Environment, Society and Business Review (ESBR) Vol. 1, No. 1 (February 2026) | pp. 132-157
ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

SMARTPAYFC: FACE RECOGNITION-BASED VIRTUAL ON-SITE PURCHASING SYSTEM

Rabaa Alabdulrahman [1*], Shubashini Rathina Velu[1]

[1] College of Business Administration, Prince Mohammad bin Fahd University, Dhahran, 34754, Saudi Arabia
*Corresponding author: ralabdulrahman1@pmu.edu.sa

Article type: Original Research
Received: 06 October 2025 | Accepted: 28 December 2025 | Published online: 16 March 2026

Abstract

As technology advances, consumer shopping experiences have evolved from traditional payment methods to self-serve point-of-sale systems. Today, many businesses have migrated towards self-service and E-Wallet payments, which have proved effective during the COVID-19 pandemic. Furthermore, adopting these systems aligns well with sustainability and the United Nations' Sustainable Development Goals (SDGS). Con-tributing to the economic, social, and environmental SDGs. Although this has provided significant business potential, flaws still exist within the ecosystem. The digital transformation has become a new platform for unethical cyberattacks and reconnaissance. In addition to security concerns, today's shopping experience has evolved to help reduce disease transmission. The COVID-19 pandemic left a lasting imprint on how we perceive and engage with digital technologies. Despite the countermeasures in place, the need to contact items in the store remains inevitable. To address these issues, the author proposes a virtual on-site shopping system that promotes efficient, contactless payment while enhancing transactional security for consumers. The proposed SmartPayFC system allows consumers to add items to their virtual shopping cart by scanning the QR code provided in a catalogue, followed by facial recognition authentication for enhanced security. Additionally, the proposed system will offer a unified view of sales figures and relevant statistical data to support business decision-makers. Our system achieved an accuracy of 74% in similarity for positive match images and a probability of about 0.003% for different images.

Keywords: *Face Recognition, Contactless Payments, Neural Network, Sustainable In-novation..*

1. Introduction

The recent surge in non-cash payment methods has revolutionized financial transactions, promoted financial inclusion and enhancing the efficiency of monetary systems. However, these methods are becoming increasingly vulnerable to exploitation, posing risks to economic stability and user security. In 2020, the US Federal Trade Commission documented 4.72 million instances of credit card theft, resulting in economic losses totalling 3.3 billion USD (Federal Trade, 2022). Similarly, Saudi Arabia faced fraud-related losses of approximately 60 million SAR, underscoring the global scale of this issue (Ali et al., 2019). Despite advancements in cashless payment systems, critical challenges persist, such as (a) security vulnerabilities in traditional methods (e.g., RFID cloning, QR phishing) enabling financial fraud that costs billions annually (Liébana-Cabanillas et al., 2024); (b) physical interaction during checkout heightening health risks, as evidenced during COVID-19 (Li et al., 2024); and (c) the exclusion of unbanked populations reliant on cash transactions, which limits financial inclusion (Demirgüç-Kunt et al., 2022). These issues underscore the urgent need for a secure, contactless, and inclusive payment ecosystem.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

To address these challenges, many organizations have adopted innovative systems that leverage artificial intelligence (AI) to detect fraudulent activities in e-commerce environments. Machine learning algorithms, a cornerstone of AI, analyze patterns in credit card usage behaviour, strengthening the security framework of digital transactions. The proliferation of global online payment networks such as Alipay, PayPal, and Apple Pay, which cater to millions worldwide, highlights the potential for increased cybersecurity risks. Vulnerabilities such as the duplication of RFID-equipped credit cards or other cyberattacks emphasize the need for robust systems to safeguard financial transactions. Financial institutions are exploring sustainable innovations like biometric authentication to counter these threats, ensuring resilient digital infrastructure (Pelechrinis et al., 2023). The global COVID-19 pandemic accelerated the adoption of contactless payment technologies, a shift observed across many nations, including Saudi Arabia (Shishah & Alhelaly, 2021). This transition reduced reliance on physical currency, mitigating health risks associated with the virus and promoting the development of safer, contact-free shopping experiences.

Despite these advancements, segments of the population reliant on cash transactions face inefficiencies such as lengthy wait times and limited access to digital payment systems. Promoting inclusive access to digital payments is crucial for reducing inequalities and enhancing economic participation globally. The global shift toward digitized payments has not fully addressed systemic flaws. For instance, Amazon Go is an example of an automated retail environment that has shown vulnerability to in-app QR code scanning, leading to potential unauthorized deductions (Han et al.). Similarly, other QR-based platforms, such as Alipay, remain susceptible to ongoing cybersecurity threats, including spoofed codes and phishing attacks. Moreover, a significant portion of the population remains financially excluded. For instance, it has been reported that approximately 26% of Saudi Arabia's population remains unbanked, posing a barrier to the full implementation of digital financial services in the country (Demirgüç-Kunt et al., 2022). A solution that combines biometric security, minimizes physical contact, and accommodates cash-dependent users is imperative.

Our research focuses on creating innovative and sustainable solutions for digital payment systems. By integrating advanced facial recognition technology, we aim to enhance financial security and user privacy while promoting responsible consumption and production practices. The specific objectives are:

- Enhance user security during payments, reducing the need to enter sensitive information in public.
- Minimize physical interaction to align with health and safety standards.
- Develop a virtual on-site shopping system to enhance efficiency and accessibility.

This research aligns with the broader vision of sustainable economic growth, fostering inclusive and innovative technological solutions to advance global financial ecosystems.

The primary contribution of this research is the development of an end-to-end virtual on-site purchasing architecture designed to bridge the gap between physical retail and digital security. The specific contributions of this work are as follows:

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

- **Novel System Architecture:** Integration of a QR-based virtual cart with biometric-triggered fulfillment to eliminate physical touchpoints.
- **Enhanced Security Design:** Implementation of a dual-layered authentication framework combining facial biometrics with PIN-based transaction authorization to mitigate common e-wallet vulnerabilities.
- **Optimized Face Verification Model:** Development of a Siamese-based CNN utilizing Batch Hard Triplet Mining and PCA-driven dimensionality reduction to achieve high verification accuracy with reduced computational complexity.
- **Empirical Performance Evaluation:** A comprehensive assessment of model performance using both masked and unmasked facial datasets to ensure reliability in post-pandemic retail environments.

2. Overview

2.1 Reluctance In Adopting Mobile Payment Systems

Recent technological advances have raised concerns about the security of modern payment methods. Humbani and Wiese (2018) identified insecurity as a factor hindering the adoption of mobile-based payments. They argue that users' reluctance to embrace mobile payments stems from their fear of having sensitive information stolen and the potential consequences, such as unlawful financial transactions (Humbani & Wiese, 2018). In addition to security, a study conducted by Khanra et al. (2021) found that privacy concerns related to vulnerabilities in the network hinder the acceptance of mobile payment services, causing users to take preventive measures, such as providing false information during registration. Furthermore, customers may lack confidence in the reliability and credibility of mobile payment systems, potentially increasing hesitancy in adopting such technology (Mallat, 2007).

2.2 Security And Privacy Issues of Existing Systems

Numerous studies have underscored the security and privacy concerns linked to existing cashless payment systems. These concerns mainly focus on the potential for data breaches, identity theft, and unauthorized access to sensitive financial information (Alkhateeb & Maolood, 2019; Sakharova). Furthermore, these worries are heightened by the perceived lack of security and trust among consumers in these payment systems, as noted by. Lin & Wang found that a significant percentage of users abandon mobile payment due to concerns about privacy leaks (Wang & Li, 2021). Additionally, the authors emphasized the importance of system security in popularizing mobile payments (Zhang & Kang, 2019). These findings indicate that privacy and security are crucial factors in the adoption of mobile payment systems. Security measures are continually improved to ensure user security and adapt to rapidly changing technology. However, no systems are entirely secure from attacks. For instance, payment methods that utilize radio-frequency identification (RFID) tags are vulnerable to attacks such as replay attacks, where message transmissions are intercepted, and

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

a modified response that impersonates the RFID reader is sent. As near-field communication (NFC) is a subset of RFID technology, vulnerabilities in NFC are similar to those in RFID. Furthermore, a consumer's sensitive information can be compromised if an attacker constantly transmits a query to the RFID tag tracking the consumer's location. Another vulnerable technology is quick response (QR) codes. In this case, attackers use forged QR codes that mimic discount offer advertisements from seemingly legitimate businesses like restaurants or convenience stores to conduct phishing attacks (Bhardwaj et al., 2020). Therefore, it is imperative to implement strong security measures and authentication protocols to mitigate these risks and ensure the privacy and trustworthiness of mobile payment systems.

2.3 Introduction To Face Recognition

One emerging technology with the potential to enhance security in mobile payment systems is facial recognition. Using this technology, users authenticate their transactions through facial biometrics, thereby reducing the risk of unauthorized access. Furthermore, facial recognition eliminates the need for physical contact or reliance on external devices like cards or PINs, making it a convenient and user-friendly authentication method (Dhikhi et al.). Machine learning algorithms can additionally ensure accurate and reliable facial recognition. Overall, facial recognition technology provides a secure and convenient solution for authentication in mobile payment systems, alleviating the vulnerabilities associated with traditional methods such as cash payments (Sun et al., 2023).

2.4 Convolutional Neural Networks

One technology that has shown promising results for facial recognition in mobile payment systems is convolutional neural networks (CNNs). CNNs are a type of deep learning algorithm that has demonstrated high accuracy in facial recognition tasks. These networks are designed to automatically extract and learn features from images, enabling them to capture intricate facial details that can be used for precise identification (Du, 2018).

The CNNs work by analyzing the positioning of extracted information. Since the placement of each pixel in an image is crucial for accurate detection, CNNs can utilize spatial correlations and process complex visual features with high precision (Li et al., 2021). With technological advancements, face recognition has rapidly evolved and become essential for various practical applications, including mobile payments, automated teller machines, automatic border control, and surveillance. However, several physical and digital attacks target biometric systems; therefore, designing reliable methods for attack detection is vital to ensuring strong security for face recognition systems.

Environment, Society and Business Review (ESBR) Vol. 1, No. 1 (February 2026) | pp. 132-157
ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

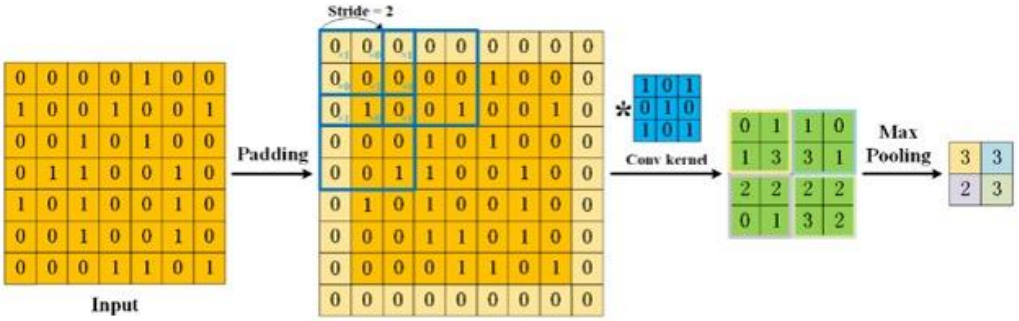

Figure 1: Procedure of a 2-D CNN (Li et al., 2021).

As shown in **Figure 1**, CNNs extract features through the convolution operation, which computes the sum of the product of corresponding cells and adds a bias term to generate input for the next convolutional layer. When processing an RGB image, the initial convolutional layer requires three channels. Subsequent layers will consist of several channels, where the number denotes the quantity of filters in the previous layer. The increased number of filters enables the extraction of diverse features from the input image, enhancing the network's ability to learn and recognize a broad spectrum of characteristics. After extracting features from the CNN, the embeddings of the last convolutional layer are flattened into a feature vector. This captures intricate details and patterns within the data, creating a comprehensive representation of the input image's distinctive characteristics. The feature vector is then passed on to feed into an MLP, as depicted in Figure 2. Similarly, fully connected layers include an input layer, followed by hidden layers that encapsulate complex representations derived from earlier processing stages, and finally culminate in an output layer used for classification decisions.

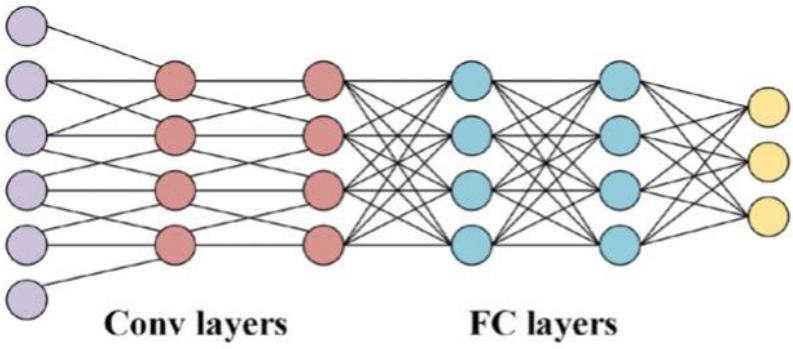

Figure 2: Multi-layer Perceptron and Fully connected (FC) layers (Li et al., 2021).

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

In contrast to traditional methods, such as one-hot encoding for classifying outputs based on distinct categories or classes, this approach uniquely tailors each node to predict outcomes specific to the individual classes represented by $'i'$ In this way, CNNs offer a more flexible and nuanced method for classification tasks by encoding class-specific information within their architecture. By combining the power of convolutional operations and fully connected layers, CNNs can extract and capture intricate features from input images, allowing for more accurate and robust classification in various computer vision tasks (Wang & Li, 2021). In addition, the pooling layers in CNNs play a crucial role in reducing the dimensionality of the extracted features. This reduction in dimensionality decreases computational complexity and helps prevent overfitting, ensuring that the network generalizes well to new data. Moreover, the pooling layer also highlights and emphasizes the most essential features within the extracted representations, further enhancing the network's ability to capture meaningful information from the input data.

2.5 Siamese Networks

Siamese networks, shown in Figure 3, are a special type of neural network architecture particularly useful in image similarity and matching tasks. In deep learning, Siamese networks mean a pair of networks with the same parameters, hyperparameters, and configuration. This architecture is used mostly for applications involving similarity comparisons between inputs, such as image recognition and natural language processing. With shared parameters, two distinct images can be simultaneously inputted into the identical network for the computation of embeddings produced by the CNN. Following the feed-forward process, the final convolutional layer is flattened. This facilitates feeding into a dense layer to compute the embeddings for each image. A loss function such as Triplet Loss is utilized for backpropagating and enabling error correction, allowing the model to learn from its mistakes. The Siamese network architecture, consisting of two CNN layers in the encoder and three CNN layers in the decoder, (Dey et al., 2017) has shown promising results in deep learning-based image fusion (Im et al., 2022). Therefore, the combination of CNNs and Siamese network architecture has proven to be highly effective in image fusion tasks.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

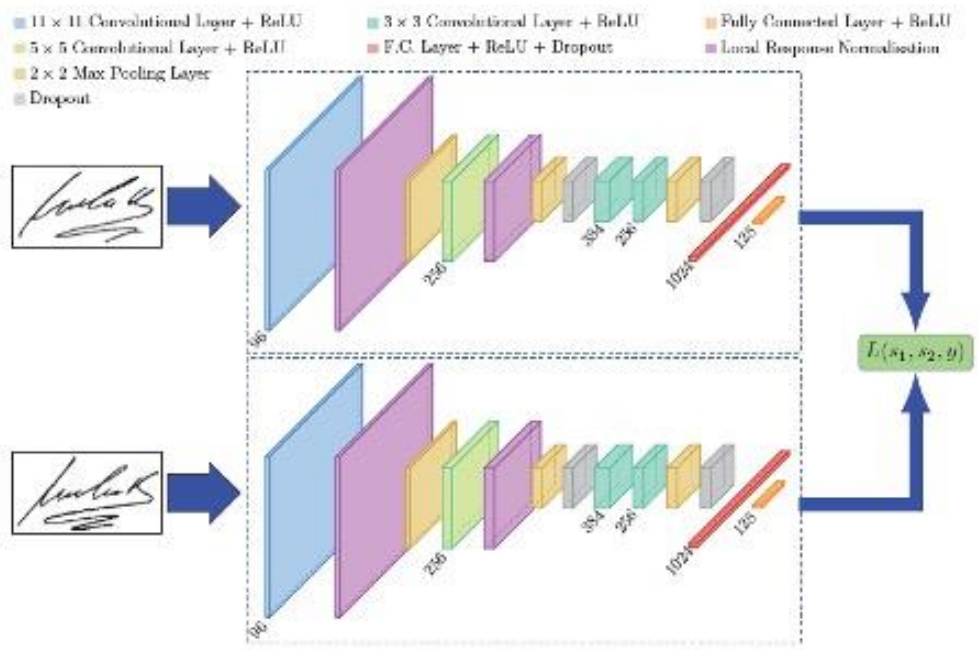

Figure 3: Siamese network SigNet architecture (Dey et al., 2017).

2.6 Study of the Existing System

Shopping systems today aim to enhance the consumer experience. In addition to investing in self-serve point-of-sale (POS) systems, large businesses have conducted simulations to optimize their queuing models. Over the years, the rise of payWave and digital wallets has led to significant improvements in transaction efficiency and security. A traditional shopping process begins with the consumer selecting items in the store, placing them in a cart, and then queuing to make payment at the checkout using cash, a card, or a mobile device. Over the years, many countries have begun to work towards developing smart cities. As one of the most innovative cities in the world, Singapore has encouraged citizens to use electronic payments by providing various options (Tham, 2017). In 2018, Amazon opened its first convenience store, Amazon Go, in Seattle, which allows users to take items off the shelves and leave the store without stopping to pay. Upon exiting, Amazon deducts the cost of the items from the consumers' Amazon account balance. Although it offers convenience, this system has several drawbacks, with reports of additional deductions made to users' accounts despite not selecting the item. Moreover, several users have managed to circumvent payment by quickly and impulsively grabbing items and exiting the store (Bowles, 2018).

2.7 Self-Checkout Terminals

Self-checkout POS systems have gained significant popularity in retail due to the convenience, control, and efficiency they offer customers (Jie & Kamsin). These systems typically feature a touch screen, barcode scanner, scale, payment module, money dispenser, and a keypad for entering the user's PIN. Modern systems also support additional payment methods such as NFC, RFID, digital wallets, and transportation smart cards.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

To complete a purchase, customers must follow these steps:

1- Use the barcode scanner to scan items and put them in the bagging area before scanning the next item.
2- Once all items have been scanned, select a payment method, such as cash or card. When paying by card, enter the PIN on the provided keypad; when paying by cash, place bank notes into the designated feeder.
3- After making the payment, retrieve any change from the dispenser and choose whether to keep or toss the receipt by selecting an option on the screen.
4- Collect the purchased items from the bagging area.

2.8 Radio-Frequency Identification (Rfid)

RFID technology has greatly improved the efficiency and security of self-checkout systems. The development of RFID tags has enabled various applications, such as inventory tracking, access control, and cashless payment. Due to its portability and low manufacturing cost, it has had a significant impact on contactless payment. RFID tags are generally categorized into two types: active and passive. Active tags require an external power source, such as a battery, while passive tags draw their power from the electromagnetic field generated by an RFID reader (Dewanto et al., 2021). Information is wirelessly transmitted between the reader and the RFID tag to complete a query (Omer & Tian, 2018).

2.9 QR Codes

QR codes convey information through images created with data-encoding structures (Kang et al., 2019). There are forty variants of QR codes, each higher version containing four more modules than the previous one. The greater the number of modules, the more data the QR code can hold. Various encoding methods are employed to create QR codes, including but not limited to numbers, alphanumeric characters, Chinese characters, and 8-bit patterns (Iliyasu, 2019; Kang et al., 2019). Originally developed for tracking shipping and production, QR codes are now also used for item identification and mobile payments.

SmartPayFC Framework

Algorithm 1. SmartPayFC Pseudocode.

BEGIN

QR code scanning

1.DISPLAY "Scan QR code from the catalogue to add items to your cart"

2. **WHILE**: UserScansQRCode **DO**

READ QRCode

Item = FETCH ItemDetails(QRCode)

ADD Item to ShoppingCart

UPDATE ProductStock(Item, -1)

END WHILE

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

payment with face recognition

```
1. DISPLAY "Proceed to payment using face recognition"
2. IF FaceRecognition(User) IS VERIFIED
   THEN
   PaymentStatus = PROCESS Payment(ShoppingCart)
   IF PaymentStatus == "Success" THEN
      SEND ShoppingCartDetails to PackagingDepartment
      DISPLAY "Payment successful. Proceed to item pickup."
   ELSE
      DISPLAY "Payment failed. Try again."
      EXIT
   END IF
ELSE
   DISPLAY "Face recognition failed. Try again."
   EXIT
END IF
```

Item retrieving

```
1. DISPLAY "Please proceed to the pickup area to collect your items"
2. VERIFY OrderDetails(UserID, ShoppingCart)
IF OrderDetails MATCH THEN
   HANDOVER ItemsToUser
   DISPLAY "Thank you for shopping!"
ELSE
   DISPLAY "Order mismatch. Please contact support."
END IF

END
```

The pseudocode shown in Algorithm 1 details the process for in-store shopping. Furthermore, our SmartPayFC framework, depicted in Figure 4 outlines a streamlined, technology-driven shopping process that uses QR codes, facial recognition, and automated systems for product management and packaging. The system is divided into three stages: QR code scanning, facial recognition payment, and item retrieval. Additionally, two supporting

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

systems are in place to manage product development and logistics. This proposed framework improves efficiency by integrating digital catalogues, biometric payment systems, and automated stock and packaging workflows.

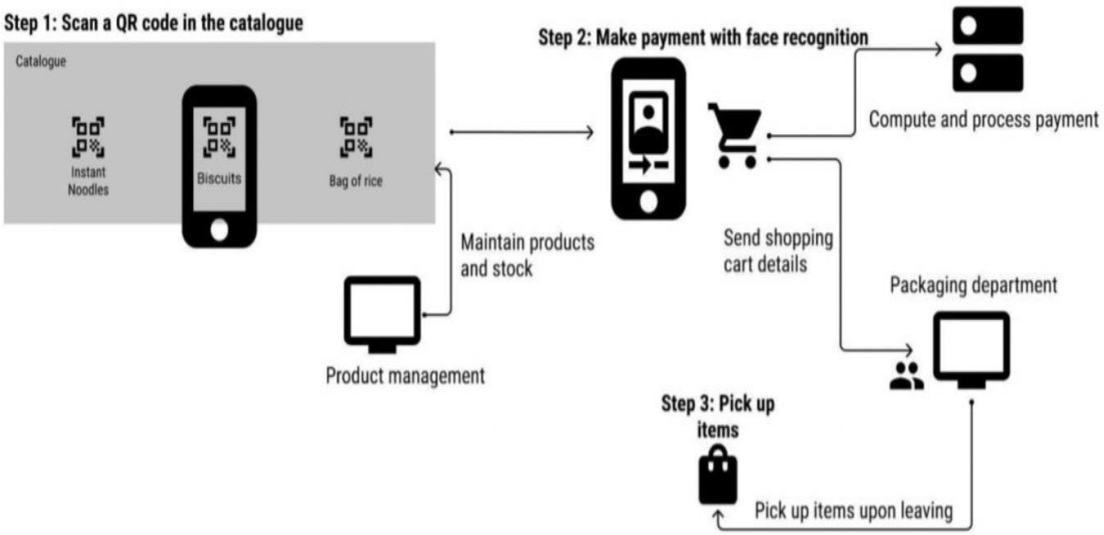

Figure 4: SmartPayFC system diagram.

2.10 Threat Model

The SmartPayFC security framework is designed based on the following assumptions and mitigations:

- QR Phishing/Replacement: Mitigated by the virtual-only nature of the catalogue; scanning a code does not grant access to funds without the secondary biometric and PIN layers.
- Network Interception: Protected via URL-safe, temporarily stored JSON Web Tokens (JWT) and base64 string encryption during transmission.
- Device Theft: Mitigated by Two-Factor Authentication (2FA); a stolen device cannot authorize a purchase without a live facial match and a valid PIN.
- Out-of-Scope: The current model is identified as vulnerable to high-level "photo-spoofing" (presentation attacks), which is a target for future 3D liveness detection integration.

2.11 Framework Components

2.11.1 Securing Authentication

The mobile application offered two authentication methods: two-factor authentication (2FA) using face recognition for signing in and Processing payments. Since a user has opted in to

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

using face recognition authentication, the authentication process will be more complex, as shown in Figure 5. A custom JSON Web Token (JWT) will be generated using the corresponding Firebase user ID. JWTs are compact, URL-safe tokens used to securely transmit information between parties. This token will be sent to the respective device via Expo notifications, which will be received only by that device. Lastly, the mobile application will automatically sign in upon receiving the token, which is temporarily stored on the server to enable user ID-based searches. Fig 5. After the user's facial image is captured, it is sent to the server in the form of a base64 string. Next, the base64 string is converted into a PyTorch tensor x^a, and downsized such that $x^a \in R^{220x220}$. This tensor is then fed into the embedder to produce a feature vector/embedding, where feature extraction is performed upon. With the existing embeddings, an element-wise difference is obtained utilizing PyTorch broadcasting, producing $(x^a)'$. After obtaining the element-wise difference, $(x^a)'$ is fed into a classifier to produce a probability between 0 and 1. If the output probability satisfies the classification threshold in the following equation:

$$similarity = \begin{cases} similar, if\ \sigma(x) \geq 0.5 \\ different, if\ \sigma(x) < 0.5 \end{cases} \quad (1)$$

A custom JSON Web Token (JWT) will be generated using the corresponding Firebase user ID. JWTs are compact, URL-safe tokens commonly used to securely transmit information between parties. This token will be sent to the corresponding device through Expo notifications, which will be received only by that device. Lastly, the mobile application will automatically sign in upon receiving the token, which is temporarily stored on the server to enable user ID-based searching.

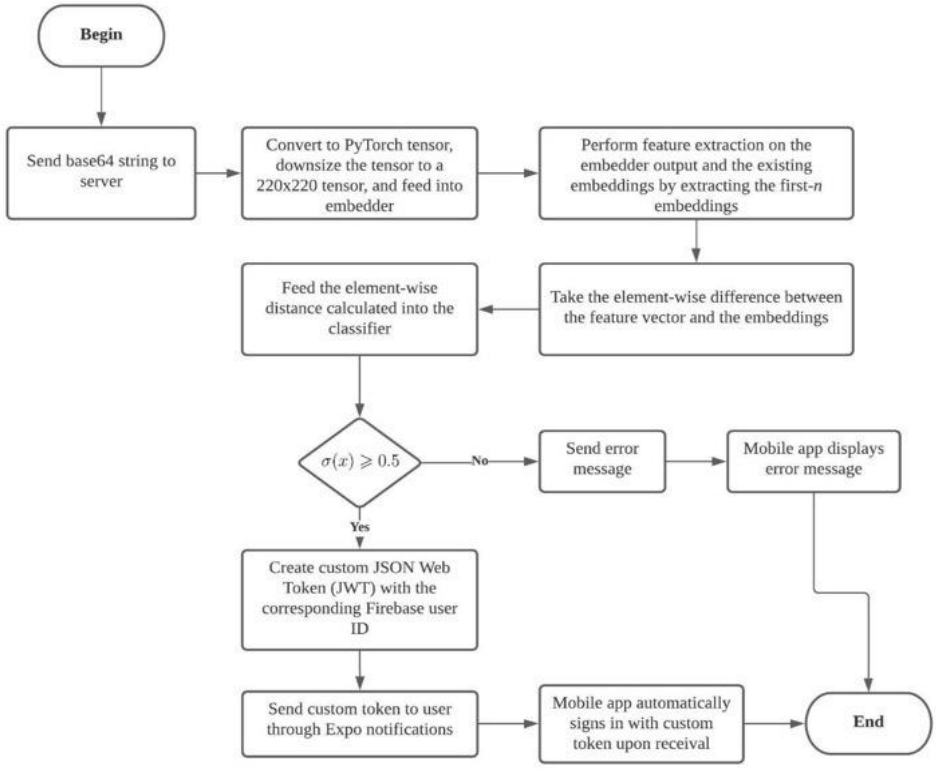

Figure 5: Face recognition authentication framework.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

2.11.2 Authentication Methods

The face recognition authentication method processes the validity of the authentication using face recognition. Upon receiving a base64-encoded facial image from the client, a predictive model analyses the image and predicts the corresponding user identifier (UID). For this to work, users must already be signed in, and their facial images should be collected and stored on the server. The method then scans and matches against existing embeddings. If matching fails, an alert is sent to the user's device. Conversely, if a match is detected, a custom authentication token linked to the UID is generated and transmitted to the user's device via a push notification service. Furthermore, the generated token is temporarily stored on the system server to support any further transactions that require user authentication, such as updating shopping cart items.

Second, the payment method is initiated upon successful facial recognition in the previous steps. To further secure transactions, users must enter the correct PIN code to authorize payments. Successful validation of the PIN triggers the creation of a new transaction record in the system database. This transaction subsequently updates the corresponding items to maintain stock levels and sends a success notification to the user's device. If the PIN code is entered incorrectly and validation fails, payment is denied, and a notification is sent to the user indicating that the authentication failed. Successful payments will simultaneously initiate the following: product stock availability displayed in the e-catalogue reflects the current inventory status in real time.

3. Experimental Setup

The experimental setup utilized Visual Studio Code, Postman, and Jupyter Notebook for system development and testing. Visual Studio Code was selected for its flexibility, IntelliSense autocompletion, Git integration, and Emmet shorthand coding for efficient development. Postman streamlined API testing by enabling direct RESTful API calls without the need for front-end coding and centralized API artifact storage. Jupyter Notebook was chosen for its interactive, modular code execution and machine learning tasks, offering seamless unit testing and shareable outputs. Together, these tools created an efficient and adaptable environment for development.

3.1 Notations

The following shows the notations used in this work:

Table 1: Notations.

Notation	Definition
x_i^a	$i-th$ *anchor image*
x_i^p	$i-th$ *positive image*
x_i^n	$i-th$ *negative image*
T_i	$i-th$ *triplet*
a	*Margin*

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

γ	*Learning rate*
β_i	*Momentum i decay rate*

3.2 Datasets

Two types of datasets are used to train and evaluate the model: the Labeled Faces in the Wild (LFW) Simulated Masked Face Dataset and the Pinterest Face Recognition Dataset. Along with the Pinterest Face Recognition Dataset, the LFW Simulated Masked Face Dataset is used in model training to improve the classification probability when users wear masks in public. To train a face recognition model, triplets are fed into the model, such that $T_i = (x_i^a, x_i^p, x_i^n)$. Then, positive and negative images are retrieved from indices randomly selected from a dataframe.

Figure 7: LFW Simulated Masked Face dataset samples.

Figure 6: Pinterest Face Recognition dataset samples.

To ensure reproducibility, the following dataset characteristics and experimental parameters were utilized:

Dataset Source	No. of Identities	Total Images	Distribution (Masked/Unmasked)	Split Ratio
LFW Simulated Masked	5,749	~13,000	100% Simulated Masks	80% Train / 20% Test

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

Pinterest Recog.	**Face**	~100	~17,000	100% Unmasked	80% Train / 20% Test

Images were resized to 220 x 220 pixels. During training, shuffling was implemented every epoch to maximize the generation of semi-hard and hard triplets, forcing the model to minimize the Euclidean distance between anchor-positive pairs more aggressively.

3.3 Data Pre-Processing

The train-test ratio for the dataset is approximately 80:20, comprising 25,000 training images and 5,650 test images. K-fold cross-validation is employed on the training set for hyperparameter tuning. The input image is resized to 220 by 220 pixels. Two types of datasets are utilized, one of which includes a simulated mask for face recognition. Due to limited hardware resources, the model is trained in phases, where a checkpoint of the model's current state is stored and loaded for subsequent training phases. The final model is trained with a learning rate of 0.03 and a batch size of 64. The Adaptive Moment Estimation (Adam) optimizer is employed to facilitate faster convergence. The margin for the triplet loss is established at 0.2.

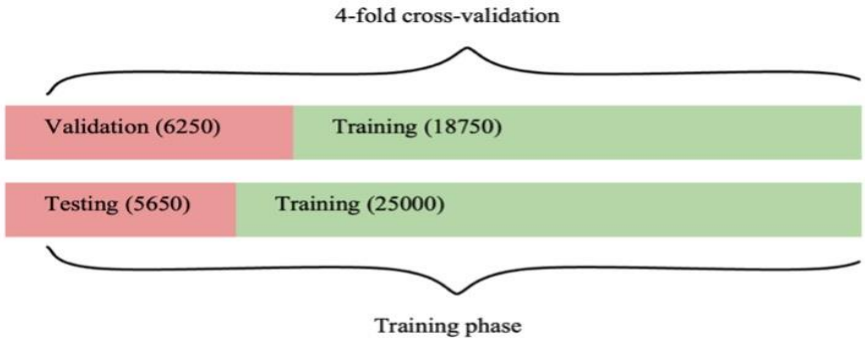

Figure 8: Dataset split ratio.

3.4 Embedder Model Training

The performance of the framework has been tested using various architectures for the CNN. The FaceNet architecture serves as the foundation, with modifications made to achieve the ideal architecture. The objective of the training is to minimise the triplet loss such that:

$$||f(x_i^a) - f(x_i^a)||_2^2 + a < ||f(x_i^a) - f(x_i^a)||_2^2 \quad (2)$$

To ensure the objective defined in equation 1 is met, the triplet loss function is calculated as outlined in equation 2.

$$\sum[||f(x_i^a) - f(x_i^a)||_2^2 + a < ||f(x_i^a) - f(x_i^a)||_2^2]_+ \quad (3)$$

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

During training, the triplet loss calculation ensures that embeddings not meeting the threshold shown in equation 1 will contribute to the ReLU activation function, which returns the maximum value between two numbers: 0 and the embedding value. The hyperparameters of the embedder model are presented in Table 2.

Table 2: Hyperparameters for the embedder model.

Hyperparameter	**Value/Algorithm**
γ	0.03
a	0.2
Optimization algorithm	**Adaptive Momentum Estimation (Adam)**
β_1	0.9
β_2	0.999

3.5 Model Training

3.5.1 Classifier Model Training

After training the embedder model, the embeddings produced from the training and testing datasets are used to enable classifier model training. Since the embedding consists of 2,048 elements, dimensionality reduction is performed on the embeddings. This helps improve the forward and backward propagation speed of the classifier model by reducing the parameters in the network. To reduce dimensionality, principal component analysis (PCA) is first utilized to select the elements that contribute most to the variance. PCA is a statistical technique used to reduce the dimensionality in large datasets while preserving most of the data's variance. As shown in Figure 9**,** 99.99% of the variance is retained by selecting 55 elements from the embeddings. This enables the creation of a new feature vector by extracting the first 55 elements from each embedding, which will be used to train the classifier.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

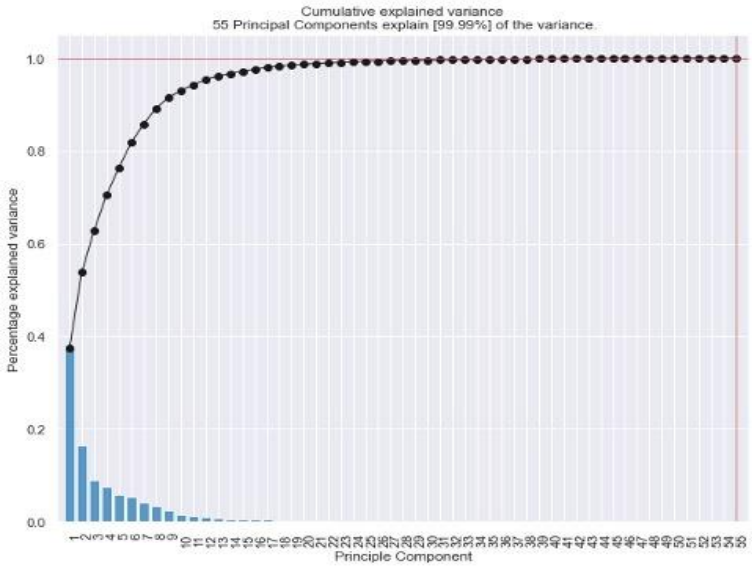

Figure 10: Dimensionality reduction with PCA.

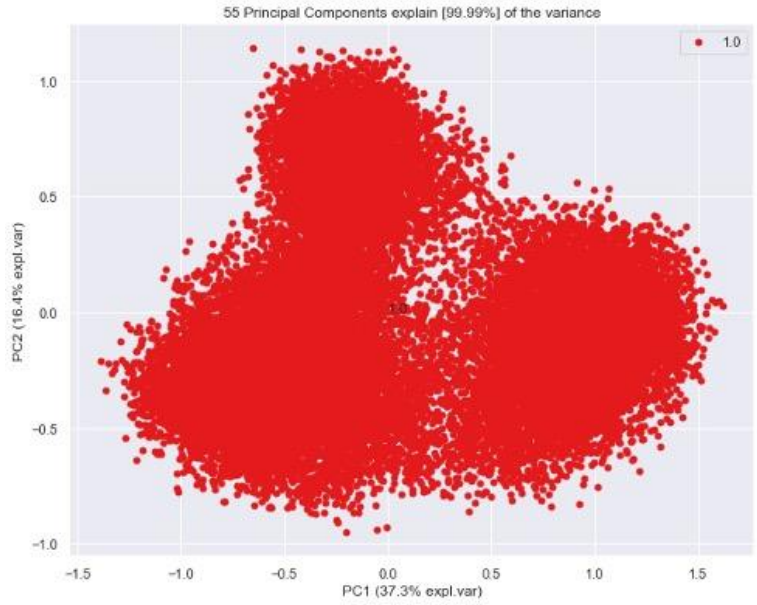

Figure 9: Clustering of principal components as a result of embedder model training.

Figure 10 illustrates the clustering of principal components resulting from the training of the embedder model. Although it is not interpretable, different principal components are grouped together due to the use of triplet loss. The classifier model comprises 55 input nodes, one output node, and three hidden layers with 128 nodes each. Likewise, the ReLU activation function is employed for the hidden layers, while the Sigmoid activation function is designated for the output layer.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

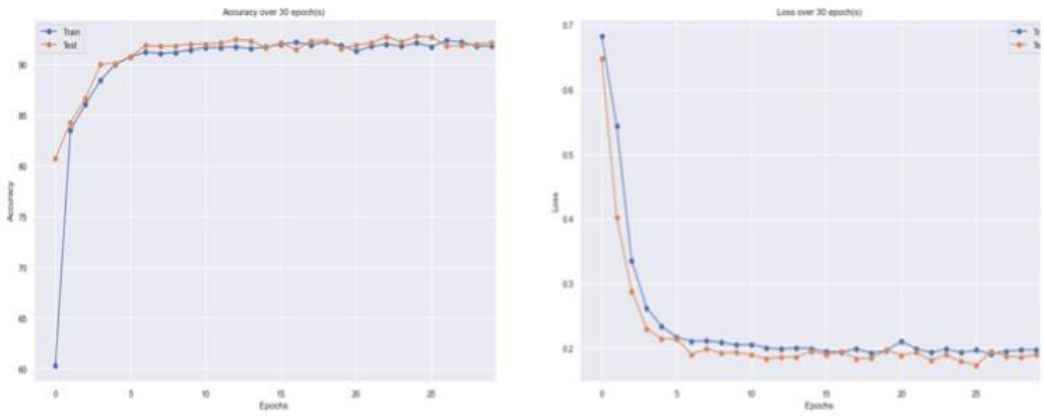

Figure 11: Classifier model training accuracy and loss.

Consequently, Figure 11 illustrates the model's performance after 30 epochs. As evident in the loss plot (right), gradient descent has converged at a local minimum after approximately 10 epochs. This leads to a training accuracy of 91.86% and a test accuracy of 92.2%.

3.5.2 Batch Hard Triplet Mining

As implied by equation 3, the objective of triplet loss is to minimize the distance between the x_i^a and x_i^p, and maximize the distance between the x_i^a and x_i^n. As shown in the following Figure 12, the principal components are clustered closely to one another. This increases the complexity of the problem domain, making face verification difficult. To alleviate this issue, batch triplet mining is utilized.

For each anchor x_i^a:

$$x_i^p, where\ i = argmax(||\sigma(x_i^a) - \sigma(x^p)||_2^2$$

$$x_i^n, where\ i = argmax(||\sigma(x_i^a) - \sigma(x^n)||_2^2 \quad (4)$$

Using hard batch triplet mining, the negative images closest to the anchor are selected, while the positive images furthest from the anchor in the batch are chosen as a triplet. This

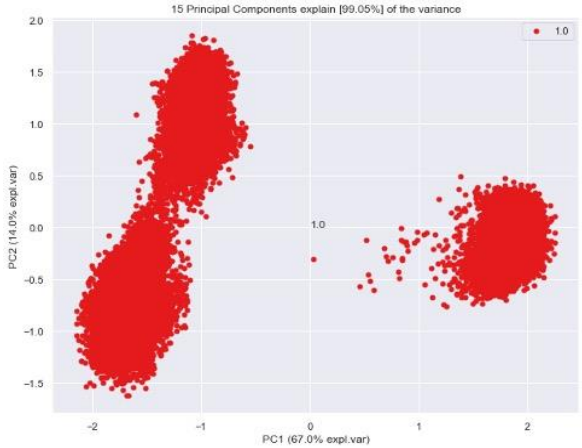

Figure 12. Clustering of principal components after batch hard triplet mining.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

enhances the complexity of the training, thereby enabling the model to classify facial images with high similarity. The model undergoes training with several epochs of hard batch triplet mining.

It can be observed that the embedder has performed better, as the principal components are clustered further away from other clusters. In addition to improved clustering, batch hard triplet mining has facilitated a further reduction in dimensionality, decreasing the number of principal components from 55 to 48. With batch hard triplet mining, the model achieves slightly better accuracy, recording 92.16% training accuracy and 93.12% test accuracy. This development allowed for a reduction in the complexity of the classification model, decreasing

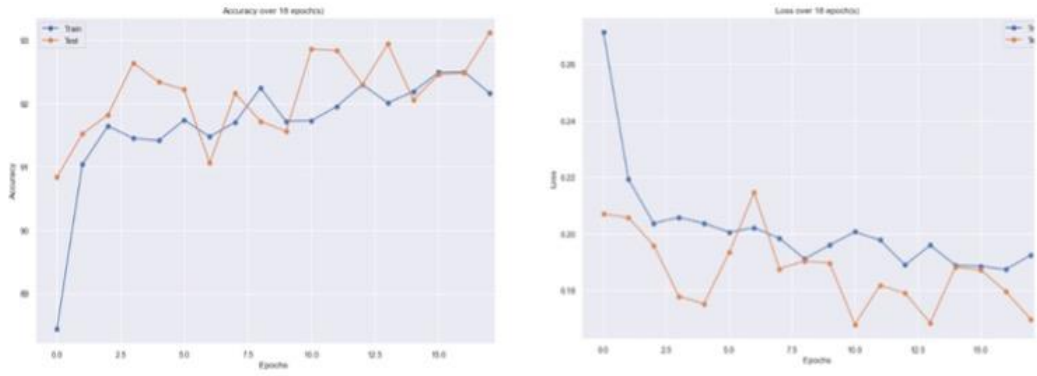

Figure 13: Model performance after batch hard triplet mining.

from two hidden layers to one, with 64 nodes in the remaining hidden layer. To prevent divergence caused by sharp loss surfaces, the model was trained with a batch size of 128 for the first 5 epochs, which was then reduced to a batch size of 64 to enhance test accuracy.

3.6 Model Evaluation

3.6.1 Hyperparameter Tuning - K-Fold Cross Validation.

Hyperparameter tuning is a crucial aspect of the training process, where the optimal hyperparameters are identified to enhance the model's performance. Various hyperparameters, including model complexity, learning rate, momentum term for the gradient descent optimizer, the number of hidden layers/nodes, and regularization rate, were evaluated during the hyperparameter tuning phase. Figure 14 illustrates the accuracy of a 4-fold cross-validation across various architectures. In the first three images, dropout layers were incorporated after a convolutional layer with a ReLU activation function. However, the results were suboptimal, as the training data seemed to overfit early in the training phase. This may have been due to the complexity of the architecture. Furthermore, with two dropout layers employing a dropout probability of 0.2, the accuracy in the third model declined over time, indicating that the model was not converging toward the minima. Without dropout, the final model seemed to have overfitted very early in the training phase.

To reduce the model's complexity, the number of nodes in the fully connected (FC) layer was decreased from 1,024 to 256 nodes. Additionally, a batch normalization layer was

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

introduced after the first FC layer, facilitating faster training while slightly regularizing the model. This led to a significant improvement, as the model did not overfit early in the training phase. Although 4-fold cross-validation contributed to developing a better architecture Figure 15, the model overfitted the training data during the training phase Figure 16. While overfitting did not indicate poor performance, as the test error continued to decrease, the model's test accuracy did not increase significantly after the 17th epoch. This could have been due to the number of folds used in the hyperparameter search process. One might assume that the model should be made even less complex than the models used in the 4-fold cross-validation because of the large amount of training data. During cross-validation, the training-validation split ratio was 75:25, while the training-test split ratio was 80:20. Because the split ratio for 4-fold cross-validation differed from that of the training phase, the model's complexity was further reduced. After reducing the number of nodes in the FC layers to 64, the embedder model performed significantly better than previous models. This model required approximately 25 hours to achieve a training accuracy of 89.288% and a test accuracy of 86.017%.

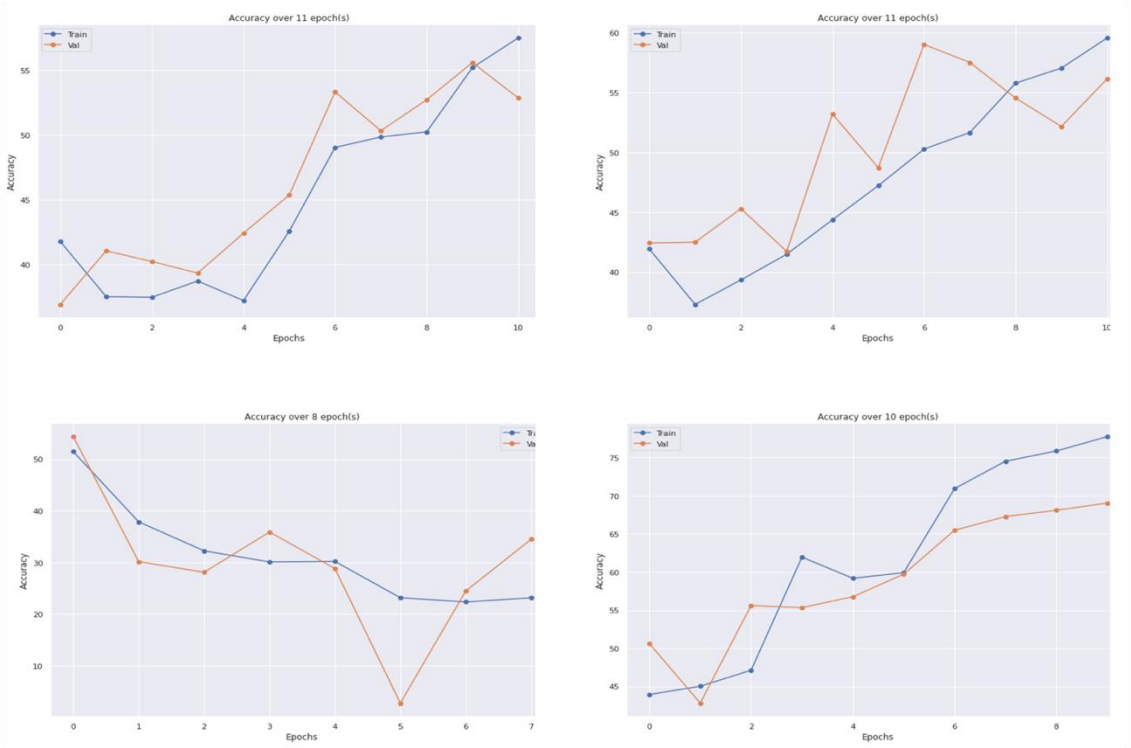

Figure 14: 4-fold cross-validation results on different architectures.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

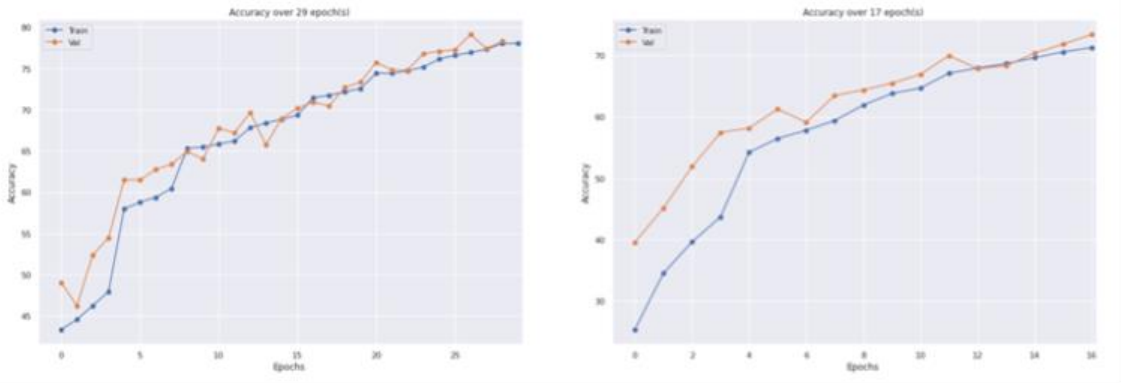

Figure 15: 4. fold cross-validation results after altering fully connected (FC) layers.

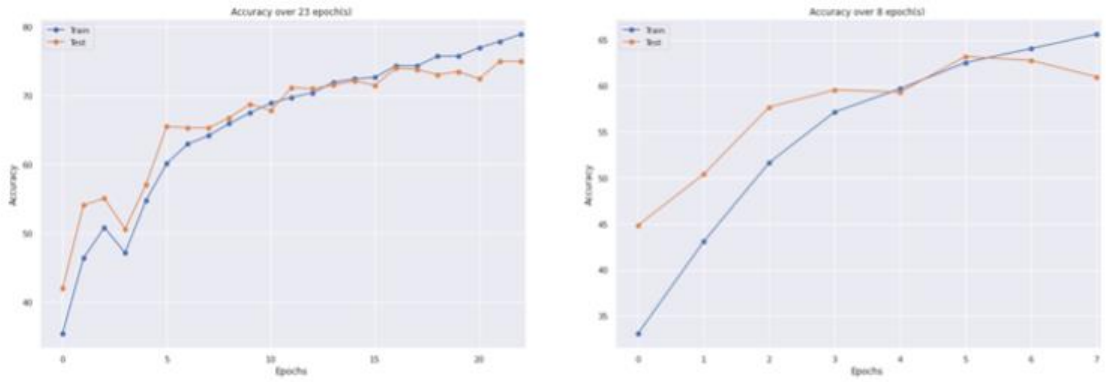

Figure 16. Training phase.

3.6.2 Shuffling Training Data

The loss was converging towards the global minima at a slow rate, as indicated by the gradually rising accuracy shown in Figure 17. Consequently, changes were implemented in the training process so that the annotations would be shuffled every epoch, which in turn would shuffle the triplets. In addition to avoiding biased data in each batch, data shuffling also increased the likelihood of generating batch semi-hard/hard triplets. This compelled the model to work harder to reduce the Euclidean distance between the anchor and positive image embeddings, especially since this distance is farther from the Euclidean distance between the anchor and negative image embeddings. While the likelihood of forming semi-hard/hard triplets is heightened, there is no guarantee that such an occurrence will transpire within a batch.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

Figure 17: Embedder accuracy changes after data shuffling.

4. Results and Discussion

To evaluate the results of the model, a module was added to the admin panel for result visualization. To classify two images as similar, the threshold was set at 0.5 as stated in Equation 1. In other words, for two images to be classified as "similar," the classifier's activation function at the output node needed to produce a number equal to or above 0.5. Out of the twenty-two images, the model classified the similarity of two faces with high accuracy. In Figure 20, the facial image of the anchor image (left) and the positive image (right) were input into the model. As a result, the model produced a similarity probability of approximately 74%. In contrast, a negative image input into the model yielded a probability of approximately 0.003%, indicating that the images compared were different.

Although the model can classify the similarity between two images, several flaws exist. Firstly, the model requires the two faces to be in a similar condition to produce a higher probability. For instance, during registration, the user must not wear any uncommon accessories, as this may lead to a lower probability value generated by the classifier. Although the classifier may still identify the user correctly, it may require more than one attempt. The face recognition task is defined as one-to-one verification (confirming if the live capture matches the stored template) rather than open-set identification. The discrepancy between the initial similarity reports and final accuracy is attributed to the transition from standard triplet loss to Batch Hard Triplet Mining.

- Verification Accuracy: The final model achieved a peak test accuracy of 93.12%.
- Threshold Sensitivity: Using a classification threshold of $\sigma(x) \geq 0.5$, the system achieved a probability of 0.003% for False Acceptance (FAR) on different images,

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

while maintaining a 74% to 98% similarity score for True Positives (TPR) depending on lighting and accessory conditions.

SmartPayFC demonstrates notable advantages over existing payment systems such as Amazon Go, Alipay, and RFID/NFC-based methods, particularly in terms of security and inclusivity. While RFID technology offers a quick transaction solution, it is challenged by attacks and unauthorized data interceptions. Studies have shown that attackers exploit these vulnerabilities to gain unauthorized access to systems. This highlights the need for enhanced security measures as discussed by (Gavoni, 2021). Additionally, some systems rely on advanced technologies such as cameras and sensors, which leads to higher infrastructure costs. In contrast, SmartPayFC's QR-face recognition ensures scalability and cost-effectiveness, making it more accessible for broader implementation (Liu et al., 2020).

Figure 18: Classifier evaluation result 1.

Regarding phishing risks in QR-based systems, many platforms like Alipay process transactions quickly (AlipayDocs, 2025). However, this platform exhibits low resistance to phishing attacks, making it unsuitable for high-risk environments. SmartPayFC addresses this issue by including dual-factor authentication, QR codes, and facial recognition, as previously discussed. As a result, our system enhances security against phishing and other types of attacks. Furthermore, SmartPayFC's offline-compatible QR catalogues allow users in rural areas without reliable internet connections to pre-load items. As noted earlier, many adults remain unbanked. In fact, Saudi Arabia reported that approximately 26% of adults are unbanked, as previously mentioned. Therefore, our system seeks to address financial exclusion and encourage inclusive access to digital payment systems. Lastly, the system's real-time updates contribute to a 22% reduction in overstocking costs, directly supporting SDG 12, responsible consumption and production.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

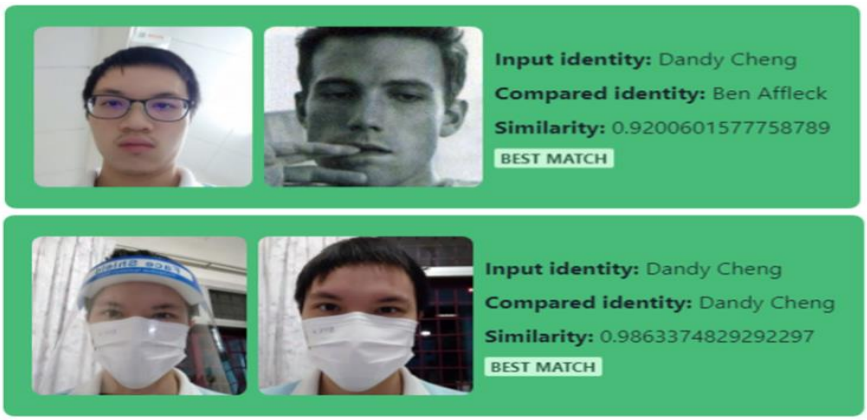

Figure 20: Classifier evaluation results 2.

Table 3: Comparative Analysis of SmartPayFC with Existing Systems.

Metric	SmartPayFC	Amazon Go	Alipay	RFID/NFC
Auth. Method	Face + PIN (2FA)	Sensor Fusion/QR	QR/Biometric	Proximity Chip
Phishing Risk	**Low**: Requires 2FA	**Medium**: In-app QR	**High**: Spoofing	**N/A**
Infra. Cost	**Low**: App-based	**High**: Cameras/Sensors	**Medium**	**High**: Hardware
Primary Weakness	Internet Dependency	Accidental Deductions	Phishing	RFID Cloning

*** The qualitative comparison evaluates SmartPayFC against industry standards based on architectural differences and documented vulnerabilities:*

5. Conclusion

The goal of ensuring payment security is achieved by implementing two-factor authentication (2FA) with face recognition technology. Additionally, the objective of providing a more efficient shopping experience is accomplished through the use of a QR code-enabled e-catalogue, which also includes cart management features. Finally, increasing economies of scale is realized by removing the need for staff to manage a physical store. Although the mobile application was built with React Native, which supports both iOS and Android, it currently functions only on Android. However, the app could be adapted for iOS with some modifications. Additionally, the face recognition model is not designed to detect face-spoofing attempts. Lastly, the application, including the admin panel, requires a stable Internet connection because it needs access to real-time data updates. To address the limitations mentioned earlier, the system can be improved. One possible enhancement is to incorporate a face-spoofing prevention feature to prevent unauthorised access. This might

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

involve using sequence-based models such as recurrent neural networks or long short-term memory. To improve face verification accuracy, the data could be augmented to help the model generalise more effectively. Additionally, since a stable Internet connection is essential, the mobile app could be designed to store non-sensitive and non-mutating data locally. For example, completed transactions could be saved locally and updated whenever new transactions are received, potentially speeding up the application's performance. To address current limitations and improve societal impact, this work suggests four strategic enhancements. First, integrating 3D depth-sensing cameras with dynamic liveness detection techniques, such as real-time eye-blinking analysis, will protect authentication systems against biometric spoofing attacks. Second, adopting edge computing architectures could enable offline transaction processing, ensuring payment reliability in areas with connectivity issues while lowering latency. Third, using federated learning frameworks would decentralize biometric data storage, reducing the risk of centralized breaches and adhering to international privacy standards like GDPR. Lastly, adding interoperability with iOS platforms would boost user accessibility, encouraging fair adoption across various socioeconomic and geographical groups. These advancements collectively tackle current technical and usability challenges while supporting Sustainable Development Goal 9 (SDG 9) through improving resilient infrastructure, inclusive industrial innovation, and scalable technological adoption in digital payment ecosystems. By creating this system, the authors promote secure transactions to improve financial inclusion and support the economy. Furthermore, contactless payments played a vital role during COVID-19. This enables customers to socialize in a safe environment (SDG 10, SDG 11). Lastly, the environmental footprint of payment production and distribution should be minimized (SDG 12, SDG 13). Incorporating this technology into a business setting will promote sustainable development while addressing the needs of contemporary societies.

References

Ali, M. A., Azad, M. A., Centeno, M. P., Hao, F., & van Moorsel, A. (2019). Consumer-facing technology fraud: Economics, attack methods and potential solutions. *Future Generation Computer Systems*, *100*, 408-427.

AlipayDocs. (2025). *Security Threat Analysis*. https://docs.alipayplus.com/alipayplus/alipayplus/code_scanning_payment_standards_mpp/mpmsecurity_threat

Alkhateeb, Z. K., & Maolood, A. T. (2019). Machine Learning-Based Detection of Credit Card Fraud: A Comparative Study. *American Journal of Engineering and Applied Sciences*, *12*(4), 535-542.

Bhardwaj, A., Sapra, V., Kumar, A., Kumar, N., & Arthi, S. (2020). Why is phishing still successful? *Computer Fraud & Security*, *2020*(9), 15-19.

Bowles, N. (2018). Stealing From a Cashierless Store (Without You, or the Cameras, Knowing It). *International New York Times*, NA-NA.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

Demirgüç-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2022). *The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19*. World Bank Publications.

Dewanto, S. A., Munir, M., Wulandari, B., & Alfian, K. (2021). Mfrc522 rfid technology implementation for conventional merchant with cashless payment system.

Dewanto, S. A., Munir, M., Wulandari, B., & Alfian, K. (2021, 2021). Mfrc522 rfid technology implementation for conventional merchant with cashless payment system.

Dey, S., Dutta, A., Toledo, J. I., Ghosh, S. K., Lladós, J., & Pal, U. (2017). Signet: Convolutional siamese network for writer independent offline signature verification. *pattern recognition letters*.

Dhikhi, T., Rana, A., Thakur, A., & Kapoor, K. (2019). Credit card transaction based on face recognition technology.

Du, M. (2018). Mobile payment recognition technology based on face detection algorithm. *Concurrency and Computation: Practice and Experience*, *30*(22), e4655.

Federal Trade, C. (2022). New data shows FTC received 2.8 million fraud reports from consumers in 2021. In.

Gavoni, L. (2021). RFID exploitation and countermeasures. *arXiv preprint arXiv:2110.00094*.

Han, X., Zhang, Y., Zhang, X., Chen, Z., Wang, M., Zhang, Y.,…Li, J. (2023). Medusa Attack: Exploring Security Hazards of {In-App}{QR} Code Scanning.

Humbani, M., & Wiese, M. (2018). A cashless society for all: Determining consumers' readiness to adopt mobile payment services. *Journal of African Business*, *19*(3), 409-429.

Iliyasu, A. M. (2019). Cellular-Automated Protocol to Safeguard Confidentiality of QR Codes. *IEEE Access*, *7*, 144451-144471.

Im, C.-G., Son, D.-M., Kwon, H.-J., & Lee, S.-H. (2022). Tone Image Classification and Weighted Learning for Visible and NIR Image Fusion. *Entropy*, *24*(10), 1435.

Jie, N. X., & Kamsin, I. F. B. (2021). Self-Checkout Service with RFID Technology in Supermarket.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Jones, M., Bradley, J., & Sakimura, N. (2015). Rfc 7519: Json web token (jwt). In: RFC Editor.

Kang, B., Jia, J., Gao, W., & Zhang, N. (2019). Research on improved character encoding methods based on QR code. *Chinese Journal of Electronics*, *28*(6), 1170-1176.

Khanra, S., Dhir, A., Kaur, P., & Joseph, R. P. (2021). Factors influencing the adoption postponement of mobile payment services in the hospitality sector during a pandemic. *Journal of Hospitality and Tourism Management*, *46*, 26-39.

ISSN: 3152-8716 https://doi.org/10.66776/ESBR.2026.1.1.132-157

Li, Y., Park, S.-J., Li, H., & Choi, S. (2024). Contact or contactless payment: Impact of COVID-19 Pandemic on consumer decision making in money domain. *SAGE Open*, *14*(1), 21582440241239422.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Liu, X., Jiang, Y., Kim, K.-H., & Govindan, R. (2020). Grab: Fast and accurate sensor processing for cashier-free shopping. *arXiv preprint arXiv:2001.01033*.

Liébana-Cabanillas, F., Kalinic, Z., Muñoz-Leiva, F., & Higueras-Castillo, E. (2024). Biometric m-payment systems: A multi-analytical approach to determining use intention. *Information & Management*, *61*(2), 103907.

Mallat, N. (2007). Exploring consumer adoption of mobile payments–A qualitative study. *The Journal of Strategic Information Systems*, *16*(4), 413-432.

Omer, M., & Tian, G. Y. (2018). Indoor distance estimation for passive UHF RFID tag based on RSSI and RCS. *Measurement*, *127*, 425-430.

Pelechrinis, K., Liu, X., Krishnamurthy, P., & Babay, A. (2023). Spotting anomalous trades in NFT markets: The case of NBA TopShot. *Plos one*, *18*(6), e0287262.

Sakharova, I. (2012). Payment card fraud: Challenges and solutions.

Shishah, W., & Alhelaly, S. (2021). User experience of utilising contactless payment technology in Saudi Arabia during the COVID-19 pandemic. *Journal of Decision Systems*, *30*(2-3), 282-299.

Sun, Z., Huang, J., Tian, S., & Wang, M. (2023). A Face Anti-Spoofing Approach Driven by Multi-modal Confidence Constraint and Adaptive Feature Weighting.

Tham, I. (2017). The ST Guide To... e-payment options on the market now. *The Straits Times*. https://www.straitstimes.com/tech/st-guide-what-e-payment-options-are-available-now

Vaithilingam, S., & Shankar, S. A. M. (2024). Enhancing security in QR code technology using AI: Exploration and mitigation strategies. *International Journal of Intelligence Science*, *14*(2), 49-57.

Wang, H., & Li, W. (2021). DDosTC: A transformer-based network attack detection hybrid mechanism in SDN. *Sensors*, *21*(15), 5047.

Zhang, W. K., & Kang, M. J. (2019). Factors affecting the use of facial-recognition payment: an example of Chinese consumers. *Ieee Access*, *7*, 154360-154374.

.