FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

IntelligentGuard: Smart Doorbell with Deep Learning for Secure User Recognition and Instant Notifications

Md Reazul Islam[1], Khondokar Oliullah[2*], Dr. Rajarshi Roy Chowdhury[3], Shaikat Das Joy[4], and M M Fazle Rabbi[5]

[1]*Assistant Professor, Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh; reazul@aiub.edu*
[2]*Lecturer, Department of Information and Communication Technology, Comilla University, Bangladesh; oliullah@cou.ac.bd*
[3]*Assistant Professor, Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh; rajarshi@aiub.edu*
[4]*Lecturer, Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh; skdas@aiub.edu*
[5]*Assistant Professor, Bangladesh University of Business and Technology, Dhaka, Bangladesh; rabbi@bubt.edu.bd*

***Correspondence:** *Khondokar Oliullah; Email: oli.it.ju@gmail.com; Phone: +880-1892352367*

ABSTRACT- In the modern world, daily activities are heavily reliant on the Internet. This study aims to provide users with a simple, personalized technology that effectively manages visitor interactions. The primary objectives are to operate the doorbell intelligently and notify users about visitors by sending a notification with an image of an unknown visitor. This system introduces a low-cost Internet of Things (IoT) smart doorbell designed to enhance home security, utilizing a Raspberry Pi and a camera sensor. Camera sensor is used to capture images in front of the doorbell, which are then processed by the Raspberry Pi and sent to the server. The pre-trained deep learning model kept on a remote server checks the captured image whether it is known or unknown; if it is unknown, a new photo is taken. The proposed hybrid CNN-LSTM model achieves 98% accuracy on the collected dataset and 96.1% accuracy on the Human Faces dataset from Kaggle. This device is utilized to remotely monitor activities outside the door and receive notifications when visitors approach the doorbell using a user-friendly mobile application. This modern security doorbell requires minimal installation tools and lacks interior wiring. In the proposed system, the ESP32-CAM sensor detects visitors and announces their names if they are recognized as known individuals; otherwise, they are labeled as strangers. All relevant details, including timestamps, are promptly sent to the user's mobile application for real-time monitoring.

Keywords: Doorbell, Face Recognition, Deep Learning, Internet of Things (IoT), Human Face Detection.

ARTICLE INFORMATION

Author(s): Md Reazul Islam, Khondokar Oliullah, Dr. Rajarshi Roy Chowdhury, Shaikat Das Joy, and M M Fazle Rabbi;
Received: 07/10/2024; **Accepted**: 21/05/2025; **Published**: 30/06/2025;
E- ISSN: 2347-470X;
Paper Id: IJEER 0710-07;
Citation: 10.37391/ijeer.130220
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-13/ijeer-130220.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Home automation and connected devices are becoming essential in modern households, enhancing both convenience and security [1]. The rapid growth of Internet of Things (IoT) technology has enabled homeowners to adopt responsive, cost-effective, and intelligent solutions [2,3], particularly in terms of security. One such advancement is the smart doorbell system [4], which not only improves security but also enhances user experience by integrating with other smart devices [5,6]. Traditional doorbells, once simple switches, have evolved into modern touchpads, sensors, and IoT-based systems [7,8]. However, the limited functionalities of traditional doorbells still leave homeowners vulnerable to burglary or unwanted visitors. Additionally, their placement, often near the ground or in hard-to-reach locations, can make it difficult to see or hear them. This paper introduces an IoT-based smart doorbell solution that offers real-time visitor recognition, emergency alerts, and instant notifications, addressing contemporary security challenges with cutting-edge technology.

The system provides a robust, user-friendly, and secure smart doorbell solution, eliminating the need for complex installations or interior wiring by utilizing IoT components like a Raspberry Pi and a camera sensor. An intelligent recognizer module, developed using deep learning techniques, identifies whether a visitor is known or unknown based on the images captured by the camera. These images are processed on the Raspberry Pi and evaluated by the pre-trained model to make a decision. If the visitor is recognized, the system announces their name. Conversely, if the visitor is a stranger, a new image is captured, and the homeowner is promptly notified *via* a mobile application.

The primary advantage of this study is its ability to provide real-time updates to users about visitors at their door. The smart doorbell, connected using Wi-Fi connectivity, communicates directly with the homeowner's mobile application, sending

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

notifications that include the visitor's image along with detailed information such as the date and time of their arrival. This functionality offers significant convenience, allowing homeowners to remotely monitor their front door and decide whether to engage with the visitor. The system's design, which minimizes the need for wiring and extensive installation, further enhances its appeal, making it accessible to a wider range of users.

This study outlines the architecture and design of the proposed model, highlighting its use of deep learning (DL) for image recognition, its IoT-based framework for real-time communication, and its built-in emergency response capabilities. By combining advanced technology with a user-friendly interface, it provides a reliable, cost-effective, and easy-to-install solution for modern homeowners seeking to enhance the security of their homes. In an era where safety and convenience are of utmost importance, the system represents a significant step forward in the evolution of smart home security technologies. The key contributions of this work are as follows:

- Developing an IoT system for home security by integrating the proposed DL model.
- Evaluating and comparing the efficiency of the applied deep model in person or human face recognition.
- Implementing the research findings into a practical mobile application that not only classifies known persons but also notifies the users, offering a user-friendly interface.

The research is organized as follows: *Section 2* reviews related work, while *section 3* details the materials and methods. *Section 4* presents the experimental setup and results. Finally, *section 5* concludes the paper and outlines future directions.

2. RELATED WORKS

This section reviews existing research and technologies relevant to smart doorbell systems and home security solutions. It explores various approaches, including the use of IoT devices, facial and voice recognition, and mobile notifications. Previous works highlight the integration of smart doorbells with mobile applications, advanced security features, and accessibility for specialized user groups. This section outlines key advancements and limitations in these systems, providing context for the contributions and innovations introduced by the proposed system.

Numerous offences fall under the umbrella of doorstep crime [9], including pressure sales, fraudulent traders, distraction burglaries, and fraud. In some cases, the offender gains initial entry and then uses distraction techniques to remain on the property, accessing more areas to commit a break-in [10]. Several advanced doorbell systems are already available on the market, each offering a variety of features and functions. Many of these systems utilize smartphone applications for communication and notifications. For instance, Park [11] proposed a security system that integrates smartphones and home networks, where closed-circuit television (CCTV) footage is sent to the homeowner's phone when a visitor presses the doorbell. Alerts are sent via a real-time short message service (SMS) server, but the system is limited to a single webcam and lacks face and voice recognition.

Thabet [12] introduced a smart doorbell system that captures multiple images of a visitor to perform facial recognition. The scanned image is compared against a pre-existing database, classifying the visitor as a familiar or a stranger. Notifications are sent to the owner's smartphone based on the result. This system uses an ARMv7 Cortex-A7 Raspberry Pi [13] board with OpenCV for image processing, utilizing a PCA-based face recognition algorithm.

Ennis [14] developed an intelligent doorbell called 'Doorstep,' which identifies when someone is at the door and sends notifications to elderly users. The elderly person can review the image or request assistance from a carer, who also receives the notification and can advise whether to allow the person entry.
Kumari [15] designed a doorbell system tailored for the hearing-impaired, which captures an image when the doorbell is pressed and sends it to wearable devices via Bluetooth. The system also uses a GSM module [16] to send notifications and includes a vibrating alert and an LCD display to show the visitor's image. All data is stored on a server for future retrieval.
Pinjala [17] introduced a doorbell security system with an integrated camera. When the doorbell is pressed, a notification is sent to the owner's smartphone, and the camera activates, allowing live viewing of the visitor [18]. The homeowner can unlock the door by entering a PIN via the app or send a voice message.

Pawar [19] proposed a face recognition and IoT-enabled smart home security system that uses a Raspberry Pi with passive infrared and ultrasonic sensors. A motion detection camera captures the visitor's image, and real-time face recognition is performed using a local binary pattern (LBP). If recognized, the door unlocks; otherwise, the doorbell rings [20], and the homeowner receives SMS and email alerts with the visitor's image. Eras [21] presented a smart doorbell proof-of-concept using Bluetooth Mesh technology to extend its range and relay event notifications throughout a building. The network nodes for the system were custom-designed based on the program's requirements. Giorgi [22] introduced a doorbell system that combines voice and iris recognition to verify the identity of the person ringing the doorbell. This system integrates a traditional doorbell with a cutting-edge AXIOM Cyber-Physical Board, preventing biometric [23] data from being stored or transmitted to the cloud.

In summary, this section reviews various smart doorbells and home security systems [24-30] that utilize facial recognition and mobile alerts through IoT technology. While these systems [31-33] offer advanced features, they often face challenges such as inconsistent performance and high energy consumption. Additionally, homeowners in remote areas without data service may not be able to view video footage or receive notifications. Many existing systems offer features like image capture and smartphone connectivity but often lack comprehensive integration and ease of installation. This study presents the development of an intelligent, low-energy doorbell capable of reliably detecting human presence and sending automated

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

notifications, regardless of the homeowner's location. A summary of existing works is presented in *table 1*.

Table 1. Comparison of Existing Smart Doorbell Systems and the Proposed Method

Study	Key Features	Limitations Identified	How the Proposed Method Fill Some Research Gaps
Park [11]	Sends CCTV image via SMS	No facial/voice recognition, single webcam	Uses CNN-LSTM for robust face recognition with sequence input
Thabet [12]	PCA-based face recognition using OpenCV	Limited accuracy, lacks deep learning	Employs CNN-LSTM deep learning for higher accuracy
Ennis [14]	Image alerts for elderly users	No recognition or automation	Adds automated recognition and decision-making
Kumari [15]	Hearing-impaired alerts via Bluetooth	Manual review required, no AI-based recognition	Adds AI-powered recognition and automatic classification
Pinjala [17]	Live video with voice message option	No intelligence in visitor classification	Uses deep learning to classify known/unknown visitors
Pawar [19]	LBP-based recognition, SMS alerts	Traditional algorithms, prone to inaccuracies	CNN-LSTM improves recognition accuracy significantly
Giorgi [22]	Voice + iris recognition	Requires complex hardware, no cloud access	Uses common, low-cost hardware with edge computing

3. MATERIALS AND METHODS

This section provides a detailed overview of the materials and methodologies utilized throughout this study. It covers the selection and configuration of IoT devices, and sensors, and integrates these components to form a cohesive smart doorbell system. Specifically, it outlines the hardware employed, such as Raspberry Pi and ESP32-CAM [34] sensors, which serve as the backbone for image capture and real-time data processing. The section also explores the deep learning model design, focusing on the CNN-LSTM hybrid architecture used for accurate face recognition.

3.1. IoT Devices and Sensors

The hardware setup for this system includes an ESP32-CAM module and a Raspberry Pi. An LED display and a speaker are also incorporated to ensure user-friendly configuration and interaction. The Raspberry Pi, a cost-effective single-board computer, features a quad-core ARM Cortex-A72 processor and 8GB of RAM, making it powerful enough for a variety of IoT applications. Its extensive connectivity options, including USB ports, HDMI, and a 40-pin GPIO header, further enhance its suitability for such use cases. Meanwhile, the ESP32-CAM is a compact camera module equipped with a 2-megapixel OV2640 sensor, along with built-in Wi-Fi and Bluetooth capabilities, enabling seamless wireless communication. This combination of hardware offers both processing power and efficient communication, critical for real-time visitor recognition in the smart doorbell system. Moreover, the system employs MQTT with TLS/SSL encryption, ensuring that all transmitted data, such as facial recognition results, visitor notifications, and image uploads remain secure and protected from unauthorized access. When the system detects a visitor, it publishes an encrypted message containing the visitor's identity (or an alert for unknown faces) to an MQTT broker. The homeowner's mobile app, subscribed to the relevant topics, receives these updates in real-time. Additionally, authentication mechanisms like username-password pairs and certificate-based authentication further enhance security by preventing unauthorized devices from accessing the system. With MQTT's lightweight nature, the system ensures low-latency, energy-efficient communication, making it ideal for real-time IoT applications while maintaining strong privacy protections for user data.

3.2. Deep Learning Approach

The face recognition process in the proposed model involves several key stages: data acquisition, preprocessing, augmentation, and deep learning-based classification, as depicted in Figure 1. The model is specifically designed to deliver efficient and reliable face recognition by utilizing a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. This structured approach enhances the ability of model to accurately identify faces from a variety of input images with optimal performance.

3.2.1. Data Acquisition

The data collection process involved gathering images of known individuals captured by their smartphone cameras. Additionally, we sourced images from the social media profiles of these selected individuals, capturing them from various angles and under different lighting conditions in real-world environments. These collected images were then stored for further augmentation, preprocessing, and subsequent model training.

3.2.2. Data Preprocessing

The stored image data undergoes a preprocessing phase that includes converting the images into NumPy arrays, labeling them, shuffling to remove biases, and splitting the dataset into training and testing sets. These steps are crucial for efficiently preparing the data for the machine learning model, ensuring reliability. This process ensures that the dataset is balanced and ready for augmentation and training, ultimately enhancing the model's ability to generalize across diverse inputs.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

3.2.3. Data Augmentation

Data augmentation is a technique used to expand the training dataset by generating modified versions of existing data. This process involves applying small transformations to the dataset or creating new data points using deep learning methods, which enhances the diversity and robustness of the model during training. Rotation, cropping, and vertical flipping techniques are employed to expand the dataset with varied dimensions, helping to prevent overfitting. By training the model on this augmented dataset, it becomes better equipped to generalize to unseen data, leading to improved performance.

3.2.4. CNN-Based Feature Extraction

In our proposed model, the processed and augmented images are fed into a CNN-based mode to extract the features of images. The CNN model consists of multiple convolutional layers (Conv2D) with 32, 64, and 128 kernels used to analyze the images. Every kernel is followed by a Rectified Linear Unit (ReLU) activation function defined in *equation (1)*.

$$f(x) = \max(0, x) \tag{1}$$

Pooling layers, placed after convolutional layers, reduce the size of feature maps. Max pooling, the most common method, partitions images into 2 × 2 non-overlapping regions and selects the maximum value from each region, reducing feature map size by four times. This technique helps prevent overfitting and lowers computational cost by decreasing parameters. In contrast, average pooling calculates the average of the 2 × 2 regions instead of selecting the maximum value, providing another method of subsampling. The max pooling approach can be defined in *equation (2)*:

$$MaxP(x)_{i,j,l} = {}^{max}_{a,b} X_{i.S_x+a,j.S_y+b,l} \tag{2}$$

where X is input, S_x= horizontal stride and S_y= vertical stride.

The Flatten layer, which comes at the end of the feature extractor, converts the two-dimensional feature maps into a one-dimensional vector so that it may move on to the following stage.

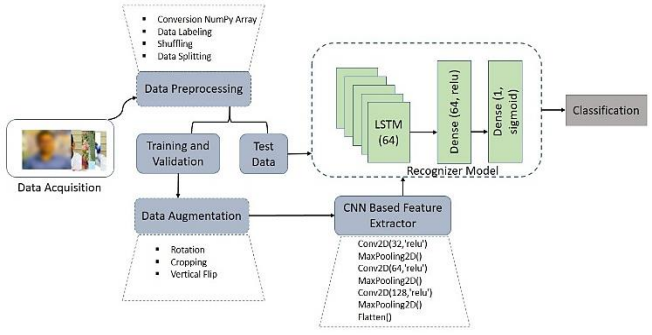


Figure 1. An abstract design of the proposed deep learning model

3.2.5. Recognizer Model (CNN-LSTM Hybrid)

In this step, the feature vector generated is passed to a hybrid model incorporating 64 LSTM units, which capture temporal dependencies in the data, improving the model's ability to recognize patterns within image features. After processing through the LSTM layer, the model transitions to a dense layer with 64 units and ReLU activation. Finally, the image is classified as belonging to either a known or unknown person via a dense output layer with a sigmoid activation function, as defined in *equation (3)*. This robust recognizer, combining CNN and LSTM networks, achieves high accuracy in face recognition applications.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3}$$

The final output of the proposed deep learning model is to classify the images as known or unknown visitors. This information is then provided to the user after completing the identification process. The overall architecture ensures an efficient model for detecting visitors.

Overall, the stored dataset undergoes pre-processing and augmentation before being input into a CNN model to extract image features. These extracted features are then passed to the proposed recurrent neural network, which ultimately determines whether the image belongs to a known or unknown individual.

3.3. Deep Proposed Framework

The proposed intelligent doorbell system is designed for addressing real-time recognition of visitor face using IoT devices and deep learning model. This framework involves three primary phases: data gathering, model training, and recognition. Firstly, the camera module captures the facial image of visitor which is then stored in the cloud-based dataset. Secondly, the proposed deep neural network-based recognizer model is learned using the dataset gathered in the previous step. In the final phase, the system takes a new photo of the visitor and compares it with the trained dataset with the help of the recognizer model. If the newly capture image is matched with existing image dataset, then the system assigns an ID to it. If the face is not matched, the image is sent to the user mobile application using a cloud server and an alert is generated for further action. Therefore, the proposed framework ensures real-time identification and notification, which make it more efficient for home security and monitoring application.

Figure 2 illustrates the flowchart of the proposed system, outlining each process step by step. Initially, the camera captures an image of the visitor, and the system detects and normalizes the face to ensure uniform size and orientation for subsequent analysis. The system then extracts unique facial features from the image, which are compared against a database of registered individuals. If the system recognizes the face, it generates a face ID, and the person's name, along with the timestamp, is sent to a real-time database. This information is then transmitted to a mobile application, where the user receives a notification with the details.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

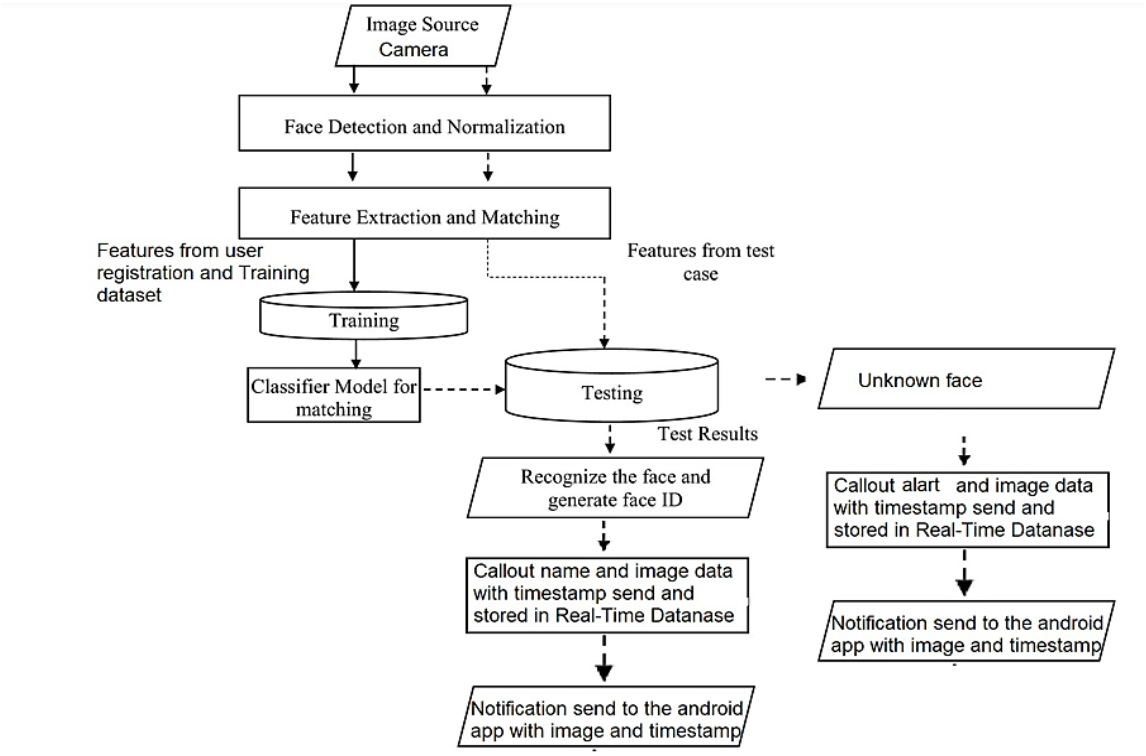


Figure 2. A flowchart of the proposed smart doorbell

If the face is unrecognized, the system triggers an alert. An image of the unknown person, along with a timestamp, is sent to the real-time database, and the mobile app is notified to take further action. This entire flow ensures seamless real-time face recognition, distinguishing between known and unknown individuals, and promptly updates the user via mobile notifications.

Figure 3 presents the architecture of a smart doorbell system for user recognition and instant notifications, divided into three main phases: Data Gathering, Training the Recognizer, and Recognition. Initially, Data Gathering starts with capturing images of individuals using a camera module attached to the Raspberry Pi. These image files are stored in a database along with user identification numbers. The next step is to train the recognizer by transmitting the collected dataset to a remote database, where a model is trained using OpenCV Python libraries to recognize known faces. This model is built by the trainer based on the accumulated dataset. During the recognition process, when a person stands at the doorbell, the system identifies them if they are recognized. If the person is unknown, a new photograph is captured and uploaded to the database for future recognition. The user's mobile application receives notifications with pictures of the visitors for confirmation. Additionally, the system allows real-time monitoring and control of visiting individuals from a distance, significantly enhancing home security with minimal wiring required.

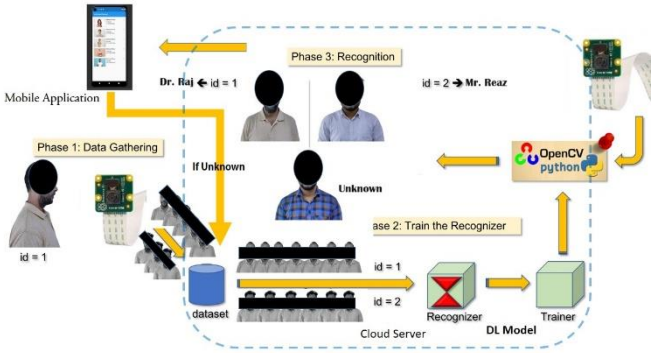


Figure 3. Architecture of the proposed doorbell system

3.4. Performance Evaluation Metrics

Before implementing the proposed model on simulated data, it is important to validate it using an existing dataset and various evaluation metrics. We assess the model's performance through key metrics such as Accuracy (4), Precision (5), Recall (6), and F1-score (7), utilizing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These metrics are critical for evaluating the system's overall reliability and accuracy in facial recognition.

$$Aaccuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (4)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

$$F1 - score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (7)$$

4. RESULTS AND DISCUSSION

In this section, we identify the most effective model for an IoT-based intelligent doorbell by analyzing the performance of various deep learning models. First, we develop an IoT-based unknown visitor detection model using the proposed framework. The collected dataset is preprocessed and augmented, and the samples are randomly divided into training and testing sets. During training, we calculate the validation loss to ensure it does not increase as the training loss decreases. The applied models are evaluated using metrics such as accuracy, precision, recall, and F1-score to demonstrate the efficiency of the proposed model in real-world scenarios.

4.1. Experimental Setup

To evaluate the proposed system, the experimental setup plays a critical role as it integrates both hardware and software components. The hardware consists of the Raspberry Pi, serving as the primary processing unit, and the ESP32-CAM camera module functioning as an IoT device. The basic circuit diagram is shown in *figure 4*. On the software side, the system includes the proposed hybrid deep learning model and a mobile application as core components. The system configuration for the study was an 8th-generation Intel Core i7 processor (6600U) clocked at 3.1 GHz, with 16 GB of RAM. For software, Python and OpenCV were used for image processing and face recognition, with data transmitted and stored in the cloud. All experiments were carried out in a controlled environment to ensure reliable data collection and analysis.

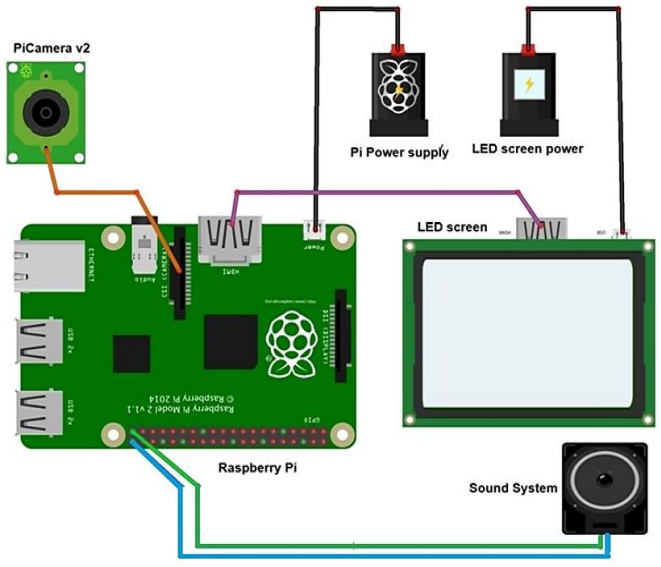


Figure 4. Circuit Diagram of the proposed system

Table 2 summarizes the key training parameters used to ensure reproducibility of the proposed CNN-LSTM model. The model is trained using the Adam optimizer, which is well-suited for adaptive learning rate adjustments. The binary crossentropy loss function is employed, appropriate for binary classification tasks such as identifying known vs. unknown visitors. The model is trained for 30 epochs with a batch size of 32, which balances training speed and performance. The sequence length is kept variable, allowing the system to adapt to different input durations, making it more flexible in handling varying numbers of sequential image frames. These settings collectively support stable and repeatable training outcomes.

Table 2. Training Parameters for Reproducibility

Parameter	Value
Optimizer	Adam
Loss Function	Binary Cross entropy
Epochs	10
Batch Size	32
Sequence Length	Variable

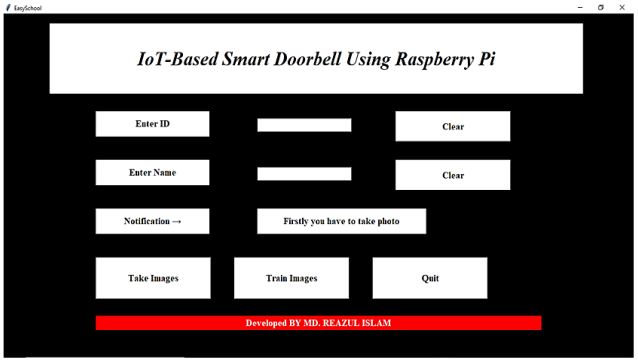


Figure 5. Data Collection system User Interface

4.2. Data Collection and Augmentation

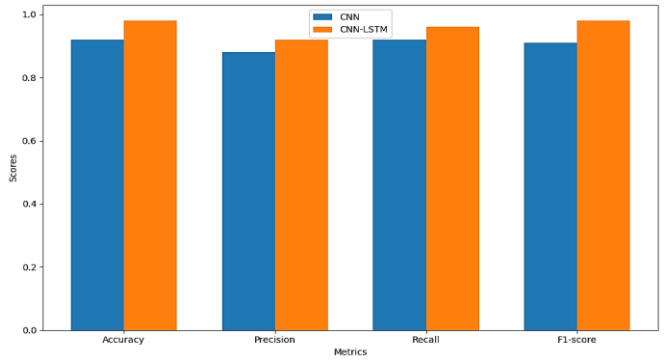
To train and evaluate our proposed CNN-LSTM model, we initially collected 225 images of a known person. These images were captured from five different perspectives: front, left, right, up, and down. This ensured the dataset contained sufficient variability in facial orientations, which is critical for accurate recognition in real-world conditions. We use user registration to train our system, as shown in Figure 5. To do this, we provide the user's name and ID. Then, pictures of the person are taken to train the system. Once someone is registered, the system will recognize them as a known visitor. To make the data more varied and avoid overfitting, we applied data augmentation techniques like rotating, cropping, and flipping the images to increase the dataset. Data augmentation is a common practice in machine learning, where new data is virtually generated from existing samples to improve the model's robustness. Following these augmentations, the number of images increased from 225 to 1,800. The dataset was then split into two parts: 80% (1,440 images) for training and 20% (360 images) for testing. This approach ensures the model is evaluated on an entirely separate dataset, minimizing bias and producing reliable performance metrics.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

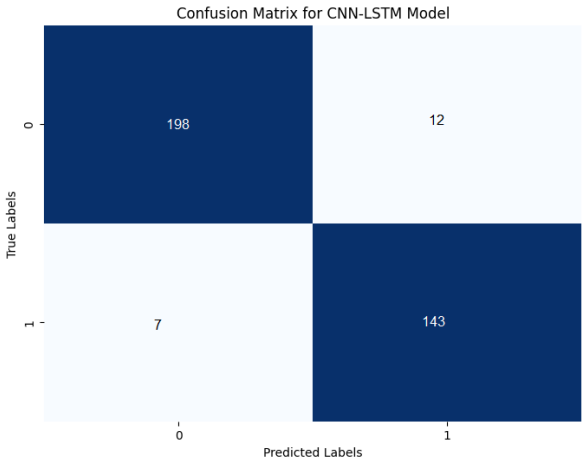
To further evaluate the efficiency and generalization capability of the proposed model, an additional dataset [36] was utilized. The preprocessing phase involved systematically cropping the facial regions from the original images to ensure consistency and focus on relevant features. Duplicate images were identified and removed to eliminate redundancy and avoid bias during training. Following preprocessing, data augmentation techniques such as rotation, flipping, and scaling were applied to enhance variability within the dataset. This process resulted in an expanded and balanced dataset comprising 10,827 images, with 8,661 samples allocated for training and 2,166 for testing. These steps ensured that the model was trained on a diverse and representative dataset, ultimately improving its robustness and performance in real-world conditions.

4.3. Performace Evaluation

In this section, we evaluate the performance of our proposed model using various metrics such as accuracy, precision, recall, and F1-score. First, we assess whether the applied machine learning models can effectively differentiate between known and unknown individuals. Second, we explain how the proposed model was trained using the collected dataset. Additionally, we provide an overview of the mobile application associated with this smart system. Finally, we compare our proposed model to other existing systems, highlighting both the strengths and limitations of each approach.



(a)



(b)

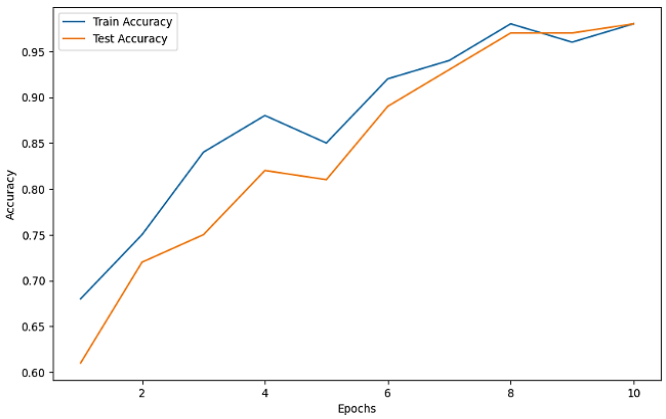
Figure 6. Comparison of CNN and CNN-LSTM; (a) Performance on Accuracy, Precision, Recall, and F1-Score; (b) Confusion Matrix of CNN-LSTM model

Table 3 provides information about the applied models to indentify which model is outperformer to recognize the visitors. The proposed hybrid model surpasses the traditional CNN model by 5.7% in accuracy, 5% in both precision and recall, and 8% in F1-score. *Figure 6(a)* figures out a comparison between two models, CNN and CNN-LSTM, in terms of accuracy, precision, recall, and F1-score. The CNN model achieves an accuracy of 92%, while the CNN-LSTM model performs better with 98% exceeding the CNN model. For precision, CNN scores 88%, but CNN-LSTM improves to 93%, showing it is better at identifying correct positives. It checks how well the model avoids false positives. Only 7% vistors can be wrongly detected by our proposed model. In terms of recall, CNN gets 92.3%, while our proposed model (CNN-LSTM) improves to 97%, indicating it can detect more correct cases than the CNN model. Moreover, our proposed model outperfroms with 99% of F1-score showing overall bertter performance, while CNN achieves 91% in terms of F1-score. This indicates that the CNN-LSTM model consistently performs better than CNN, especially in accuracy and F1-score, making it more reliable for handling complex data.

Table 3. Performance comparison of the applied DL models

Model	Accuracy	Precision	Recall	F1-score
CNN	92.3%	88%	92%	91%
EfficientNetB4	95.5%	91.5%	96%	95%
CNN-LSTM	98%	93%	97%	99%

The confusion matrix depicted in *figure 6(b)* summarizes the model's classification performance by showing how well it distinguishes between the positive and negative classes. In this case, the matrix indicates that the model correctly identified 198 instances as positive, while it incorrectly predicted 12 instances as positive when they were actually negative. Additionally, 7 instances that were truly positive were incorrectly classified as negative, and 143 instances were accurately predicted as negative. Overall, the model demonstrates strong performance with most predictions being correct and only a few misclassifications.



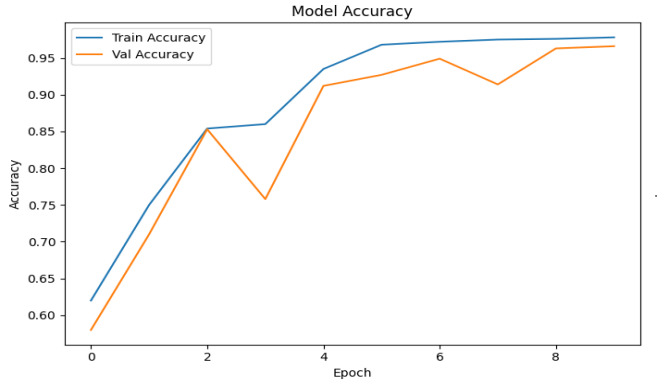
(a)

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

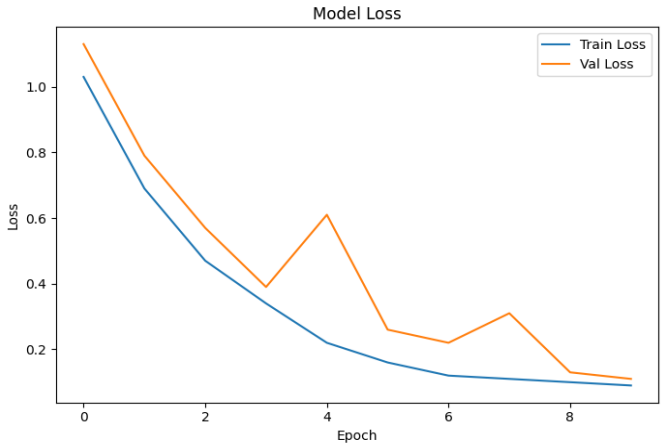
(b)

Figure 7. Learning Curve of our proposed model: (a) Accuracy graph of CCN-LSTM for training and test data; (b) Loss graph of CCN-LSTM for training and test data

Figure 7(a) demonstrates the accuracy metrics of the model over 10 epochs, with accuracy on the y-axis and the number of epochs on the x-axis. The blue curve represents the training accuracy, which begins at 68% and reaches 98%, though it fluctuates to 85% around the 5th epoch, showing that the model steadily improves in classifying the training data. Similarly, the orange curve, indicating testing accuracy, shows a strong progression from 61% to around 98%, demonstrating significant improvement over time. The diagram highlights a balanced model training approach, where achieving high training accuracy does not compromise the model's ability to generalize well on testing data, ensuring dependable performance for real-world applications in the smart doorbell system. *Figure 7(b)* depicts the training and test loss curves for the proposed CNN-LSTM model over 10 epochs. Both the training loss (in blue) and the test loss (in orange) show a consistent downward trend, indicating that the model is learning effectively as the loss decreases with each epoch. Initially, the test loss starts higher than the training loss, reflecting the expected behavior of the model performing better on training data early in the process. Over time, both curves decrease steadily, with occasional fluctuations, suggesting that the model is gradually minimizing the loss for both training and test sets. By the end of the training, the gap between the training and test loss narrows, which indicates that the model is generalizing well without overfitting. This figure demonstrates that the CNN-LSTM model successfully reduces error over the course of training, validating its performance on the task.



(a)



(b)

Figure 8. Learning Curve of our proposed model using the dataset [36].

Figure 8 presents the training and validation performance of the proposed model over 10 epochs. The left plot shows a steady increase in both training and validation accuracy, with the model achieving over 96% accuracy by the final epoch, indicating strong learning capability. The right plot displays a significant decrease in both training and validation loss, suggesting effective convergence of the model. Although there are minor fluctuations in validation performance, both accuracy and loss curves indicate that the model generalizes well without significant overfitting, validating its robustness and reliability in classification tasks.

Table 4. Classification performances of the proposed model for dataset [36]

Class	Precision	Recall	F1-Score	Support
Known	0.96	0.97	0.97	1100
Unknown	0.95	0.94	0.95	1066
Accuracy			0.961	2166
Macro Avg	0.96	0.96	0.96	2166
Weighted Avg	0.96	0.96.	0.96	2166

The precision and recall scores shown in *table 4* demonstrate that the proposed model effectively distinguishes between known and unknown individuals, with minimal false positives and false negatives. An F1-score of 0.96 further indicates a well-balanced trade-off between precision and recall, reinforcing the model's reliability in binary classification tasks. These outcomes are consistent with the validation accuracy of approximately 96% observed in the performance curves, confirming the model's robustness and high predictive quality in real-world scenarios.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

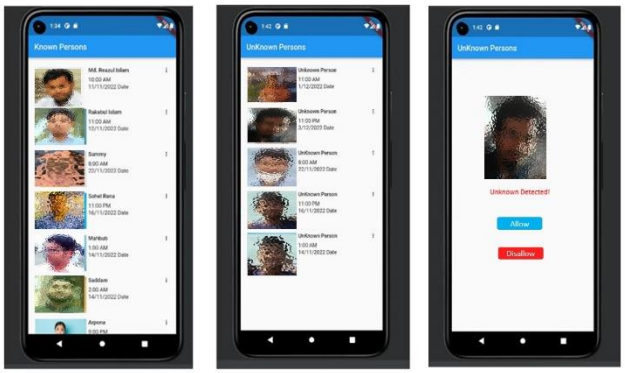


Figure 9. Mobile application interface of the proposed system

The system provides continuous tracking and oversight of all activities to the owner through a mobile application. It allows the owner to review the list of registered users and monitor visitors in real time shown in *figure 9*. When a visitor approaches, the system automatically captures their image and uses a pre-trained deep learning model to identify whether the person is known or unknown. If recognized, the visitor's name is included with the photo and sent to the owner's mobile application. If the visitor is unrecognized, a notification alerts the owner of the unknown individual. In special cases, the owner can grant emergency access to the visitor *via* the mobile application.

The real-time performance of our proposed system was evaluated through 70 live test cases, where it successfully identified visitors 67 times, achieving an accuracy of 95.7%. These results demonstrate the system's reliability in correctly distinguishing between known and unknown individuals in real-world conditions. The high success rate highlights the effectiveness of our CNN-LSTM hybrid model, ensuring accurate face recognition with minimal errors. This real-time evaluation further validates the system's robustness, efficiency, and practicality for smart home security applications.

The inference time of our proposed system ranges between 0.3 to 0.5 seconds, ensuring real-time face recognition with minimal delay. This fast processing speed demonstrates the efficiency of our CNN-LSTM hybrid model, which balances accuracy and computational performance. Unlike cloud-based systems that rely on external servers, our approach processes data locally on a Raspberry Pi, reducing network latency and ensuring instant visitor identification. The low inference time makes our system highly suitable for real-time security applications, providing quick notifications and responses while maintaining energy efficiency on resource-constrained IoT devices.

The proposed DL-based Intelligent Doorbell system is exceptional because of its high accuracy and other features. It is comparable with high performing model such as YOLO (98.27% accuracy) [26] with accuracy of 98%. Moreover, it outperfomes the LSTM, DeSGCBQNN and other existing models mentioned in the Table 5 in terms of efficiency to detect the visitors. Additionally, it integrates the emerging technology IoT and DL model that enhance its capability and functionality. In comparison with many other models, our approach integrates a mobile application, providing real-time tracking and alerts, which is crucial for practical face recognition systems. The mobile application is user-friendly with capability of real-time monitoring which make this system outperformer among other existing systems. In this way, our proposed system is now more adaptable and useful in real-world situations, particularly for security and monitoring needs, thanks to this combination. Overall, this system is positioned as a superior solution because to its complete approach, which bridges the gap between user accessibility, accuracy, and technology integration.

Table 5. Comparison with the existing studies

Ref.	Model	Dataset	Accuracy	IoT Enabled	Mobile Application
[22]	EfficientNet B4	Collected	97%	Yes	No
[23]	Alexnet	Collected	97.5%	No	No
[24]	Federated DL	-	-	Yes	Yes
[25]	DeSGCBQNN	UNSW-NB15 dataset	95%	Yes	No
[26]	YOLO	Roboflow	98.27%	Yes	No
[27]	LSTM model	MobiAct	95.87%	Yes	No
Proposed	CCN-LSTM	Collected	98%	Yes	Yes
	CCN-LSTM	Human Faces [36]	96.1%	Yes	Yes

4.4. Comparative Analysis

Our proposed IoT-based smart doorbell system is more cost-effective than the Nest Hello, as it is built using affordable components like a Raspberry Pi and ESP32-CAM. Unlike Nest Hello [35], which requires a subscription for advanced features such as facial recognition and cloud storage, our system can process and store data locally, reducing long-term expenses. Additionally, our system is less resource-intensive, running efficiently on low-power hardware while utilizing MQTT for lightweight communication, making it ideal for IoT applications. By leveraging edge computing, we minimize reliance on cloud-based processing, which reduces latency and operational costs.

In terms of accuracy and speed, our system uses a CNN-LSTM hybrid model, achieving 98% accuracy in facial recognition. While Nest Hello employs Google's AI-powered facial recognition, its exact accuracy is undisclosed. Our system's flexibility allows further optimization using public datasets like VGGFace and CelebA, potentially enhancing performance even more. While Nest Hello benefits from Google's cloud infrastructure, which may provide faster processing, our system's local processing reduces dependency on external servers, improving real-time responsiveness. With additional improvements in dataset diversity and hardware optimization, our system can match or even surpass commercial alternatives

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 2 | Pages 353-363| e-ISSN: 2347-470X

while ensuring data privacy and affordability. *Table 5* highlights how our system is more cost-effective, customizable, and privacy-focused, while Nest Hello offers seamless cloud integration but comes with higher costs and reliance on Google's ecosystem.

Table 5. Comparison table between our proposed IoT-based smart doorbell system and the Nest Hello

Feature	Proposed IoT-Based System	Nest Hello (Google Nest Doorbell)
Cost	Lower, built with Raspberry Pi & ESP32-CAM	Higher, premium hardware & subscription fees
Subscription	No subscription required	Requires subscription for advanced features (e.g., facial recognition)
Processing	Local (Edge Computing)	Cloud-based (Google AI)
Resource Efficiency	Optimized for low-power devices	Requires high-performance servers
Accuracy	98% (CNN-LSTM model)	Unspecified (Google AI)
Latency	Lower (local processing reduces delays)	Dependent on internet & cloud processing speed
Privacy & Security	Local data storage, secure MQTT communication	Cloud storage, potential data privacy concerns
Customization	Fully customizable & open-source	Limited customization options
Real-Time Alerts	Instant notifications via MQTT & mobile app	Alerts via Google Home app
Hardware Dependence	Works on Raspberry Pi, ESP32-CAM	Proprietary Google hardware
Internet Dependency	Can function offline with local processing	Requires an internet connection for cloud features

Moreover, our proposed system stands apart from YOLO-based [26], LBP [19], and OpenCV-powered [12] smart doorbells through its innovative hybrid architecture and practical deployment advantages. Technically, it leverages a CNN-LSTM model that processes sequences of image frames, allowing it to learn temporal patterns and make more accurate decisions compared to traditional models that analyze static images. This sequence-based recognition enhances robustness against variations in lighting, pose, and facial occlusion. Unlike computationally intensive YOLO models, our system is optimized for low-resource environments like Raspberry Pi, achieving an efficient inference time of 0.3–0.5 seconds. Practically, it incorporates secure, real-time communication using the MQTT protocol, enabling fast notifications to users' mobile devices. Additionally, the system is highly cost-effective, utilizing affordable hardware such as Raspberry Pi and ESP32-CAM, and its performance is bolstered through data augmentation techniques that improve generalization. These combined strengths make our system not only technically superior but also more feasible for real-world deployment than conventional smart bell solutions.

5. CONCLUSIONS

In this study, a smart doorbell system is proposed and implemented that leverages DL and IoT technologies to enhance home security through real-time visitor recognition and instant notifications. The system integrates a Raspberry Pi and ESP32-CAM to provide an inclusive security solution. A camera sensor is used to capture images in front of the doorbell, which are then processed by the Raspberry Pi and sent to the server. The pre-trained DL model kept in remote server checks the captured image whether it is known or unknown; if it is unknown, a new photo is taken. The proposed framework effectively classifies visitors as known or unknown by utilizing a pre-trained deep learning model, delivering real-time notifications to the homeowner's mobile application. The proposed hybrid model combining CNN and LSTM achieved 98% accuracy and outperformed other performance metrics. This system offers real time recognition of visitors and triggers a notification if the visitor is identified as unknown providing an additional layer of security.

In conclusion, the proposed solution not only achieves high performance in visitor recognition but also offers a scalable, low-cost, and reliable system for home security. In future, we can continue to enhance the capabilities of the system by integrating more advanced features such as voice recognition, multi-camera support and contribute to the advancement of home security.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, M.R.I. and K.O.; methodology, M.R.I and K.O.; software, K.O.; validation, S.D.J., R.R.C. and M.F.R..; formal analysis, R.R.C.; investigation, R.R.C.; resources, R.R.C.; data curation, K.O.; writing—original draft preparation, S.D.J.; writing—review and editing, M.F.R.; visualization, M.R.I.; supervision, M.F.R.; project administration, M.R.I. All authors have read and agreed to the published version of the manuscript".

Funding Source: This research received no external funding

Acknowledgments: The authors would like to thank to American International University-Bangladesh, Comilla University and Bangladesh University of Business and Technology for their support and resources.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Anupriya, S.R. and Muthumanikandan, V., A survey on exploring the effectiveness of iot based home security systems. In 2023 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-10). IEEE. 2023.

[2] Rahim, A., Zhong, Y., Ahmad, T., Ahmad, S., Pławiak, P. and Hammad, M., Enhancing smart home security: anomaly detection and face recognition in smart home IoT devices using logit-boosted CNN models. Sensors, 23(15), p.6979. 2023.

[3] Rajeswari, Vinod Kumar, N., Suresh, K.M., Sai Kumar, N. and Girija Sravani, K., IoT-Based Smart Home Security Alert System for

Continuous Supervision. Machine Learning for VLSI Chip Design, pp.51-63. 2023.

[4] Sayeduzzaman, M., Hasan, T., Nasser, A.A. and Negi, A., An Internet of Things-Integrated Home Automation with Smart Security System. Automated Secure Computing for Next-Generation Systems, pp.243-273. 2024.

[5] Sattaru, P.K., Burugula, K.V., Channagiri, R. and Kavitha, S., 2023, January. Smart Home Security System using IoT and ESP8266. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 469-474). IEEE.

[6] Islam, M. R., Oliullah, K., Kabir, M., Rahman, A., Mridha, M. F., Khan, M. F., & Dey, N. Machine learning-driven IoT device for women's safety: a real-time sexual harassment prevention system. Multimedia Tools and Applications, 1-30. 2024.

[7] Singh, O., Singh, N., Singh, A., & Vinoth, R. AI, IoT, and Blockchain in Fashion: Confronting Industry Applications, Challenges with Technological Solutions. International Journal of Communication Networks and Information Security, 16(4), 393-410. 2024.

[8] Salvi, S., Dhar, R., & Karamchandani, S. (2021). IoT-Based Framework for Real-Time Heart Disease Prediction Using Machine Learning Techniques. In Innovations in Cyber Physical Systems: Select Proceedings of ICICPS 2020 (pp. 485-496). Springer Singapore.

[9] Force, O.C.T., 2023. Annual Report and Threat Assessment.

[10] Verma, Y.K., Yarlagadda, N., Bhandari, J.K. and Dheep, R., 2024. Smart Door Bell Using IoT: Implementation and Design. In Intelligent Circuits and Systems for SDG 3–Good Health and well-being (pp. 393-402). CRC Press.

[11] Park, W.H. and Cheong, Y.G., IoT smart bell notification system: Design and implementation. In 2017 19th International conference on advanced communication technology (ICACT) (pp. 298-300). IEEE. 2017.

[13] Thabet, A.B. and Amor, N.B. Enhanced smart doorbell system based on face recognition. In 2015 16th international conference on sciences and techniques of automatic control and computer engineering (STA) (pp. 373-377). IEEE, 2015.

[14] Baobaid, A., Meribout, M., Tiwari, V.K. and Pena, J.P., Hardware accelerators for real-time face recognition: A survey. IEEE Access, 10, pp.83723-83739. 2022.

[15] Ennis, A., Cleland, I., Patterson, T., Nugent, C.D., Cruciani, F., Paggetti, C., Morrison, G. and Taylor, R, Doorstep: A doorbell security system for the prevention of doorstep crime. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 5360-5363). IEEE, 2016.

[16] Premsai, I. and Thiyagu, T.M, IoT based Wireless Alert System for Individuals with Impaired Hearing. In 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 662-666). IEEE, 2024.

[18] Baballe, M.A., Accident Detection System with GPS, GSM, and Buzzer. TMP Universal Journal of Research and Review Archives, 2(1), pp.28-36. 2023.

[19] Pinjala, S.R. and Gupta, S., Remotely accessible smart lock security system with essential features. In 2019 international conference on wireless communications signal processing and networking (WiSPNET) (pp. 44-47). IEEE. 2019.

[20] Sayeduzzaman, M., Hasan, T., Nasser, A.A. and Negi, A., An Internet of Things-Integrated Home Automation with Smart Security System. Automated Secure Computing for Next-Generation Systems, pp.243-273. 2024.

[21] Pawar, S., Kithani, V., Ahuja, S. and Sahu, S., Smart home security using IoT and face recognition. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE, 2018.

[22] Mostakim, M.N., Mahmud, S., Jewel, M.K.H., Rahman, M.K. and Ali, M.S., Design and development of an intelligent home with automated environmental control. International Journal of Image, Graphics and Signal Processing, 10(4), p.1. 2020.

[23] Eras, L., Domínguez, F. and Martinez, C., Viability characterization of a proof-of-concept Bluetooth mesh smart building application. International Journal of Distributed Sensor Networks, 18(5), p.15501329221097819. 2022.

[24] Giorgi, R., Bettin, N., Ermini, S., Montefoschi, F. and Rizzo, A., An iris+ voice recognition system for a smart doorbell. In 2019 8th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE. 2019

[25] Chauhan, K. and Chauhan, R.K., Design and development of two levels electronic security and safety system for buildings. International Journal of Electronic Security and Digital Forensics, 12(3), pp.279-292. 2020.

[26] Anu and Bhatia, D., 2014. A smart door access system using finger print biometric system. International Journal of Medical Engineering and Informatics 2, 6(3), pp.274-280.

[27] Ali, H.H., Naif, J.R. and Humood, W.R., A new smart home intruder detection system based on deep learning. Al-Mustansiriyah Journal of Science, 34(2), pp.60-69. 2023.

[28] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 11, pp. 408-412, 2020

[29] Gudipalli, A., Kejriwal, V., Patel, V., Gupta, R. and Dixit, U., COVID bell–A smart doorbell solution for prevention of COVID-19. Paladyn, Journal of Behavioral Robotics, 14(1), p.20220115. 2023.

[30] Gaikwad, V., Rathi, D., Rahangdale, V., Pandita, R., Rahate, K. and Rajpurohit, R.S., Design and Implementation of IOT Based Face Detection and Recognition. Computing & Intelligent Systems, pp.923-933. 2024.

[31] Dalal, S., Lilhore, U.K., Sharma, N., Arora, S., Simaiya, S., Ayadi, M., Almujally, N.A. and Ksibi, A., Improving smart home surveillance through YOLO model with transfer learning and quantization for enhanced accuracy and efficiency. PeerJ Computer Science, 10, p.e1939. 2024.

[32] Kulurkar, P., Dixit, C.K., Bharathi, V.C., Monikavishnuvarthini, A., Dhakne, A., Preethi, P., AI based elderly fall prediction system using wearable sensors: A smart home-care technology with IOT, Measurement: Sensors,Volume 25, p.100614. 2023.

[33] Chaudhari, D.A. and Umamaheswari, E., Technology-Enabled Medical IoT System for Drug Management. International Journal of Communication Networks and Information Security, 16(1), pp.19-31. 2024.

[34] Raihan, P.A., Tullah, R., Julianti, M.R., Ramdhan, S., Ak, M.F. and Fazilla, S., IoT-Based Biometric Attendance System Using Arduino and ThingsBoard. International Journal of Communication Networks and Information Security, 15(4), pp.103-117. 2023.

[35] Tello, A.B., Xing, J., Patil, A.L., Patil, L.P. and Sayyad, S., Blockchain Technologies in Healthcare System for Real Time Applications Using IoT and Deep Learning Techniques. International Journal of Communication Networks and Information Security, 14(3), pp.257-268. 2022.

[36] Salikhov, R.B., Abdrakhmanov, V.K. and Safargalin, I.N., Internet of things (IoT) security alarms on ESP32-CAM. In Journal of Physics: Conference Series (Vol. 2096, No. 1, p. 012109). IOP Publishing. 2021.

[37] White, C., & Gilmore, J. N. Imagining the thoughtful home: Google Nest and logics of domestic recording. Critical Studies in Media Communication, 40(1), 6-19. 2023

[38] Human faces. (2020, September 21). Kaggle. https://www.kaggle.com/datasets/ashwingupta3012/human-faces.