLOv8 adds a layer of security. Consequently, developing Yourvault touchless locker security system is expected to be a practical solution, replacing conventional methods involving physical touch. This system holds potential for implementation in public locker settings, offering heightened security levels and optimal user convenience. For further research, there is an expectation for system optimization to achieve faster speeds while maintaining robust user facial data security. Additionally, in-depth research is warranted, involving collecting more varied data, including diverse age groups, ethnicities, and other unique characteristics.

## FUNDING INFORMATION

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abdul Haris Rangkuti	✓	✓		✓	✓	✓			✓	✓	✓	✓	✓	
Evawaty Tanuar				✓	✓	✓				✓	✓	✓		
Febriant Yapson	✓		✓		✓				✓		✓			
Felix Octavio Sijoatmodjo	✓		✓	✓					✓		✓			
Varyl Hasbi Athala		✓	✓			✓				✓	✓			

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review &amp; Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at <https://www.kaggle.com/datasets/yohanes07/yourvault-face-recognition-datasets>.




## REFERENCES

- [1] R. Tanadi and R. Yusuf, "Face recognition implementation on public spaces usability and performance review," in *2020 6th International Conference on Interactive Digital Media*, 2020, pp. 1–5, doi: 10.1109/ICIDM51048.2020.9339602.
- [2] M. Z. Iqbal and A. G. Campbell, "From luxury to necessity: Progress of touchless interaction technology," *Technology in Society*, vol. 67, 2021, doi: 10.1016/j.techsoc.2021.101796.
- [3] H. F. Alqahtani *et al.*, "Automated smart locker for college," in *2020 3rd International Conference on Computer Applications & Information Security*, 2020, pp. 1–6, doi: 10.1109/ICCAIS48893.2020.9096868.
- [4] A. Pramono, M. A. Febriantono, I. A. Agustina, I. B. A. Wijaya, T. I. W. Primadani, and S. A. Budiman, "Manufacturing a smart locker security system for public spaces using E-KTP as a primary access," in *2022 International Conference on ICT for Smart Society*, 2022, pp. 1–6, doi: 10.1109/ICISS55894.2022.9915146.
- [5] J. Sa-ngiampak *et al.*, "LockerSwarm: an IoT-based smart locker system with access sharing," in *2019 IEEE International Smart Cities Conference*, 2019, pp. 587–592, doi: 10.1109/ISC246665.2019.9071664.
- [6] B. Chander, S. Pal, D. De, and R. Buyya, "Artificial intelligence-based internet of things for industry 5.0," *Internet of Things*, pp. 3–45, 2022, doi: 10.1007/978-3-030-87059-1\_1.
- [7] P.-J. Lin and C.-T. Ho, "Smart lock security system based on artificial internet of things," in *2020 IEEE Eurasia Conference on IOT, Communication and Engineering*, 2020, pp. 79–81, doi: 10.1109/ECICE50847.2020.9302010.
- [8] V. G. Menon, S. Jacob, P. Sehdev, M. R. Khosravi, and F. Al-Turjman, "An IoT-enabled intelligent automobile system for smart cities," *Internet of Things*, vol. 18, 2022, doi: 10.1016/j.iot.2020.100213.
- [9] W. Lin and S. Hu, "Design and implementation of an offline face recognition locker," *Journal of Physics: Conference Series*, vol. 1634, no. 1, 2020, doi: 10.1088/1742-6596/1634/1/012131.
- [10] P. Hadke, M. Khandagale, A. Pawar, and V. Rakh, "Face and liveliness detection based smart bank locker," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 175–178, 2022, doi: 10.48175/ijarsct-2635.
- [11] X. Yujie, "Facial-recognition smart lockers hacked by fourth-graders," *Sixth Tone*. 2019. Accessed: Oct. 17, 2019. [Online]. Available: <https://www.sixthtone.com/news/1004698>
- [12] A. Gupta, R. Narula, P. Garg, D. Joshi, and N. Upadhyay, "Touchless operations using hand gestures detection," *Recent Innovations in Wireless Network Security*, vol. 5, no. 3, pp. 31–40, 2023.
- [13] K. M. Kavana and N. R. Suma, "Recognition of hand gestures using mediapipe hands," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 6, pp. 4149–4156, 2022.
- [14] S. I. Serengil and A. Ozpinar, "LightFace: a hybrid deep face recognition framework," in *2020 Innovations in Intelligent Systems and Applications Conference*, 2020, pp. 1–5, doi: 10.1109/ASYU50717.2020.9259802.
- [15] S. Tamang, B. Sen, A. Pradhan, K. Sharma, and V. K. Singh, "Enhancing COVID-19 safety: exploring YOLOv8 object detection for accurate face mask classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 892–897, 2023.
- [16] A. Firmansyah, T. F. Kusumasari, and E. N. Alam, "Comparison of face recognition accuracy of ArcFace, Facenet and Facenet512 models on deepface framework," in *2023 International Conference on Computer Science, Information Technology and Engineering*, 2023, pp. 535–539, doi: 10.1109/ICCoSITE57641.2023.10127799.
- [17] T. E. Min, D. Lee, J. Han, and A. Tan, "Robustness of facial recognition to noise," *Wayback Machine*, pp. 1–9, 2023, [Online]. Available: [https://web.archive.org/web/20230420033224/https://www.dst.gov.sg/ydsp/projects/files/reports/ROBUSTNESS\\_OF\\_FACIAL\\_RECOGNITION\\_TO\\_NOISE.pdf](https://web.archive.org/web/20230420033224/https://www.dst.gov.sg/ydsp/projects/files/reports/ROBUSTNESS_OF_FACIAL_RECOGNITION_TO_NOISE.pdf)
- [18] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5202–5211, doi: 10.1109/CVPR42600.2020.00525.
- [19] A. Ghimire, N. Werghi, S. Javed, and J. Dias, "Real-time face recognition system," *arXiv-Computer Science*, pp. 1–2, 2022.
- [20] B. Kanawade *et al.*, "Automated human recognition in surveillance systems: an ensemble learning approach for enhanced face recognition," *Ingenierie des Systemes d'Information*, vol. 28, no. 4, pp. 877–885, 2023, doi: 10.18280/isi.280409.
- [21] R. A. Asmara *et al.*, "Face recognition using ArcFace and FaceNet in Google Cloud platform for attendance system mobile application," in *Proceedings of the 2022 Annual Technology, Applied Science and Engineering Conference*, Dordrecht: Atlantis Press International, 2022, pp. 134–144, doi: 10.2991/978-94-6463-106-7\_13.




- [22] N. Li *et al.*, “Chinese face dataset for face recognition in an uncontrolled classroom environment,” *IEEE Access*, vol. 11, pp. 86963–86976, 2023, doi: 10.1109/ACCESS.2023.3302919.
- [23] M. Geremek and K. Szklanny, “Deep learning-based analysis of face images as a screening tool for genetic syndromes,” *Sensors*, vol. 21, no. 19, 2021, doi: 10.3390/s21196595.
- [24] M. Jin, “A study of face alignment methods in unmasked and masked face recognition,” *M.Sc. Thesis*, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2023.
- [25] O. Yakovleva, A. Kovtunenkov, V. Liubchenko, V. Honcharenko, and O. Kobylin, “Face detection for video surveillance-based security system,” in *COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems*, 2023, pp. 69–86.

## BIOGRAPHIES OF AUTHORS






**Abdul Haris Rangkuti**    is a lecturer and a researcher at the School of Computer Science in Bina Nusantara University. He received the B.S. degree in management informatics and the M.Sc. and Ph.D. degree in computer science both from Institut Pertanian Bogor and University of Gadjah Mada in 1996, 2006 and 2020. His research interests are multimedia processing (image, audio, video processing), computer vision, and medical disease recognition. He involved in international project such as from ADB and World Bank as analyst system and team leader. He ever carried out some project from department/minister as analyst system or team leader. He can be contacted at email: rangku2000@binus.ac.id.






**Evawaty Tanuar**    was born in the city of Jambi and earned a bachelor's degree in computer science in 2006. In 2008, she continued her studies in IT strategic planning at the University of Wollongong, Australia. Her career has always been in the field of education. After completing her further studies, she joined the Department of Computer Science at Bina Nusantara University as a course coordinator. She also had the opportunity to assist at the Kalbi's Institute. Now she is trusted to serve as the chair of the Department of Computer Science at Bina Nusantara University, Bandung Campus. She can be contacted at email: etanuar@binus.edu.






**Febriant Yapson**    is an undergraduate student at the Department of Computer Science at Bina Nusantara University, expected to graduate in 2025. He is pursuing a bachelor's degree in computer science with a keen interest in computer vision and the internet of things. Throughout his studies, he has engaged in multiple projects and research activities, developing skills in system analysis and effective teamwork. He can be contacted at email: febriant.yapson@binus.ac.id.



**Felix Octavio Sijoatmodjo**    is an undergraduate student at the Department of Computer Science at Bina Nusantara University and is expected to graduate in 2025. He is currently pursuing a bachelor's degree in computer science. His academic research interests include computer vision and the internet of things. He has participated in several projects and research initiatives, gaining experience in system analysis and teamwork. He can be contacted at email: felix.sijoatmodjo@binus.ac.id.



**Varyl Hasbi Athala**    was born in Pontianak. He attended Bina Nusantara University for his honours, Bachelor of Science in computer science which completed in 2022. He has been working as a researcher in Bina Nusantara University since 2020. In the past, he has published many international papers and conferences with the topic of image classification, image retrieval, and object detection. Currently, his research interests include machine learning and image processing. He can be contacted at email: varyl.athala@binus.ac.id.