electronics

MDPI

Article

A Deep-Learning-Driven Aerial Dialing PIN Code Input Authentication System via Personal Hand Features †

Jun Wang [1], Haojie Wang [2], Kiminori Sato [3] and Bo Wu [3,*]

[1] Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology, Hachioji 192-0982, Tokyo, Japan; g2123002c8@edu.teu.ac.jp
[2] School of Information Engineering, Chang'an University, Xi'an 710021, China; 2022124138@chd.edu.cn
[3] School of Computer Science, Tokyo University of Technology, Hachioji 192-0982, Tokyo, Japan; satohkmn@stf.teu.ac.jp
* Correspondence: wubo@stf.teu.ac.jp

† This paper is an extension version of the conference paper: Bo Wu, Hiroki Sato, and Kiminori Sato, An Aerial Virtual Dialing PIN Code Input Authentication System Design via Infrared-based Hand Tracking Device, In Proceedings of Computer Information Systems, Biometrics and Kansei Engineering 2023, Tokyo, Japan, 22–24 September 2023.

Abstract: The dialing-type authentication as a common PIN code input system has gained popularity due to the simple and intuitive design. However, this type of system has the security risk of "shoulder surfing attack", so that attackers can physically view the device screen and keypad to obtain personal information. Therefore, based on the use of "Leap Motion" device and "Media Pipe" solutions, in this paper, we try to propose a new two-factor dialing-type input authentication system powered by aerial hand motions and features without contact. To be specific, based on the design of the aerial dialing system part, as the first authentication part, we constructed a total of two types of hand motion input subsystems using Leap Motion and Media Pipe, separately. The results of FRR (False Rejection Rate) and FAR (False Acceptance Rate) experiments of the two subsystems show that Media Pipe is more comprehensive and superior in terms of applicability, accuracy, and speed. Moreover, as the second authentication part, the user's hand features (e.g., proportional characteristics associated with fingers and palm) were used for specialized CNN-LSTM model training to ultimately obtain a satisfactory accuracy.

Keywords: hand tracking; Media Pipe; deep learning; CNN-LSTM; non-contact authentication; dialing-type authentication; two-factor authentication

Academic Editor: George A. Tsihrintzis

Received: 24 November 2024
Revised: 24 December 2024
Accepted: 29 December 2024
Published: 30 December 2024

Citation: Wang, J.; Wang, H.; Sato, K.; Wu, B. A Deep-Learning-Driven Aerial Dialing PIN Code Input Authentication System via Personal Hand Features. *Electronics* **2025**, *14*, 119. https://doi.org/10.3390/electronics14010119

1. Introduction

As modern technology continues to evolve, there is a growing need for authentication methods. The worldwide outbreak of New Crown Pneumonia has accelerated the urgent need for contactless authentication. This trend is not only influenced by epidemic prevention and control but also reflects society's pursuit of a more efficient and hygienic lifestyle [1]. In contrast to the traditional biometric authentication methods, such as fingerprint recognition [2] and vein recognition [3], which are easily affected by the external environment and require the user to come into direct contact with the device, a simple and secure contactless authentication method consistent with our daily living habits is needed to meet the current social demand for hygienic safety and efficient access.

The dialing-type authentication as a common PIN code input system has gained popularity due to its simple and intuitive design. However, this type of system requires user contact with the system interface and is susceptible to the threat of shoulder hacking

(observing another person inputting a PIN code over the shoulder) because it can be used by anyone who knows the code [4]. A shoulder surfing attack describes a situation where the attacker can physically view the device screen and keypad to obtain personal information. Therefore, new contactless authentication technology to solve the safety and convenience problems is needed.

On the other hand, with the rapid development of Internet of Things (IoT) technology, research and techniques related to human motion capture and analysis [5–9] have become very common. In addition, hand feature recognition has been extensively studied using features such as hand geometry and finger proportionology, which lays a solid foundation for further exploration in this field [10]. In particular, hand-tracking devices continue to mature and can now track fingers and output their coordinates with high precision and low costs. For example, Leap Motion is an advanced hand-tracking device that uses infrared technology to recognize hand and finger movements in three-dimensional space with an astonishing 1/100th of a millimeter accuracy. Also, Leap Motion has been used in many research areas such as biometric identity verification [11] and sign language [12], where it has achieved remarkable results.

In addition, Media Pipe as a deep-learning-driven solution has a hand-tracking module, which can also capture finger and hand movements with high precision and output the corresponding coordinates [13]. As a camera-based solution, Media Pipe can easily run on most devices such as smartphones and tablets. Moreover, the use of Media Pipe can allow users to easily leverage their existing devices for gesture recognition and interaction without additional investment in specialized hardware devices [14].

Therefore, based on the use of "Leap Motion" devices and "Media Pipe" solutions, in this paper, we try to propose a new two-factor dialing-type input authentication system powered by aerial hand motions and features without contact. To be specific, the design of the aerial dialing system will be given first. Then, as the first authentication part, we will construct two types of hand motion input subsystems using both Leap Motion and Media Pipe separately and compare them by the experiment-based measures of FRR (False Rejection Rate) and FAR (False Acceptance Rate). Moreover, as the second authentication part, the user's hand features (e.g., proportional characteristics associated with fingers and palm) will be used for the training of a CNN-LSTM model to increase the accuracy of authentication. These two parts will combine as the two-factor authentication system we try to propose in this research, which is called the deep-learning-driven aerial dialing PIN code input authentication system (DADAS).

The contributions of this paper are as follows:

A non-contact authentication system based on air dialing has been designed.

Two hand-tracking technologies, Leap Motion and Media Pipe, have been compared to identify the more suitable approach for our system.

Utilizing the user's hand features, CNN-LSTM technology is employed to achieve user identity authentication, thereby enhancing the security of the system.

By integrating hand tracking with deep learning technology, real-time two-factor authentication is successfully implemented.

The rest of this paper is organized as follows: Section 2 will outline the relevant work on contactless authentication and related research on CNN-LSTM deep learning models. Then, in Section 3, we will look at the deep-learning-driven aerial dialing PIN code input authentication system (DADAS). Also, in Section 4 is the hand-motion-based authentication part of the two-factor authentication, which describes mechanisms designed to explain how to use a wireless dial-up system. Section 5 is the hand-feature-based authentication part of the two-factor authentication. It includes hand feature extraction and model training.

Section 6 is the DADAS test results. Finally, Section 7 is the conclusion, contributions, and future work.

2. Related Works

In this section, we will introduce relevant research about contactless authentication systems and the issues of authentication systems that use CNN-LSTM deep learning models.

2.1. Authentication System: Leap Motion-Based Authentication System

Leap Motion based on infrared cameras is a common hand-tracking device that is widely used in authentication systems. Based on the CNN method, Yamamoto et al. proposed a personal authentication method using the Leap Motion sensor for airborne digital writing and verified the feasibility and effectiveness of the method through experiments, obtaining a high authentication accuracy rate. Moreover, using a convolutional neural network as a machine learning method makes full use of the advantages of deep learning technology in digital recognition and authentication [15].

On the other hand, Ata et al. proposed biometric technology based on hand tremors, taking identity verification as one of the main concerns of the paper, emphasizing the importance of enhancing security in the age of information security. Moreover, a variety of feature extraction methods are used for comparative analysis to select the best method, which increases the reliability and practicability of the method [16].

In addition, Sato et al. Proposed a method for personal authentication based on the aerial click operation of the virtual touch panel, which combined the PIN code and selected finger information for dual authentication, increasing the security of authentication. By using the aerial click operation and two-factor authentication method, the security of personal authentication is effectively improved, and the possibility of attack methods such as peeping and snooping is reduced [17].

Two-factor authentication mechanisms are widely used in literature to combine different authentication elements, such as hand movements and hand information, to improve the security of the authentication system. These studies provide strong technical support for the establishment of new personal authentication systems.

2.2. Authentication System: Media Pipe-Based Authentication System

The following studies explore the reliability and validity analysis of hand-tracking techniques based on the Media Pipe framework in different application domains.

Latreche et al. describe a Media Pipe-based measurement system for the reliability and effectiveness analysis of human rehabilitation exercise [18]. Working with physical therapists, Latreche et al. collected data from about 50 healthy volunteers to determine the reliability of the Media Pipe-based shoulder measurement system. This proves that it is also feasible to use Media Pipe to collect hand ratio characteristics for identity authentication.

Amprimo et al. mainly discussed the reliability and effectiveness of hand-tracking technology based on Google Media Pipe Hand (GMH) and GMH-D frameworks in clinical applications [19]. Through the experimental results, the authors found that the two frameworks show a high consistency in both time and spectral characteristics. If we use Media Pipe to collect hand features, then we can learn from the research methods and evaluation criteria of this paper and evaluate the performance and reliability of the system by comparing it with the gold-standard system.

Harris et al. developed a simple user guide application that uses the Media Pipe framework [20]. The various gesture data were then trained, each gesture was recognized, and the message was conveyed through the gesture in the system user guide application. Users can archive information in the user guide based on recognized gestures. Therefore,

the identity authentication system based on Media Pipe collecting hand proportion characteristics can be used in combination with this study to improve the security and accuracy of identity authentication.

The authors both explore the potential uses of hand-tracking technology based on the Media Pipe framework for reliability and effectiveness analysis, clinical applications, and user guidance. They introduce the possibility of identity authentication through the collection of hand features and suggest ways to evaluate and compare system performance.

2.3. CNN-LSTM Model-Based Authentication System

The following studies introduce the identity authentication technology based on deep learning, which provides a reference for this paper.

For example, Bajaber et al. studied touch gesture biometrics systems that authenticate users by analyzing their touch behavior on a touch device [21]. On the dataset, the best results achieved by using the CNN-LSTM model are 84.87% accuracy in training, 78.28% accuracy in validation, and 78.35% accuracy in testing. The ability of deep learning methods to automatically discover the features of touch gestures is explored, and the effectiveness of this method in touch authentication is proved.

Liu et al. provide a viable multi-factor user authentication solution [22]. This system combines multiple biometrics and pattern-based passwords to achieve multi-factor authentication with a single gesture operation, increasing security and improving the user experience. In addition, an anti-deception scheme is proposed to enhance the system's ability to resist attacks. Therefore, we refer to the methods and techniques in this paper to achieve a more reliable and secure multi-factor user authentication system.

Garcia et al. compared several machine learning algorithms, including long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and long short-term memory convolutional neural networks (CNN-LSTMs) [23], which adjust hyperparameters. The results show that the correct classification rate of the CNN-LSTM is 84.76%, which is slightly higher than the other networks.

Raghavendra et al. proposed an authentication mechanism that combines facial recognition and facial movement to enhance security by analyzing the facial movement of the user while saying the password. The model overcomes the problem that traditional facial recognition systems may be decoded by photos, masks or glasses, and is not affected by language barriers [24].

Gao et al. proposed an automatic ear needle segmentation method based on deep learning, which mainly includes three stages: ear contour detection, anatomical part segmentation and key point localization, and image post-processing. The mAP of anatomical part segmentation and key point location of this method are 83.2% and 98.1%, respectively, and the running speed is significantly improved [25].

These documents provide identity authentication technologies based on deep learning methods, including analyzing touch behavior for identity authentication, multi-factor identity authentication schemes, and the application of machine learning algorithms in identity authentication. It provides useful references and support for our research.

2.4. Advances in Contactless Biometrics for Authentication

Contactless hand biometrics has gained significant attention due to its non-invasive nature and applicability in diverse scenarios. Gonzalez-Soler et al. advanced contactless hand biometrics, highlighting its potential as an alternative to traditional biometric methods such as fingerprint and facial recognition, especially for forensic applications. This paper studies the application of deep neural networks in hand recognition and achieves high precision under controlled conditions [26]. Li et al. addresses the limitations of traditional hand

contour feature extraction methods in handling challenges such as skin color, occlusion, and lighting in flight simulation environments. It proposes an improved method combining skin color processing, hand key point detection, and an eight-way seed-filling algorithm for image segmentation [27]. Imura et al. proposes a hand-gesture-based biometric authentication method as an alternative to traditional passwords, addressing computer security concerns. Using a 3D motion sensor (Leap Motion), the system captures and analyzes fingertip and finger joint movements to enhance authentication performance [28].

3. Deep-Learning-Driven Aerial Dialing PIN Code Input Authentication System

In this paper, we try to propose a novel contactless PIN authentication system that supports both Leap Motion and Media Pipe as input methods. This system, which we refer to as the deep-learning-driven aerial dialing PIN code input authentication system (DADAS), combines contactless PIN code input with hand-feature-based user identification.

As illustrated in Figure 1, the DADAS incorporates two core authentication modules: a PIN-based dialing module and a CNN-LSTM model-based hand feature module. In this dual-authentication framework, both modules operate concurrently to enhance security and reliability.

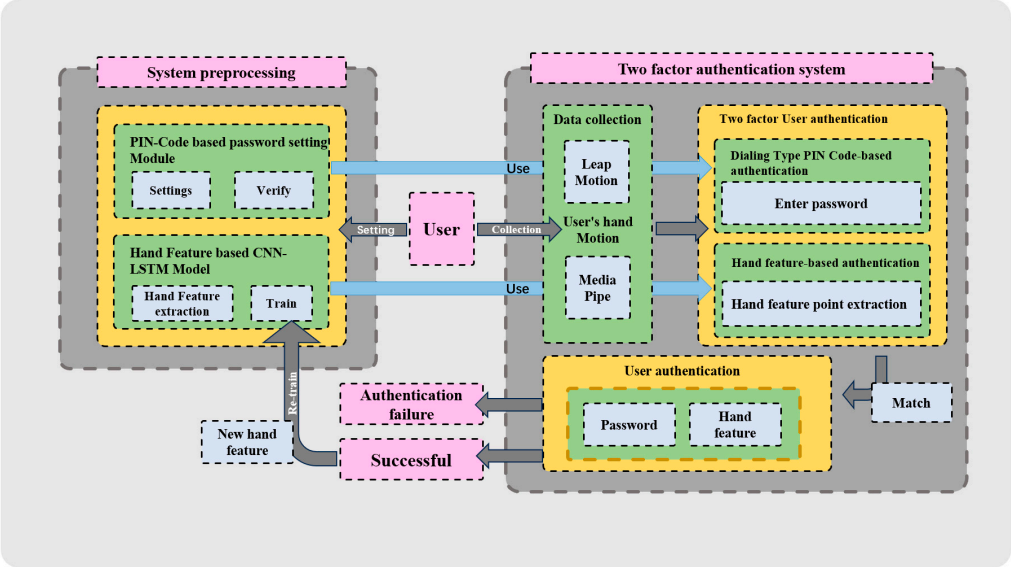


Figure 1. DADAS system design.

In the pre-processing phase, the PIN-based dialing module allows the user to set and confirm the password and authenticate the password entered by the user. At the same time, the hand features module based on the CNN-LSTM model extracts key hand features of users and uses deep learning technology to train these data to generate AN accurate authentication model.

In the two-factor authentication phase, the user's hand movements are captured through Leap Motion and Media Pipe for password entry and authentication. At the same time, the hand feature authentication system extracts the user's hand features and passes them to the trained CNN-LSTM model for authentication. For the user to successfully complete the authentication, the PIN-based dialing module and the hand-feature-based module must simultaneously verify the user's identity.

If the authentication is successful, then the system will continue to use the user's hand data for model learning and optimization, thereby further improving the accuracy

and adaptability of recognition. If authentication fails, then the user will be prompted to re-enter the hand data and will be asked to reset the password if necessary.

Through the close combination of these two core modules, DADAS implements more efficient and secure two-factor authentication, providing users with a reliable and smooth authentication experience. The details of the system will be introduced in the following chapters.

The following sections, Sections 4 and 5, will explore the two important components of two-factor authentication in detail, respectively: Section 4 focuses on the authentication method based on hand motions, while Section 5 discusses the authentication method based on hand features.

4. Two-Factor Authentication: Hand-Motion-Based Authentication

This section will describe the mechanism design to explain how to use the aerial dialing system without touch.

4.1. Design of the Dialing-Type Authentication System

In the system, the user needs to use Leap Motion or Media Pipe to manipulate a carousel displayed on the screen and enter PIN codes by aligning the digits with the entry point. Users can rotate the digits by turning their index finger in the air and entered The PIN code by aligning the digit in the screen. Because the dial can be operated without touching the screen, this type of dial is called an "aerial dial".

We have meticulously engineered a set of airborne dials on the screen by emulating traditional dials, as illustrated in Figure 2. The rotational direction and the digit count are also exhibited on the screen. To ensure the compatibility and effective integration of the two frameworks, the development environment of the Leap Motion and Media Pipe frameworks is Windows 10.

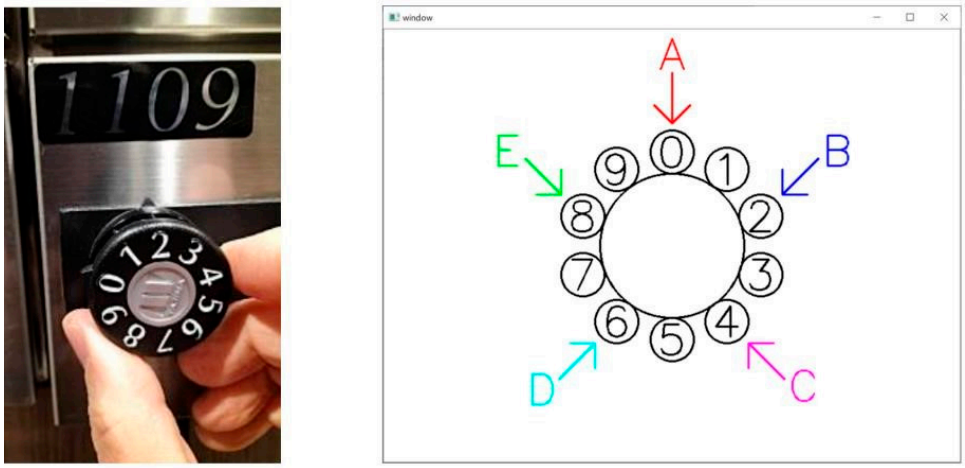


Figure 2. Aerial virtual dials simulating traditional dials.

As shown in Figure 3, an over-the-air virtual dial pad is displayed on the screen, which can be operated by the user through Leap Motion or Media Pipe utilizing free rotation of the index finger (or other fingers) in any direction. Unlike traditional dialing methods, the system gives different fingers unique functions such as insert, delete, and reset. The user uses these fingers to enter a four-digit PIN code at each of the five input position points (A through E). It is worth noting that the position of the position points (A to E) changes randomly each time the PIN is entered, increasing the security of the system.

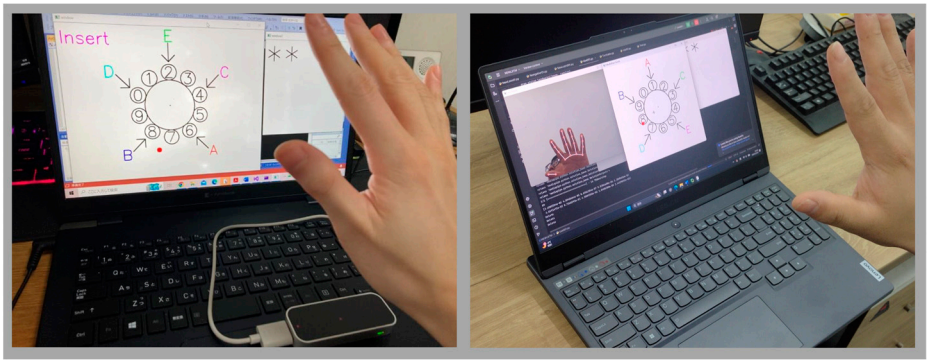
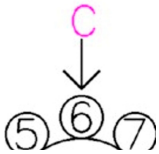
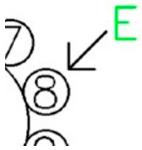


Figure 3. How users operate aerial virtual dial pad by moving their fingers in the air through Leap Motion or Media Pipe.

For example, when the user tries to use the PIN code "0862", he/she must enter it in the pre-registered order at the five-entry points A, B, C, D and E. If the order is "D → E → C → A", then the user needs to turn the dial to the position corresponding to the one shown in Table 1.

Table 1. Step-by-step instructions for operating the aerial virtual dial pad.

Digits	Process	Diagrammatic Drawing	Digits	Process	Diagrammatic Drawing
First digit	0 for D		Third digit	6 for C	
Second digit	8 for E		Fourth digit	2 to A	

The system uses a combination of numeric and positional input methods, and the user needs to register the PIN code and the corresponding input position point sequence in advance. Here, there can be duplications of numbers and position points (A, B, C, D, and E). With this design, even if a third party observes the user from behind, it is impossible to determine which digit the user has entered at which position. For example, Figure 2 shows that the digit for B is 2, but a third party might see 0 for A or 4 for C. This means that the third party has only a 1/5 chance of guessing each digit. If four consecutive guesses were required, then the probability of authentication success would be only 1/625 (4 squares of 5 = 625 possible combinations), or 0.16%, making it difficult to detect.

4.2. Mechanism of the Dialing-Type Authentication System

The section will describe the mechanism to explain how to use the aerial dialing system.

Insert Function. Users can trigger a red dot display by placing their hands vertically the Leap Motion and Media Pipe. This red dot indicates the position of the user's index fingertip on Leap Motion and Media Pipe. A black dot is also displayed in the center of the screen, indicating the center position of the virtual carousel in the air. The user can rotate the virtual carousel clockwise or counterclockwise by moving the index finger so that the

red dot rotates clockwise or counterclockwise around the center point (black dot), as shown in Figure 4.

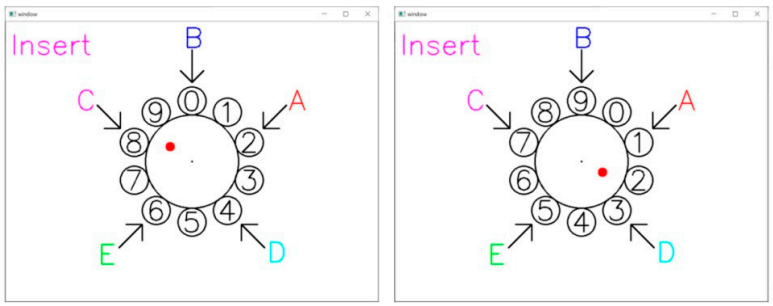


Figure 4. Insert function for aerial dial interaction. The red dot represents the user's fingertip, tracked by Leap Motion, while the black dot marks the center of the aerial dial.

In this paper, a system called "AIRCLICK" was used to enter PIN codes. An "AIRCLICK" is a finger thrusting forward in the air, similar to clicking a mouse in the air. The system determines whether an "AIRCLICK" has occurred by detecting whether the speed of the index finger (V_{index}) exceeds the speed of the palm of the hand (V_{palm}).

$$V_{index} - V_{palm} > th \tag{1}$$

As shown in Equation (1), a threshold value (th) of 130 was determined based on previous experiments, and when the speed of the index finger exceeds this threshold, the system recognizes an "air click" even if the index finger moves slightly. After matching with the user's predefined PIN, an "air tap" in any position (the position of the red dot is irrelevant) will result in the entry of a digit.

The red dot on the tip of the index finger changes to a blue dot in approximately one second when performing an air-click operation (Figure 5). The red dot changing to a blue dot indicates a successful input. Once the red dot changes to a blue dot, the user can perform operations other than input, but cannot trigger the input again. This design prevents a single click from being mistaken for two inputs. It takes about a second to go from the blue dot back to the red dot, and once the red dot reappears, the input is ready.

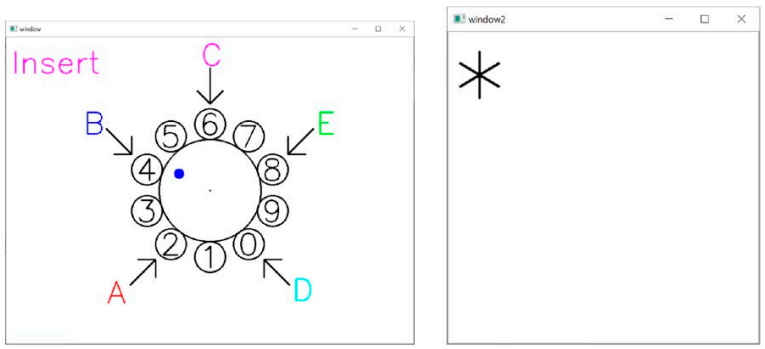


Figure 5. Aerial dialing and feedback screen immediately after input (first digit of input).

On the feedback screen, the entered PIN code is displayed as "*". After entering four digits, the result is displayed on the feedback screen. If the registered PIN matches the entered PIN, then "Accept" is displayed in blue. If there is no match, then "Reject" is displayed in red. Figure 6 shows the feedback screen when the entered four-digit PIN matches the registered PIN.

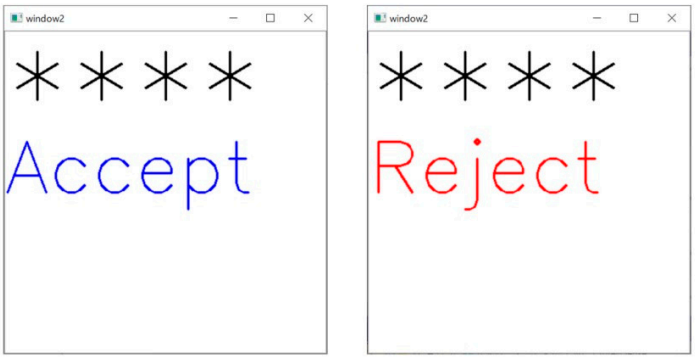


Figure 6. "Accept" and "Reject" screens (asterisk "*" represents the entered PIN values).

Delete Function. Most general PIN code entry systems have a button that allows the user to delete the PIN code if the user accidentally enters a value that is different from the PIN code that he or she wants to enter. Without this feature, if even a single digit is entered incorrectly, the user must start over again, which is very time-consuming and inconvenient. Although the system does not have a "button for deleting one digit", it does have a "function for deleting one digit". This is achieved by "air-clicking" with the ring finger. If a wrong PIN code is entered during PIN code entry, then a digit can be deleted by air-clicking with the ring finger using the same action as when entering the PIN code.

As shown in Figure 7, when deleting, the word "Delete" is displayed in green letters on the feedback screen, and the number of "*" is reduced by one. When deleting, the position of A to E on the aerial virtual dial is randomly changed. A green dot appears on the aerial virtual dial screen at the tip of the ring finger for less than one second. As with the blue dot, the user cannot delete it again until the green dot disappears.

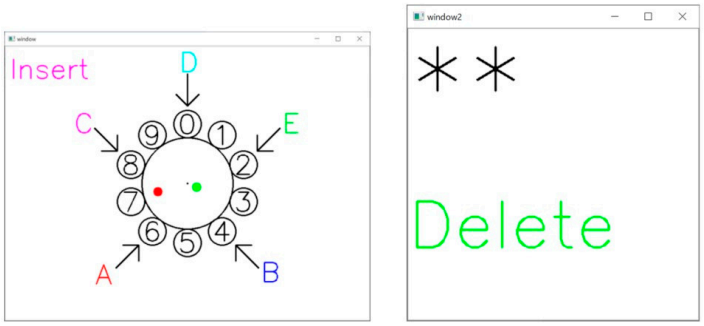


Figure 7. Delete function. A green dot appears on the screen along with the word "Delete", indicating that the password has been reduced by one character (asterisk "*" represents the entered PIN values).

Reset Function. On the other hand, the system has a function that deletes and initializes all entered PIN codes after the four-digit PIN code has been entered. Similar to delete, air-clicking with the ring finger deletes all entered PIN codes and returns to the initial state. The same finger is used for both deletion and resetting. Once four digits have been entered and the result is displayed, it is no longer possible to delete a single digit, but it can be reset.

As shown in Figure 8, when reset, the word "Reset" is displayed in light blue letters on the feedback screen, and all "*" marks on the screen are deleted. When resetting, the positions of A to E on the aerial virtual dial are randomly switched. A light blue dot appears on the aerial virtual dial screen at the tip of the ring finger for less than one second. After resetting, the PIN code can be entered again from the beginning.

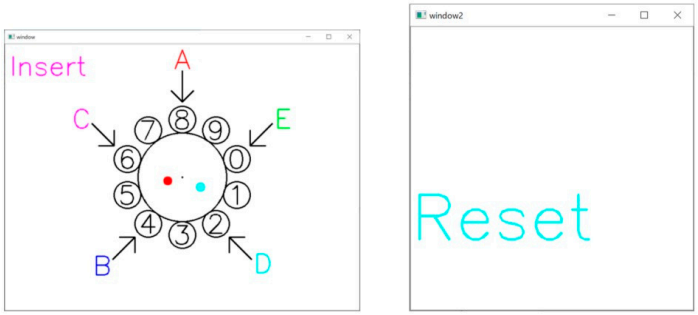


Figure 8. Aerial dial and feedback screen immediately after reset (reset).

4.3. Leap Motion and Media Pipe

In this paper, the advantages of Leap Motion and Media Pipe are used to develop the over the air dialing interface. This module allows users to manipulate a carousel displayed on the screen and enter PIN codes through intuitive hand gestures, thereby enhancing user experience and security.

Leap Motion is a high-precision device that uses infrared LEDs to recognize hand and finger movements in three-dimensional space at 1/100th of a millimeter intervals [29]. Its superior ability to recognize hand and finger movements has led to a wide range of applications in various research projects, including biometric identity verification and sign language. Figure 9 shows the appearance of Leap Motion.

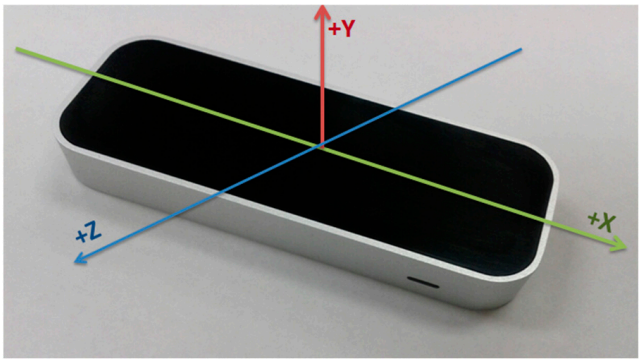


Figure 9. Leap Motion device.

Leap Motion illuminates the user's fingers with infrared LEDs and uses the feedback data to pinpoint the position of the user's hand, fingers, and joints in 3D space [30], obtaining their coordinates in the X, Y, and Z axes. Not only that, but Leap Motion also measures the velocity of the fingers in the X, Y, and Z axes, providing detailed position and motion information, as shown in Figure 10.

On the other hand, Media Pipe implements hand tracking by performing pose estimation using a deep learning model that analyzes images to determine the location of key points of the hand, including the palm, fingers, and wrist, to obtain coordinates in the X, Y, and Z axes. By calculating the change in key points between frames, information about the velocity of the hand in the X, Y, and Z axes can be obtained, and the skeletal data of the fingers can be further inferred. This enables Media Pipe to output hand-tracking results, including hand position coordinates, finger poses, and velocity information, providing a flexible, versatile, and scalable hand-tracking solution for a variety of application scenarios.

As shown in Figure 11, the key nodes of the hand are obtained by Media Pipe, and the proportion of the user's hand is obtained by calculating and comparing the vector distances between the nodes, which is used as the basis for recognition.

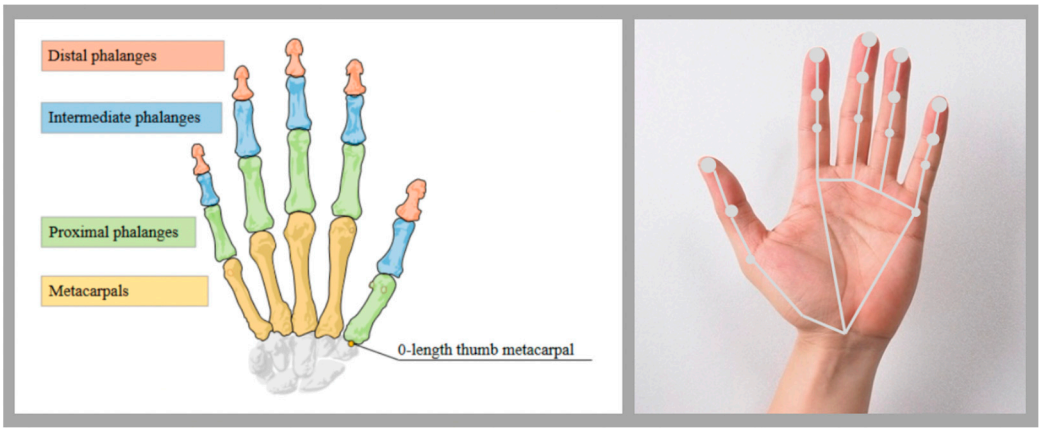
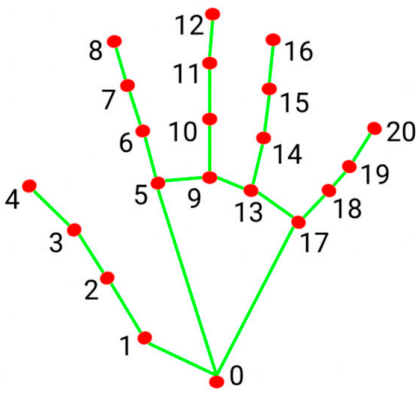


Figure 10. Hand anatomy and tracking points visualized by Leap Motion.



0.WRIST	11.MIDDLE_FINGER_DIP
1.THUMB_CMC	12.MIDDLE_FINGER_TIP
2.THUMB_MCP	13.RING_FINGER_MCP
3.THUMB_IP	14. RING_FINGER_PIP
4.THUMB_FINGER_TIP	15.RING_FINGER_DIP
5.THUMB_FINGER_PIP	16.RING_FINGER_TIP
6.INDEX_FINGER_DIP	17.PINKY_MCP
7.INDEX_FINGER_DIP	18.PINKY_PIP
8.INDEX_FINGER_TIP	19. PINKY_DIP
9.MIDDLE_FINGER_MCP	20. PINKY_TIP
10.MIDDLE_FINGER_PIP	

Figure 11. Hand joint point marking and sorting and Media Pipe hand feature point name.

Through the application of actual scenarios, this system can show its comprehensive advantages in security, convenience, and health protection, and provide innovative solutions for identity authentication in many fields.

1. In systems such as ATMs that require a PIN code to be entered, users can enter the PIN code through gestures to avoid touching the device.
2. Appropriate for systems where users need to enter data remotely for security or privacy reasons, such as Safe case, and so on.
3. With the continuous development of VR technology, some systems will also have requirements for input functions. After simple optimization, the system can be seamlessly embedded into the existing VR system to provide users with a convenient and safe input method.

4.4. Experiments, Contrast, and Discussion

Based on our previous research [31], to compare the impact of different input modes (Leap Motion and Media Pipe) on the authentication system, we conducted a set of acceptance rate experiments.

To be specific, to obtain the FAR (False Acceptance Rate), we asked our collaborators to shoulder a hacker to record a video of entering a PIN and tested whether they could detect the correct PIN. Second, we conducted authentication experiments, and to obtain the FRR (False Rejection Rate), we asked participants to enter their registered PIN and measured the time it took them to enter the password, the number of successful authentications, and the number of incorrect entries.

4.4.1. FAR Experiment Results (Leap Motion)

In this experiment, we invited 16 participants to watch a video recording of a user who entered his PIN code with the aerial virtual dialer, in which the 16 participants were asked to guess the correct PIN code from the video and to write their guesses on a statement or a piece of paper. All participants had to practice for five minutes before using the system.

Table 2 shows the results of the experiment. The correct PIN code is "D4 → C5 → B6 → A7". The number of people who could figure out this PIN code from the video-recorded hacking was 0 out of 16. Therefore, in this experiment, the FAR was 0%, and the probability of getting all four digits correct is 1/625, or 0.16%. This probability is considered to be very difficult to obtain.

Table 2. Results of the FAR experiments (Leap Motion).

No.	First Digit	Second Digit	Third Digit	Fourth Digit	Positive: O False: X
Correct	**D4**	**C5**	**B6**	**A7**	**O**
1	E8	D7	E2	A7	X
2	A6	E3	A4	D5	X
3	A4	D4	B4	E9	X
4	E8	C5	E2	D5	X
5	E8	C5	D8	A7	X
6	E8	D7	B6	A7	X
7	A6	C5	D8	A7	X
8	D4	D7	A4	D5	X
9	A6	C5	B6	D5	X
10	C2	A9	B6	A7	X
11	D4	C5	E2	A7	X
12	A6	D7	D8	C1	X
13	B0	A9	A4	C1	X
14	A6	B1	A4	E9	X
15	A6	B1	D8	B3	X
16	E8	D7	A4	C1	X

Additionally, nine people in total correctly guessed at least one digit. The probability of getting at least one digit correct is 59.04%. This means that the probability of getting at least one digit correct is more than half. Certainly, the only way to break through was to get all four digits correct, but 16 of the participants had already guessed 80% of the password.

4.4.2. FRR Experiment Results (Leap Motion)

In this experiment, we invited eight collaborators to operate the aerial virtual dialer and enter their PIN codes. All eight participants were asked to enter the exact same PIN code three times, and four items were measured: FRR, number of false entries, entry time, and identity rejection rate as shown in Table 3.

Table 3. Results of the FRR experiments (Leap Motion).

No.	Number of Times	Number of Wrong Entries	Input Times	Success: O Failure: X	Rejection Rate
Subject1	1	0	25.78	O	0%
	2	0	21.03	O	
	3	0	22.03	O	
Subject2	1	0	21.58	O	0%
	2	0	20.82	O	
	3	0	21.21	O	
Subject3	1	0	11.58	O	0%
	2	0	15.56	O	
	3	1	31.03	O	
Subject4	1	0	19.25	O	0%
	2	0	15.15	O	
	3	0	16.21	O	
Subject5	1	1	23.33	O	33.33%
	2	1	35.75	O	
	3	0	20.98	O	
Subject6	1	0	16.07	O	0%
	2	0	11.75	O	
	3	0	14.95	O	
Subject7	1	0	37.01	O	0%
	2	0	17.56	O	
	3	0	16.36	O	
Subject8	1	0	20.53	O	0%
	2	0	22.5	O	
	3	0	17.35	O	

In the experiment, the specified PIN code was "B4 → A9 → E2 → A7", and if the user entered the wrong PIN code, then the user was asked to delete a character with his ring finger and re-enter the correct PIN code. The overall verification rate was 95.85%, and the error rejection rate was 4.15%. Of the eight collaborators, only one failed the validation, and the validation failed once in three times. The experimental results show that the FRR of the system is good.

In addition, four of the eight collaborators made at least one error, with between one and two errors per certification. All of the wrong entries start with the second digit. While the random rearrangement of the positioning points (A to E) may be inconvenient for the user, some cases of misinput were also found in the experiment, where the user entered the data even though no air click was made. These unintentional misinputs usually occur after input, when the user is looking for the next input point.

4.4.3. FAR Experiment Results (Media Pipe)

Similarly, in this experiment, we invited 16 participants to watch a video recording of a user entering a PIN using an over-the-air virtual dialer. The PIN was a four-digit password. Sixteen of these participants were asked to guess the correct PIN from the video and were asked to write their guesses on a piece of paper.

The result of the experiment is shown in Table 4, and the correct PIN code is B3 → E4 → C9 → A0. Out of 16 people, 0 people cracked the code through the video. So, the FAR is derived to be 0%.

Table 4. Results of the FAR experiments (Media Pipe).

No.	First Digit	Second Digit	Third Digit	Fourth Digit	Positive: O False: X
Correct	**B3**	**E4**	**C9**	**A0**	**O**
1	B3	A2	C9	D6	X
2	B3	E4	A5	D6	X
3	A5	C6	E7	D6	X
4	B6	C4	E9	C7	X
5	A5	C6	B3	B8	X
6	D1	B0	D1	A0	X
7	B3	C6	A5	B8	X
8	E7	D8	B3	C4	X
9	A5	C6	D1	E2	X
10	A4	D8	C9	B8	X
11	A4	C8	C6	D5	X
12	A1	C6	A5	D5	X
13	E4	B6	A6	B5	X
14	A5	E4	D1	C4	X
15	B3	D8	A5	E2	X
16	E7	B0	E7	D6	X

In addition, two people guessed two numbers, with a probability of 15.30%. Five people guessed one number, and the probability of getting one number right is 33.33%. A total of seven people guessed at least one number, and the probability of answering at least one number correctly out of a total of 64 numbers for 16 people is 10%. This proves that guessing all four passwords is still more difficult.

4.4.4. FRR Experiment Results (Media Pipe)

The experimental code is B3 → E4 → C9 → A0. Similarly, if the wrong PIN code is entered, then the user is asked to delete a character with the ring finger and re-enter the correct PIN code. As shown in Table 5, all of the eight collaborators passed the verification. Only one of the eight participants had an incorrect input, and the validation failed once in three times. The results of the system show that the FRR of the system is good.

Table 5. Results of the FRR experiments (Media Pipe).

NO.	Number of Times	Number of Wrong Entries	Input Times	Success: O Failure: X	Rejection Rate
Subject1	1	0	24.58	O	0%
	2	0	20.21	O	
	3	0	19.02	O	
Subject2	1	0	22.09	O	0%
	2	0	21.54	O	
	3	0	20.03	O	
Subject3	1	1	22.21	O	0%
	2	0	24.09	O	
	3	0	19.89	O	
Subject4	1	0	24.76	O	0%
	2	0	20.21	O	
	3	0	19.36	O	
Subject5	1	0	15.23	O	0%
	2	0	12.14	O	
	3	0	13.39	O	
Subject6	1	0	23.22	O	0%
	2	0	21.29	O	
	3	0	19.04	O	
Subject7	1	0	22.11	O	0%
	2	0	21.19	O	
	3	0	18.51	O	
Subject8	1	0	20.23	O	0%
	2	0	17.85	O	
	3	0	17.21	O	

4.4.5. Contrast and Discussion

Leap Motion uses infrared technology to measure the position of hands and fingers at tiny intervals with high accuracy, especially in applications that are sensitive to tiny hand movements. In contrast, Media Pipe does not require dedicated hardware and can achieve hand tracking through ordinary cameras, effectively reducing hardware costs and improving versatility and flexibility.

In the experiment comparison, one experimenter in Leap Motion's authentication system missed two out of three validations, compared to just one in Media Pipe's system. In addition, the average input time for the Leap Motion certification system was 20.64 s, while the average input time for the Media Pipe certification system was 19.97 s. The results of the experiment showed that Media Pipe was ahead of Leap Motion in both accuracy and speed of authentication.

Therefore, from multiple dimensions such as cost, versatility, accuracy, and speed, Media Pipe is more comprehensive and superior in practical applications, especially in the case of no additional sensors. This makes Media Pipe a more attractive option, especially for scenarios where cost effectiveness and flexibility are important.

5. Two-Factor Authentication: Hand-Feature-Based Authentication

In this paper, we use two main techniques, convolutional neural network (CNN) and long short-term memory (LSTM) network, from deep learning techniques to achieve the efficient verification of user identity through the collection of user hand information (proportional characteristics associated with fingers and palm) and the training of deep learning models.

5.1. Hand Feature Extraction

Based on the invitation of 10 experimenters to have their hands scanned 30 times (each scan generates 30 frames of data) we collected the target authentication video data, where each frame contains 11 hand feature values (finger joint ratio, finger ratio, and convex hull ratio in the center of the palm). As shown in Figure 12, The 11 feature points are index finger and thumb; middle finger and thumb; ring finger and thumb; little finger and thumb length ratio (four); knuckle ratio of index finger; knuckle ratio of middle finger; knuckle ratio of ring finger; knuckle ratio of little finger (four); ratio of the length of the index finger to the wrist to the length of the thumb to the wrist; ratio of the length of the little finger to the wrist to the length of the thumb to the wrist (two); one area ratio (convex surface of the hand and convex surface of the palm).

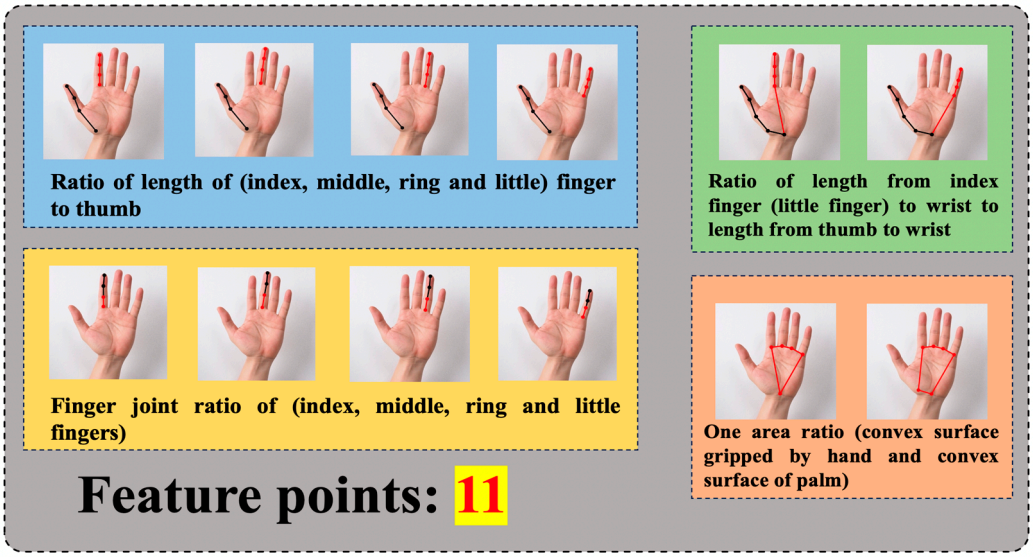


Figure 12. Hand feature points are extracted, and lines mark different finger joints and other feature points.

By feeding these data into a deep learning neural network CNN-LSTM model for learning, we ended up building a model that could verify users in real time. The model achieves efficient authentication by learning the unique characteristics of each participant's hand.

5.2. Double Authentication: Hand-Feature-Based Model Construction (CNN-LSTM)

On the one hand, we use a convolutional neural network (CNN) to process the spatial information of hand images. Through a series of convolutional and pooling layers, the CNN can effectively extract the local features of the hand and capture the important patterns in the hand image.

We introduce a long short-term memory (LSTM) network to process temporal features in the hand information (proportional characteristics associated with fingers and palm). The architecture of the LSTM layers facilitates the system's ability to capture temporal

dependencies inherent in gestures, thereby enabling our model to leverage sequential data more comprehensively.

The fully connected layer part is used to integrate the features of the CNN and LSTM, which includes two fully connected layers with ReLU activation functions. The final output layer is a dense layer with a SoftMax activation function with the number of neurons equal to the number of users. This enables the model to output classification probabilities for different users, achieving the effective identification and differentiation of user identities.

First, the model combines two neural network architectures, namely the convolutional neural network (CNN) and long short-term memory (LSTM) network. This combination is suitable for processing temporal data, such as gesture sequences.

As in Figure 13, we define the input layer, which is responsible for receiving the timing data, where the shape of the timing data is set to (None, 11, 1). Next, three convolutional layers are introduced, containing 128, 64, and 64 convolutional kernels with a window size of 3, and using the ReLU activation function. The main task of these convolutional layers is to extract key spatial features in the temporal data. A maximum pooling layer immediately follows each convolutional layer to reduce the dimensionality of the features in an orderly manner.

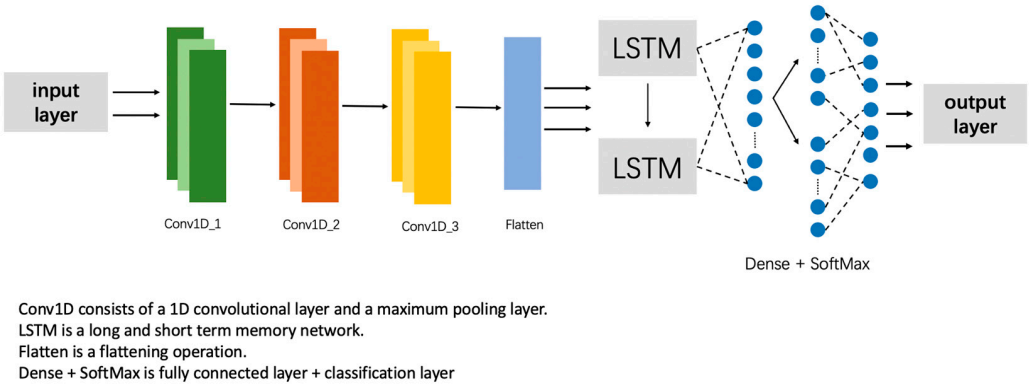


Figure 13. CNN-LSTM model for hand-gesture-based user authentication.

We then introduce a spreading layer to spread the time series data into one dimension in preparation for the subsequent LSTM layers. Next are two LSTM layers, each containing 64 hidden units. Of these two LSTM layers, the first LSTM layer is configured to return the full sequence of timing outputs, while the second LSTM layer returns the output of only the last time step.

The second half of the model consists of two fully connected layers containing 128 and 64 neurons, respectively, with ReLU activation functions. Finally, we define the output layer with several neurons equal to the number of users and a SoftMax activation function for multi-category classification. Overall, the deep learning model aims to recognize and differentiate the hand features of different users to achieve the double verification of user identity. In the design of our model, we place particular emphasis on the synergistic integration of convolutional operations and LSTM architecture. This approach aims to more effectively address the spatial and temporal dependencies inherent in time series data.

We used 10 samples of user data (each user recorded a total of 30 times, and each time 30 hand data were collected, for a total of 900 hand data per person) for training, and the learning rate was set to 0.09. During the training process, the cross-entropy loss function was minimized by using the Adam optimizer, and the model was fitted by the model.fit() function. The batch size parameter was set to 128, i.e., each batch contains 128 samples, and the entire training process was carried out in 800 rounds. During the training process, 30% of the data were used as a validation set to evaluate the performance of the model.

5.3. Experimental Results

In our previous experiments, we verified the significant advantages of the Media Pipe framework over Leap Motion in terms of cost, versatility, and speed. Therefore, in this paper, we choose Media Pipe as the framework for the dialing-type contactless PIN authentication system in the DADAS system, which completes the first level of authentication by recognizing the user's hand feature information and then enters the contactless PIN authentication system under the framework of Media Pipe to enter the password, thus realizing a higher level of two-factor authentication security.

In the next section, we describe the results of the training model and how it successfully recognizes the user's hand information (proportional characteristics associated with fingers and palm) and then enters the password through the dialing contactless PIN authentication system in the Media Pipe framework.

First, data input is performed in the model with the user identifier as 14 (the identifier is the name of the user that we arbitrarily set). We perform 30 data collections on the user's hand and save 30 frames of feature data for each collection.

Shown in Figure 14 is one instance of hand feature collection of the user under the Media Pipe-based camera. We fed the collected user feature data into the model for 800 iterations.

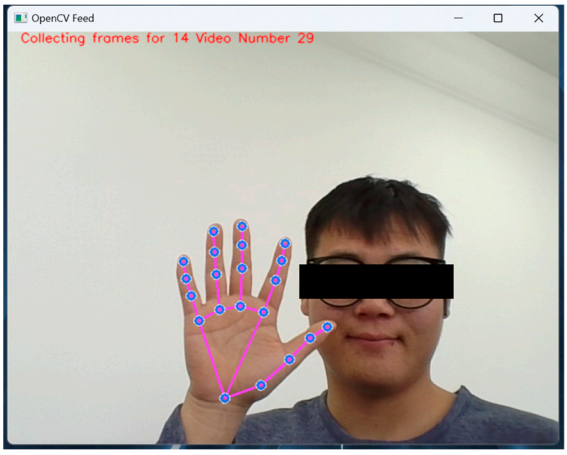
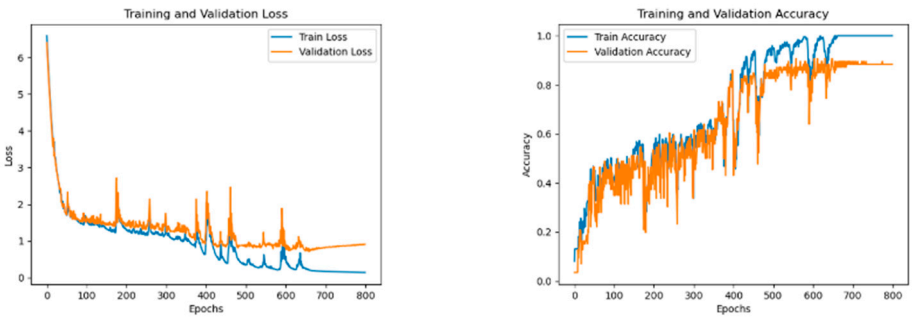


Figure 14. A collection of hand features by the user under the Media Pipe-based camera.

As can be seen from Figure 15, the accuracy rate steadily increases with the increase in iterations, while the loss rate gradually decreases and becomes stable. Specifically, the accuracy rate reached 0.8667 and the loss rate was 0.7260, indicating that our model training was successful.



(**a**) (**b**)

Figure 15. Loss rate (**a**) and accuracy (**b**) of the model.

We tested the user with identifier No. 14, which was successfully recognized as user No. 14, proving that the model's user recognition is effective. Subsequently, the user is successfully recognized and enters the dialing contactless PIN authentication phase based on the Media Pipe framework, completing the password entry and achieving double authentication.

6. System Test Results and Conclusions

Finally, we invited ten users to participate in the experiment and taught each experimenter about the DADAS in detail in advance to ensure that everyone was familiar with the DADAS process. Through the training, we ensured that each experimenter fully understood the operation process of the DADAS system, input methods, and the use of layout skills. During the teaching process, we focused on the accuracy of gesture input, how to cope with layout changes in the system, and how to effectively avoid the potential risk of bystander attacks. Through this meticulous training, we ensured the smooth running of the experiment and improved the reliability and validity of the experimental data. Based on the experimental results shown in Table 6, we see that there was a total of ten experimenters, each of whom performed ten tests. Of these tests, six participants successfully completed all 10 DADAS certifications. In addition, three experimenters missed only one out of ten DADAS certifications, while only one experimenter missed two out of ten DADAS certifications.

Table 6. Result of two-factor authentication ("○" indicates Pass, and "X" indicates Failed).

	user1	**user2**	**user3**	**user4**	**user5**
time1	○	○	○	○	○
time2	○	○	X	○	○
time3	○	○	○	○	○
time4	○	○	○	○	X
time5	○	○	○	○	○
time6	○	○	○	○	○
time7	○	○	○	X	○
time8	○	○	○	○	○
time9	○	○	○	○	○
time10	○	○	○	○	○
	user6	**user7**	**user8**	**user9**	**user10**
time1	○	○	○	○	○
time2	○	○	○	○	○
time3	○	○	X	○	○
time4	○	X	○	○	○
time5	○	○	○	○	○
time6	○	○	○	○	○
time7	○	○	○	○	○
time8	○	○	○	○	○
time9	○	○	○	○	○
time10	○	○	○	○	○

We speculate that the cause of these errors may be related to factors such as camera distance and light. While we have made great strides in accuracy, there are still some problems with identifying errors. To further improve system performance, we will work to address these potential issues in the future. In the future, we will continue to improve and expand the dataset to improve the accuracy of the system's recognition of various scenarios and situations. We will make it more concise and easier to understand in the future so that users can complete the authentication process more quickly. Deep learning models are also continuously improved to enhance the learning and generalization capabilities of the system.

7. Conclusions

In this paper, we propose a new deep-learning-driven in-flight dialing PIN code input authentication system (DADAS) based on a high-precision hand motion capture system.

Specifically, firstly, a scheme was devised to implement the in-flight dialing PIN code input system by means of hand movements. On this basis, we further compared two hand motion capture systems (Leap Motion and Media Pipe) as a part of the motion input. The FRR (False Rejection Rate) and FAR (False Acceptance Rate) experiment results show that the contactless PIN authentication system based on the Media Pipe framework is superior to the Leap Motion framework in many aspects such as cost, versatility, authentication speed, and accuracy.

In addition, in the second authentication part of synchronization, we also used the user's hand information (proportional features related to fingers and palms) and deep learning (CNN-LSTM) methods to generate user identification models to improve the security of the system. We invited 10 experimentalists, each of whom scanned their hands 30 times and generated a sequence of 30 frames of data, each containing 11 hand feature values. After training, the CNN-LSTM model verified these data and output the correct results. The accuracy rate of the model reached 0.8667 and the loss rate was 0.7260, indicating that the CNN-LSTM model can effectively learn and extract hand data and achieve accurate identity authentication.

To test the usability of the whole system, we conducted a user identification authentication experiment, in which 10 experimenters were invited to perform 10 tests each, and a total of 100 authentications were performed. The results of the experiment showed that six participants successfully completed all ten certifications, three participants missed only one certification, and one participant missed two certifications. According to these results, it can be proved that the DADAS system has achieved a high success rate in user authentication.

Aerial dialing offers an intuitive, user-friendly experience, which can enhance accessibility for users who may struggle with physical keypads, such as elderly individuals or people with mobility impairments. Although the repetition probability of hand features may be higher than that of fingerprint or iris features, through the composite authentication system, multiple features can be comprehensively judged at the same time, which effectively improves the accuracy of verification and reduces the false positive rate. The system design ensures the efficiency of the authentication process and keeps the verification time within a short range, thus balancing security and user experience. In this system, hand features are selected as the basis for recognition, and for the repeatability problem of single-person measurement, we calculate the proportion value of the hand instead of the direct value, which ensures a high accuracy. To address the lack of uniqueness of hand features, we combine the two-factor authentication of PIN codes and hand features to enhance security. Due to the non-uniqueness of hand features, it is not indexed to a specific person once lost like fingerprint data. Therefore, the proposed system can make full use of

this feature to protect the privacy of users and will not cause security problems due to the leakage of hand characteristics.

Since the dataset may not be entirely perfect or comprehensive, there remains a need to enhance the recognition accuracy of the system across various scenarios and environments. Additionally, it is possible that the user interface lacks sufficient intuitiveness, which could lead to prolonged user engagement during the authentication process and create challenges in comprehending the procedure.

As future work, we will continue to improve and expand the dataset to improve the accuracy of the system's recognition of various scenarios and situations. We also plan to further expand the application scope of contactless dynamic input systems, not only PIN input, but also more complex authentication methods. By introducing the input of letters, symbols, and custom passwords, the system will be able to support more diverse authentication requirements. At the same time, we will continue to improve the deep learning model and enhance the learning and generalization capabilities of the system.

Author Contributions: Conceptualization, J.W. and B.W.; data curation, K.S. and B.W.; formal analysis, J.W. and B.W.; funding acquisition, B.W.; investigation, J.W. and B.W.; methodology, H.W. and B.W.; project administration, J.W. and B.W.; resources, K.S. and B.W.; software, H.W.; supervision, B.W.; validation, J.W. and B.W.; visualization, J.W.; writing—original draft, J.W. and B.W.; writing—review and editing, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI [grant numbers JP21K11876].

Data Availability Statement: The data presented in this study are available on request from the corresponding author (the data are not publicly available due to privacy or ethical restrictions).

Acknowledgments: These authors contributed equally to this work: Jun Wang and Haojie Wang.

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

References

1. Baidya, J.; Saha, T.; Moyashir, R.; Palit, R. Design and implementation of a finger-print-based lock system for shared access. In Proceedings of the IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017; Volume 7, pp. 1–6.
2. Hersyah, M.H.; Yolanda, D.; Sitohang, H. Multiple Laboratory Authentication System Design Using Fingerprints Sensor and Keypad Based on Microcontroller. In Proceedings of the International Conference on Information Technology Systems and Innovation (ICITSI), Padang, Indonesia, 19–22 February 2020; Volume 8, pp. 14–19.
3. Akila, D.; Jeyalaksshmi, S.; Jayakarthik, R.; Mathivilasini, S.; Suseendran, G. Biometric Authentication with Finger Vein Images Based on Quadrature Discriminant Analysis. In Proceedings of the 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 19–21 March 2021; Volume 5, pp. 118–122.
4. Setiawan, H.A.; Rauf, R.M. Implementation of Multi-Entry Onscreen Keyboard Model on Android-Based Mobile Application to Prevent Shoulder Surfing Attack. In Proceedings of the 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE), Kasihan, Indonesia, 7–8 April 2020; Volume 3, pp. 174–178.
5. Wu, B.; Wu, Y.; Dong, R.; Sato, K.; Ikuno, S.; Nishimura, S.; Jin, Q. Behavioral Analysis of Mowing Workers Based on Hilbert–Huang Transform: An Auxiliary Movement Analysis of Manual Mowing on the Slopes of Terraced Rice Fields. *Agriculture* **2023**, *13*, 489. [CrossRef]
6. Wu, B.; Zhu, Y.; Dong, R.; Sato, K.; Ikuno, S.; Nishimura, S.; Jin, Q. Pre-braking behaviors analysis based on Hilbert–Huang transform. *CCF Trans. Pervasive Comp. Interact* **2022**, *5*, 157–182. [CrossRef]
7. Wu, B.; Wu, Y.; Nishimura, S.; Jin, Q. Analysis on the Subdivision of Skilled Mowing Movements on Slopes. *Sensors* **2022**, *22*, 1372. [CrossRef] [PubMed]
8. Wu, B.; Zhu, Y.; Nishimura, S.; Jin, Q. Analyzing the effects of driving experience on prebraking behaviors based on data collected by motion capture devices. *IEEE Access* **2020**, *8*, 197337–197351. [CrossRef]
9. Wu, B.; Wu, Y.; Aoki, Y.; Nishimura, S. Mowing Patterns Comparison: Analyzing the Mowing Behaviors of Elderly Adults on an Inclined Plane via a Motion Capture Device. *IEEE Access* **2020**, *8*, 216623–216633. [CrossRef]

10. Li, S.; Fei, L.; Zhang, B.; Ning, X.; Wu, L. Hand-based multimodal biometric fusion: A review. *Inf. Fusion* **2024**, *109*, 102418. [CrossRef]
11. Vhaduri, S.; Poellabauer, C. Multi-Modal Biometric-Based Implicit Authentication of Wearable Device Users. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3116–3125. [CrossRef]
12. Huang, X.; Wu, B.; Kameda, H. Development of a Sign Language Dialogue System for a Healing Dialogue Robot. IEEE Intl Conf on Dependable. In Proceedings of the 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), AB, Canada, 25–28 October 2021; Volume 5, pp. 867–872.
13. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.-L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
14. Liang, W.; Zhou, X.; Huang, S.; Hu, C.; Xu, X.; Jin, Q. Modeling of Cross-disciplinary Collaboration for Potential Field Discovery and Recommendation Based on Scholarly Big Data. *Future Gener. Comput. Syst.* **2018**, *87*, 591–600. [CrossRef]
15. Yamamoto, S.; Ito, S.I.; Ito, M.; Fukumi, M. Authentication of Aerial Input Numerals by Leap Motion and CNN. In Proceedings of the IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), Bali, Indonesia, 28–30 July 2018; Volume 6, pp. 189–193.
16. Ataş, M. Hand Tremor Based Biometric Recognition Using Leap Motion Device. *IEEE Access* **2017**, *5*, 23320–23326. [CrossRef]
17. Tanaka, R.; Fukumoto, S.; Kashima, M.; Sato, K.; Watanabe, M. Development of the aerial PIN code input system by the specific finger. IEICE Trans. on Fundamentals of Electronics. *Commun. Comput. Sci. A* **2017**, *J100-A*, 384–392.
18. Latreche, A.; Kelaiaia, R.; Chemori, A.; Kerboua, A. Reliability and validity analysis of Media Pipe-based measurement system for some human rehabilitation motions. *Measurement* **2023**, *214*, 112826. [CrossRef]
19. Amprimo, G.; Masi, G.; Pettiti, G.; Olmo, G.; Priano, L.; Ferraris, C. Hand tracking for clinical applications: Validation of the Google Media Pipe Hand (GMH) and the depth-enhanced GMH-D frameworks. *Biomed. Signal Process. Control.* **2024**, *96*, 123–129. [CrossRef]
20. Harris, M.; Agoes, A.S. Applying hand gesture recognition for user guide application using Media Pipe. In Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT 2021), Yogyakarta, Indonesia, 2 February 2021; Volume 1, pp. 101–108.
21. Bajaber, A.; Fadel, M.A.; Elrefaei, L.A. Evaluation of Deep Learning Models for Person Authentication Based on Touch Gesture. *Comput. Syst. Sci. Eng.* **2022**, *42*, 465–481. [CrossRef]
22. Liu, J.; Zou, X.; Han, J.; Lin, F.; Ren, K. BioDraw: Reliable multi-factor user authentication with one single finger swipe. In Proceedings of the IEEE/ACM 28th International Symposium on Quality of Service (IWQoS), Guangzhou, China, 19–21 June 2020; Volume 7, pp. 1–10.
23. Raghavendra, M.; Omprakash, P.; Mukesh, B.R.; Kamath, S. AuthNet: A deep learning based authentication mechanism using temporal facial feature movements. *arXiv* **2020**, arXiv:2012.02515. [CrossRef]
24. Gao, Z.; Jia, S.; Li, Q.; Lu, D.; Zhang, S.; Xiao, W. Deep learning approach for automatic segmentation of auricular acupoint divisions. *J. Biomed. Eng.* **2024**, *41*, 114–120.
25. Garcia, C.I.; Grasso, F.; Luchetta, A.; Piccirilli, M.C.; Paolucci, L.; Talluri, G. A comparison of power quality disturbance detection and classification methods using CNN, LSTM and CNN-LSTM. *Appl. Sci.* **2020**, *10*, 6755. [CrossRef]
26. Gonzalez-Soler, L.J.; Zyla, K.M.; Rathgeb, C.; Fischer, D. Contactless hand biometrics for forensics: Review and performance benchmark. *J. Image Video Proc.* **2024**, *26*, 1–12. [CrossRef]
27. Wang, C.Y.; Xue, P.X. Gesture Recognition Based on Key Points of Hand and Skin Color. *Comput. Syst. Appl.* **2021**, *30*, 180–185.
28. Imura, S.; Hosobe, H. A Hand Gesture-Based Method for Biometric Authentication. In *Human-Computer Interaction, Proceedings of the Theories, Methods, and Human Issues, Proceedings of the 20th International Conference, Las Vegas, NV, USA, 15–20 July 2018*; Lecture Notes in Computer Science; Kurosu, M., Ed.; Springer International Publishing: Cham, Switzerland, 2018; Volume 10901.
29. Nandy, A. *Leap Motion for Developers*, 1st ed.; Apress: New York, NY, USA, 2016.
30. Insider, B. Tracking Hands, Fingers, and Tools. Available online: https://www.buildinsider.net/small/leapmotioncs/002 (accessed on 5 January 2024).
31. Wu, B.; Sato, H.; Sato, K. An Aerial Virtual Dialing PIN Code Input Authentication System Design via Infrared-based Hand Tracking Device. In Proceedings of the Computer Information Systems, Biometrics and Kansei Engineering 2023, Tokyo, Japan, 22–24 September 2023.