Grenze International Journal of Engineering and Technology, June Issue

AI-Enabled Multi-Modal Smart Door Access System using Face Voice and Gesture Recognition

Abhishek Tripathi[1], Abhilash Mandloi[2], Chirukuri Manohar[3], Suryavamshi Dasharath[4], Manchala Abhishek[5] and Robba Hemanth Babu[6]
[1,3-6]Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Tamil Nadu, India
Email: tripathi.abhishek.5@gmail.com, 99220041064@klu.ac.in, 99220041008@klu.ac.in, 99220040913@klu.ac.in, 99220040714@klu.ac.in
[2]Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India
Email: asm@eced.svnit.ac.in

***Abstract*— The Multi-Mode Smart Door Access System planned to be designed will employ an arduino control system and ESP32-CAM to ensure that the authentication is as secure as possible without touching anything. It accomplishes this through voice checks and face checks. The mean response time was 1.8s and accuracy was 97%. False Acceptance rate (FAR) was 2 and the False Rejection rate (FRR) was 1.5. The experiment revealed that ArcFace (AUC = 0.95) was more effective than other models in observing things. The system performed good in varying degrees of light and noise, which demonstrated that it will be able to perform in intelligent homes and other settings in real time. This project is similar to SDG 9 (Industry, Innovation, and Infrastructure) and SDG 11 (Sustainable Cities and Communities).**

***Index Terms*— Internet of Things, deep learning, security, and authentication.**

I. INTRODUCTION

AI (artificial intelligence) and ML (machine learning) are becoming increasingly better. Things are now seen and handled differently. They can now be smartly automated in such fields as farming, health care and security [1], [2]. The most prevalent methods of entering a room are an RFID tag, key, or PIN code. Since these techniques are becoming simpler to duplicate and abuse, it is further crucial to possess security systems that adopt more than one technique and have the ability to alter when new data is acquired. The above-presented issues are resolved with the proposed Multi-Modal Smart Door Access System, which is based on a single smart authentication system that can distinguish between voices, faces, and gestures. Such combination allows easier identification of users, prevents the situation when they pretend to be someone they are not, and makes the site more natural and friendly. It can know who you are by seeing your face, hearing your voice and seeing the way you move. It achieves this through deep learning, voice recognition, and the observation of the movement of people with their hands. All of these things can be accomplished by lightweight ML algorithms. They ensure that the system is running on real time on low-cost hardware which already exists in other systems. The techniques have worked in the environment where IoT and real-time recognition are used [3]-[5]. The system has developed an easy and non-contact access model which is safe and well suited both in homes and hospitals as well as in any other place people work and live.

Grenze ID: 01.GIJET.12.2.21_22

It employs AI, Internet of Things, and computers, which are embedded in other devices. It reflects the transformation of the previous user-checking system, where the single approach was used, to the more open and adaptable security model that aligns with the objectives of the smart and long-term digital infrastructures.
The paper is arranged in the following manner: Section II will consider the innovative door access systems, which use the Internet of Things and other technologies. In Section III, we discuss the suggested system methodology and the way to prepare to multimodal fusion. In the fourth section, we discuss the interaction of the hardware and the software. Section V discusses the tests along with what they revealed. In part VI, we conclude the work and discuss how we might do it better in future

II. Literature Review

Over the years, smart door access systems have been improved with the introduction of biometric authentication, IoT, and embedded computing. Initial studies had indicated that machine learning and data-driven approaches play well in embedded systems, and the smart decision-making paradigm was adopted in real-time systems [1]. In order to enhance the reliability of the access, high-tech sensing systems, such as the facial recognition system that is based on radar and cameras, were introduced. Nonetheless, this complicated the system [2]. Vision-based authentication has been researched on a lot. As an example, Tripathi et al. [3] demonstrated that facial recognition was capable of operating on small systems, whereas Franco et al. demonstrated that activity and behavior recognition was capable of operating in smart environments.[4]. Single-factor biometric systems are not going away since they are very easy to use. The sensors employed by Goyal et al. [5] to control the door opening were fingerprint-based IoT door locks, which performed effectively in controlling access but slowed down during bad weather. Gudavalli et al. [6] recommended a two-stage verification to ensure the safety of things. They demonstrated that layered authentication is more resistant than those mechanisms that are independent. Guntur et al. researched on development of IoT enabled door locking system, which could be tracked and real time alerts raised. The latter, [7] and Hashim et al. [8] demonstrated the method of Raspberry Pi and Bluetooth-based smart locks that are inexpensive and simple to operate in home automation. In reference to Huang et al. [9], alternative access methods were proposed, such as using visible light to authenticate keys, whereas in reference to Alice et al. [10], low-powered microcontrollers were used in security door systems.
In order to become safer and have more possibilities to monitor the situation remotely, several researchers have investigated the possibilities of IoT-enabled smart door access and home security systems. Ramakrishna et al. [11] have shown a complex IoT-based home security system which can enable the control of the home remotely through smartphone and real time alerts. The reliability of the system in uncontrolled environments was however limited because the system depended on restricted authentication schemes. Salamah et al. adopt the face recognition option and voice command to enhance the accessibility of the system thus rendering it more useful to individuals with mobility limitations [12]. Nevertheless, the system performance may be poor in noisy environment or in case of facial occlusions. Umamaheswari et al. applied special authentication parameters and advanced matching algorithms to enable real-time decision making, but could not fight against spoofing and sensor malfunction due to the use of only one biometric modality [13] There are also reports of efforts to optimize authentication on low-power embedded systems. Wangean et al. demonstrated the ability to use Haar Cascade with OpenCV LBPH on resource constrained devices to achieve facial recognition but the accuracy was highly dependent on the change in pose and lighting [14]. Al-Mutawa et al. did not give strong biometric authentication as the primary concern of their comprehensive internet of things-based home automation model, which incorporated multiple sensors to enhance the comfort and security of the users [15]. Nandhini et al. enabled even smarter surveillance by integrating facial recognition with weapon detection as well as by prioritizing the detection of threats over ensuring safe access control [16].
Although Subiyanto et al. improved the accuracy of facial recognition with the help of a PCA-GA algorithm running on Raspberry Pi platforms [17], [18], Sandesh et al. designed a smart lock system on Raspberry Pi platforms, which showed flexibility and could be easily deployed to automate homes. Although the authors highlight the importance of IoT-controlled surveillance in institutional contexts, the automated laboratory access system introduced by Vishnoi et al. in the controlled environment is based on a limited number of authentication aspects [19]. Developed automated laboratory access structure into a controlled environment, which highlights the importance of IoT-based surveillance in an institutional setting. These findings demonstrate that multimodal biometric systems involving visual, auditory and motion-based information are the most effective method of obtaining secure, contactless, and unrestricted access. This research enhances these technological innovations with a multi-modal authentication mechanism that is capable of identifying faces, voices as well as gestures. This renders them to be more reliable, easy to use and flexible in real time.

III. METHODOLOGY

The Multi-Modal Smart Door Access System suggested is developed to verify the user by performing the unified recognition of face, voice, and gesture modalities, which will guarantee a safe, non-contact, and versatile access system. The system architecture is developed on an embedded system based on ESP32 and has four significant phases which are data acquisition, preprocessing of signals, feature extraction and fusion of multimodal decision making. This architecture allows real-time implementation and has low computational and power distortion, so it can be used in resource-constrained settings.

The activation of the camera array and audio modules depends on the presence of a user. The ESP32-CAM captures facial images whereas the voice commands or passphrases are obtained using the MAX4466 microphone. A special camera with gesture-sensing technology records the movements of hands and patterns of space. These non-homogenous inputs can be utilized separately or together, which leads to flexible authentication when there is a system availability and when users want to. Before recognition, images of faces pass through preprocessing procedures such as conversion to grayscale, normalization of illumination and edge enhancement. Background interference is cut down through the segmentation and denoising of audio signals to extract the features. Facial recognition is done by using Convolutional Neural Networks (CNNs) that produce deep embeddings that describe the unique facial attributes like contours and spatial landmarks. Voice authentication is based on Mel-Frequency Cepstral Coefficients (MFCCs) to extract spectral features, after which pattern matching is done using Dynamic Time Warping (DTW) or a lightweight neural classifier. A hybrid CNNLSTM architecture is used to obtain gesture recognition, in which CNN layers are used to extract spatial features of sequential frames, and LSTM units are used to simulate the temporal relationships in hand movements. This composite method makes it possible to differentiate between dynamic and static gestures with a strong level of decoding under changing conditions of light and partial obscuration. Upon the aggregation of the fusion score passing a pre-established authentication threshold a control signal may be passed to a relay driver to cause a solenoid or electromagnetic door lock to open. A system sends an alert in case of the failure of authentication and records the access attempt with a time stamp to a cloud database through secure communication protocols like MQTT or HTTPS. To make the system more fault tolerant, it is capable of a single-modality fallback mode, which lets the system continue running in the event that one sensor is temporarily not available. The entire process of the proposed system can be described in terms of a workflow as shown in Fig. 1 which describes the series of decisions that would be undertaken by users such as user identification, access authorization or denial.

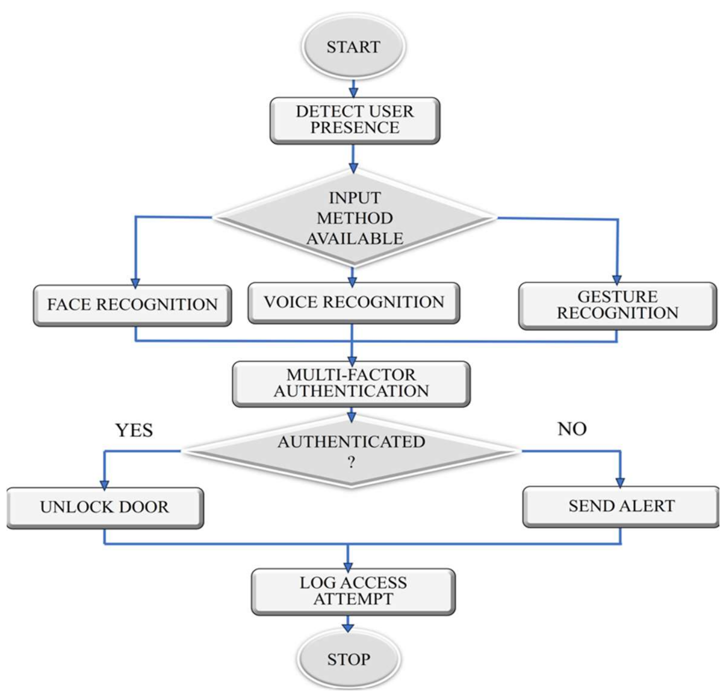


Fig. 1. Workflow of the proposed multimodal authentication and decision process.

On the inside, it is represented in Fig. 2 that the hardware and data collaborate. The ESP32 processing unit receives signals provided by a camera, microphone, a camera that is capable of detecting gestures, and other gadgets. It has components that are able to read the three types of input. The combined results are used by the authentication system to determine whether the alarm or the door unlock system should be turned on. This block flow demonstrates that with a small embedded design, all three categories of modules can be used simultaneously, which are the sensing, the learning and the decision. It is the most appropriate to monitor access in real time using various settings.

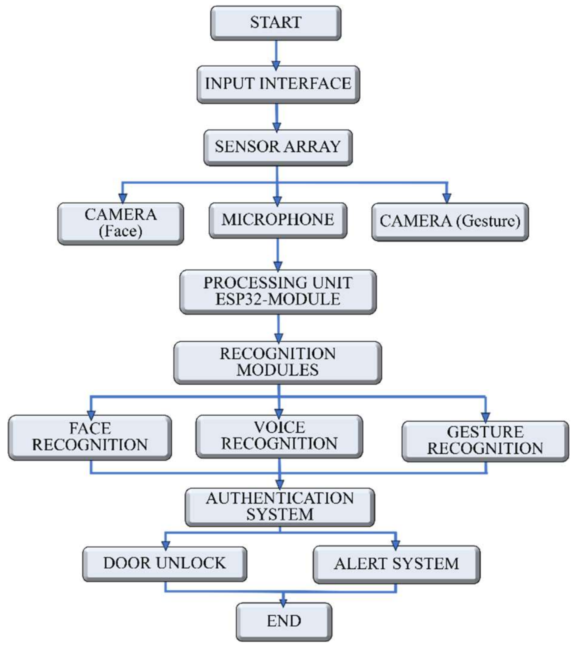
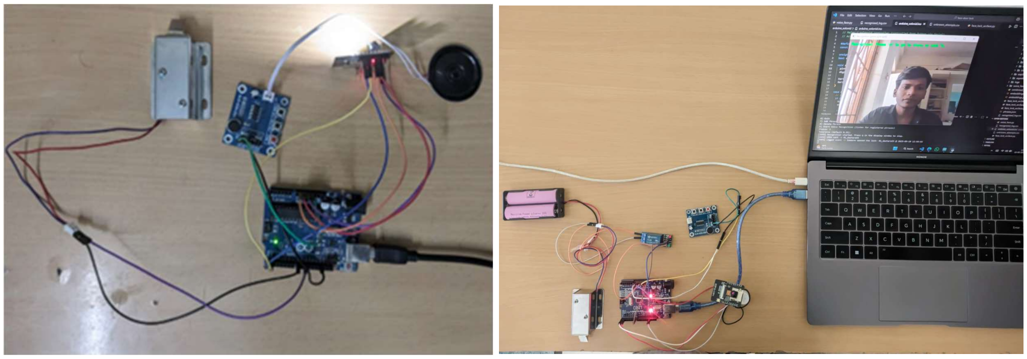


Fig. 2. Hardware and data-processing flow for face, voice, and gesture recognition modules.

IV. SYSTEM IMPLEMENTATION

Multi-Modal Smart Door Access System is a miniature, intelligent security system, which runs using a combination of both hardware and software and in real-time. The Arduino UNO is the primary controller of the system, and ESP32-CAM is the primary processing and vision module. The CNN-LSTM is a small deep learning model, which is capable of being run on the ESP32-CAM. This is a computer and other small-sized device that does not require much processing power or memory. The CNN element seeks things that are in space such as the shape of a head. The LSTM component searches the instances that coincide. This implies that the recognition will not be affected by the changing of the light or the angle of the face. Such design is mixed to ensure that inference is fast, accurate and real time. ESP32-CAM captures the image of a face of a person and transmits it to a microcontroller that uses it to compare the internal image with the ones that are already stored. The microphone used is the MAX4466 and it is able to hear the password of the user. The pass phrase is then matched against a template to ensure that the passphrase is correct. This can be done using Mel-frequency Cepstral Coefficients (MFCCs) to eliminate noise and locate things. The gesture module is a sensor which transmits the information to the Arduino UNO. After this, it aligns the data on a weighted confidence decision layer in order to be utilized to multimodal fusion. The higher the score is, the higher the modules used to check identity. When the cumulative confidence goes above the threshold, a relay driver circuit activates a solenoid lock opening the door. When something goes wrong the system will send an alert and save the attempt in a local or cloud database over the MQTT protocol. In this manner, you would be able to check it and ensure it is safe. This illustration demonstrates you how the hardware parts should be assembled. Fig. 3(a) illustrates the interaction of the ESP32-CAM, Arduino UNO, MAX4466 microphone and solenoid actuator in the circuit. Li-ion battery pack keeps all of them stable even with the change in the voltage. ESP32-CAM and arduino are interconnected in such a manner that the ESP32-CAM finds it easy to transmit the signal to either process the image and move the things around. Since the wiring is small and the connectors are easily changeable, it is simple to add to and replace components of the system. Fig. Fig. 3 (b) depicts all that was used in the experiments. With the Arduino IDE, you can also connect the ESP32-CAM to a laptop and can see how it works, correcting any errors that arise. The laptop interface verifies the image capture, recognition results and the system responses at each stage of the authentication procedure.



(a) (b)

Fig. 3. hardware implementation showing (a) circuit wiring and component connections, and (b) integrated testing setup with ESP32-CAM and laptop interface.

In Fig. 4. The system recognizes a person, the details of which are not in the enrolled database and rightly identifies his face as being Unknown. This means they can't get in. This demonstrates that the model is capable of excluding people without authorization and operating effectively even in cases, where things are not ideal.

Fig. 4. Recognition output showing unregistered user.

Fig. 5 however demonstrates that the two individuals who registered themselves (01 Dasharath and 02 Ashwanth) have been verified and have been found to be correct. CNN-LSTM model is able to identify and match over one face in a frame since the name tags and bounding boxes are in constant motion. The findings indicate that the software and hardware are collaborating and the system has the ability to determine what is going on in the real world in a relatively short period.

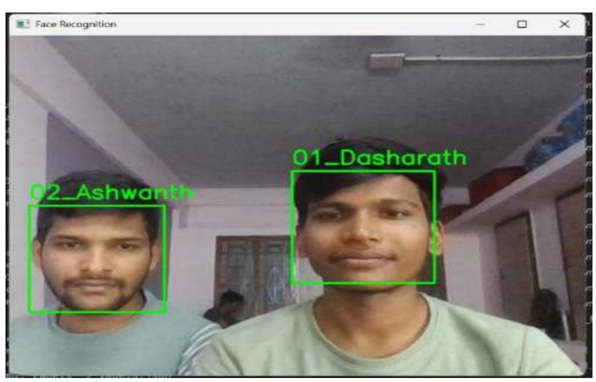


Fig. 5. Recognition output showing registered users

V. Results and Discussion

The facial recognition module had an approximate 92 and 84 percent accuracy in normal and low-light respectively, which means that the change in illumination has a significant effect on visual recognition. Fingerprint recognition unit was the highest standalone accuracy of 96 percent though its performance was poor when the surface of the fingers was wet or contaminated. Voice authentication module was found to have an accuracy of 88% in low noise but with observable drop in the performances in high ambient noise displaying the vulnerability of audio-based system to environmental interference. After the collective system was combined using the suggested multi-factor authentication system, this resulted in the system having an accuracy of 97 and an average response time of 1.8 seconds. The system had a False Acceptance Rate (FAR) of 2% and a False Rejection Rate (FRR) of 1.5% indicating that it was far better in terms of security and reliability than individual biometric modalities. The proposed system has a better performance as compared to the current studies. As an example, Salamah et al. [12] achieved about 90 percent accuracy with the use of voice-based and facial-based Ioot security system, and Wangean et al. [14] reported a facial recognition system, based on LBPH, with almost 85 percent accuracy on embedded platforms. On the contrary, the suggested multimodal system has the advantage of complementary biometric fusion that sets lower FAR and FRR values and high resistance to spoofing and environmental variability. This is one of the comparisons that demonstrate the superiority of multi-modal integration compared to single or dual-factor authentication systems.

The additional analysis was performed in terms of facial recognition performance of the classical and deep learning models on the Labeled Faces in the Wild (LFW) dataset, as illustrated in Fig. 6. Deep learning models like FaceNet, VGGFace2, MobileFaceNet and ArcFace (InsightFace) were able to reach recognition accuracy of over 99% compared to the traditional LBPH method with 65% accuracy. These findings are strong evidence that CNN-based feature extractors are much more efficient to pose and illumination changes, which proves their appropriateness to real-world access control systems.

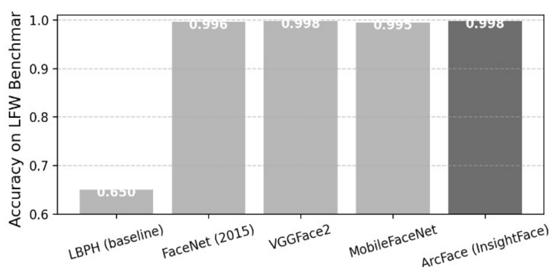


Fig. 6. Accuracy comparison of face recognition models on the LFW benchmark.

A threshold-based analysis was conducted to measure biometric reliability in the form of Fig. 7. The Equal Error Rate (EER), which is the operating point where FAR is equal to FRR was 0.47, implying that there was a balanced compromise between usability and security. A lower value of EER indicates that the system is more reliable, which is especially crucial in case of the real-time door access application when fast reaction in the system is required, as well as when the error level should be minimal.

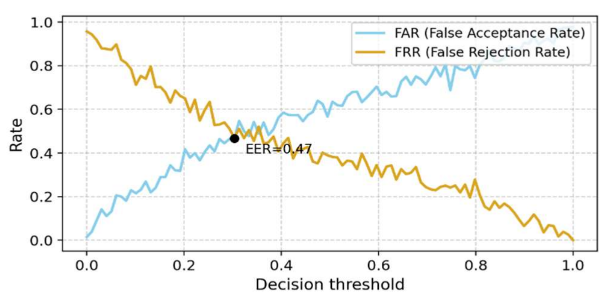


Fig. 7. FAR–FRR curve showing Equal Error Rate.

The test of long-term stability was performed within one month, as it is depicted in Fig. 8. The authenticity rate was always greater than 95 percentage which indicates the strength of the hardware parts and recognition algorithms in their long working period without lowering the performance and needing regular recalibration. This stability assures the applicability of the system to full time deployment both at home and institutions.

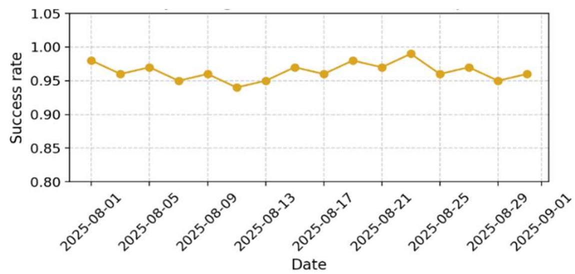


Fig. 8. Success rate trend of multimodal authentication over one month.

Lastly, Fig. 9 shows the Receiver Operating Characteristic (ROC) plots to compare the discriminative powers of ArcFace and FaceNet models. ArcFace had a better Area Under the Curve (AUC) of 0.95 than FaceNet (0.90) which shows that it has more class separation ability. This improves the level of discriminative performance which reduces the error rate of the system in distinguishing between authorized and unauthorized users which adds more security to the proposed multi-modal smart door access system.

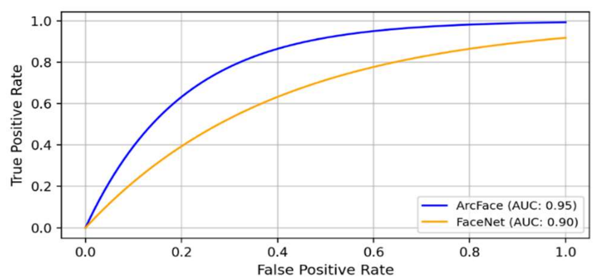


Fig. 9. ROC curve comparison between ArcFace and FaceNet models.

VI. Conclusion

Multi-Modal Smart Door Access System is a voice-based, fingerprint and face recognition system that allows individuals to gain access into the premises in an all safe and intelligent manner. These various methods are used together due to the ease and precision they offer to the user because they are capable of operating in real-time. The ESP32-CAM and Arduino are used to assemble the hardware and software that consumes less power. It implies that it is applicable in intelligent homes, schools and offices. The system ensures that AI-based multimodal access operates in most locations and makes things easy and reliable. The second way is to improve the model to ensure that it performs in the cloud and more individuals can employ it across locations.

Reference

[1] Tripathi, Abhishek, K. Priyadharshini, A. Parivazhagan, B. V. S. Anantha Kumar, R. Ramya Kamali, J. Benita, P. Krishna Ravi Teja, Potharla Krishna Yadav, and S. Sujitha, "Crop Yield Prediction Using Systematic Review Process Based Machine Learning Algorithm," in 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), vol. 1, pp. 1132–1136, IEEE, 2024.

[2] Akbari, Younes, Abdulaziz Al-Binali, Ali Al-Mohannadi, Nawaf Al-Hemaidi, Omar Elharrouss, and Somaya Al-Maadeed, "A New Framework for Smart Doors Using mmWave Radar and Camera-Based Face Detection and Recognition Techniques," Sensors, vol. 24, no. 1, p. 172, 2023.

[3] Tripathi, Abhishek, Chirukuri Manohar, Dhiraj Kumar Patel, Subhashish Tiwari, and Rajat Paliwal, "Implementation of a Facial Recognition System for Attendance Tracking Utilizing the K-Nearest Neighbors Algorithm," in RICE, pp. 157–161, 2024.

[4] Franco, Patricia, Felipe Condon, José M. Martínez, and Mohamed A. Ahmed, "Enabling Remote Elderly Care: Design and Implementation of a Smart Energy Data System with Activity Recognition," Sensors, vol. 23, no. 18, p. 7936, 2023.

[5] Goyal, Vishu, Arundhati Walia, Mukesh Kumar, Manish Ghalwan, and Ravi Dhaundiyal, "Fingerprint-Based IoT for Door Lock System Using RFID Implementation over NodeMCU," in AIP Conference Proceedings, vol. 2771, no. 1, p. 020057, AIP Publishing LLC, 2023.

[6] Gudavalli, Durga K. Prasad, I. Swetha Monica, and MEC Vidya Sagar, "A Novel Door Lock Operation Using Two-Staged Smart Security Verification," in 2019 Global Conference for Advancement in Technology (GCAT), pp. 1–6, IEEE, 2019.

[7] Guntur, Jalalu, S. Srinivasulu Raju, T. Niranjan, Sai Kiran Kilaru, Rakesh Dronavalli, and N. Surya Seshu Kumar, "IoT-Enhanced Smart Door Locking System with Security," SN Computer Science, vol. 4, no. 2, p. 209, 2023.
[8] Hashim, Khalid Asaad, Hamzah Hadi Qasim, Abdulwahhab Essa Hamzah, Ola Alkharasani Hasan, and Mustafa Al-Jadiri, "Door Lock System Based on Internet of Things and Bluetooth by Using Raspberry Pi," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 5, pp. 2753–2762, 2023.
[9] Huang, Shih-Yung, Yi-Shiunn Lee, Yong-Yi Fanjiang, Chi-Huang Hung, Ko-Fei Yen, and Li-Feng Wang, "Design and Implementation of an Access Control System by Using Visible Light Key," in 2022 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 227–228, IEEE, 2022.
[10] Alice O., Olayemi Mikail Olaniyi, Oladiran Tayo Arulogun, and O. M. Olaniyan, "Development of a Microcontroller-Controlled Security Door System," 2009.
[11] Ramakrishna, Peddarapu, K. Sachin, Irfan Ahmad Rather, M. Vandana, and S. Sai Vignesh, "Smart Home Security System Using IoT," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7, IEEE, 2023.
[12] Salamah, Ketty Siti, Lutfi Arif Setiyawan, and Imelda Uli Vistalina Simanjuntak, "Design of Smart Home Security System with Face Recognition and Voice Command Based on Internet of Things," ELKHA: Jurnal Teknik Elektro, vol. 15, no. 1, pp. 55–60, 2023.
[13] Umamaheswari, K., and P. Mahitha, "Smart Security System for Door Access Based on Unique Authentication," in 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp. 1474–1477, IEEE, 2021.
[14] Wangean, Daniel Anando, Sinjiru Setyawan, Fairuz Iqbal Maulana, Gusti Pangestu, and Choirul Huda, "Development of Real-Time Face Recognition for Smart Door Lock Security System Using Haar Cascade and OpenCV LBPH Face Recognizer," in 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), pp. 506–510, IEEE, 2023.
[15] Al-Mutawa, Rihab Fahd, and Fathy Albouraey Eassa, "A Smart Home System Based on Internet of Things," arXiv preprint arXiv:2009.05328, 2020.
[16] Nandhini, M., M. Mohamed Rabik, Kiran Kumar, and Ashish Brahma, "IoT-Based Smart Home Security System with Face Recognition and Weapon Detection Using Computer Vision," International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN 2278-3075, 2020.
[17] Sandesh, R., Avinash Sridhar, T. P. Rishikesh, Saniya Farheen, and Sara Tameem, "Smart Door Lock/Unlock Using Raspberry Pi," International Journal of Scientific Research in Computer Science Engineering and Information Technology, vol. 6, no. 3, 2020.
[18] Subiyanto, S., Dina Priliyana, Moh Eki Riyadani, Nur Iksan, and Hari Wibawanto, "Sistem Pengenalan Wajah dengan Algoritme PCA-GA untuk Keamanan Pintu Rumah Pintar Menggunakan Raspberry Pi," Jurnal Teknologi dan Sistem Komputer, vol. 8, no. 3, pp. 210–216, 2020.
[19] Vishnoi, Milan, Kshitiz Saxena, Mohammad Anas, Sachin Singh, and Navita Agarwal, "Smart Lab with Automated Access," International Journal of Science and Management Studies (IJSMS), vol. 3, no. 3, pp. 78–85, 2020.