Non-intrusive continuous user verification by care robots: MoveNet gait data

ZHANG, Ruomeng, KEISHING, Solan, MARCHANG, Jims <http://orcid.org/0000-0002-3700-6671>, MAWANDA, Raymond <http://orcid.org/0009-0002-2696-8135>, WANG, Ning <http://orcid.org/0000-0002-3264-1852> and DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>

Intelligent Sports and Health 1 (2025) 160–178

Contents lists available at ScienceDirect

Intelligent Sports and Health

journal homepage: www.keaipublishing.com/en/journals/intelligent-sports-and-health

Non-intrusive continuous user verification by care robots: MoveNet gait data

Ruomeng Zhang[a,*], Solan Keishing[b], Jims Marchang[b,c,**], Raymond Mawanda[b], Ning Wang[b,c], Alessandro Di Nuovo[b,c]

[a] School of Information, University of Warwick, Warwick, UK
[b] School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield, UK
[c] Advanced Wellbeing Research Centre, Sheffield Hallam University, Sheffield, UK

ARTICLE INFO

Keywords:
Care-robots
MoveNet
Machine learning
Authentication
Verification

ABSTRACT

The growing ageing population demands for advanced care technologies, such as care robots, to support quality living. Ensuring the safety and privacy of these vulnerable users necessitates reliable and user-friendly authentication methods. Security features should not become a burden to the user experiences. However, it is vital to continuously verify the user by the care robot in communicating and delivering its services. To offload the burden of verification to the users, it should be the robot initiating the verification process. However, relying on biometric data like voice doesn't guarantee the source and needs continual verbal input while face relies on the line of sight for continuous verification making it challenging. Thus, this study examined the use of MoveNet by the care robot for continuous identity verification for an authentication process of a user, leveraging the 17 data points of the gait data collected from body joints and face features along with the distances among the data points to verify user identity. The research evaluated the performance of various MoveNet models and machine learning algorithms to identify the most effective approach for continuous user authentication in care robots. The methodology involved collecting and analysing gait data from a controlled group of participants, implementing and testing with several MoveNet models and machine learning techniques, with a particular emphasis on neural networks. The results highlighted that integrating MoveNet with neural network models, especially the Thunder and Lightning f16 variants, achieved accurate user identification with an accuracy of 99.86 % (NN), 99.89 % (CNN),99.93 % (Random Forest) and KNN gives F1 score of 99.74 %, while SVM performs the worst with only 13.23 % F1 score. These findings provide an opportunity for the robot to seamlessly verify the user for authentication purpose using machine learning methods. A neural network is tested with all the MoveNet models (lightning, lightning int8, lightning f16, thunder, thunder int8, and thunder f16) and the paper proofs its usability in a ROS system, in average, prediction time takes between 0.99 to 1.06 s with an accuracy ranging from 99.64 % to 99.90 %. Lightning f16 and Thunder are the best performing models in terms of prediction time and accuracy.

1. Introduction

In the UK, the provision of home care for older individuals, especially those residing alone, is garnering increased focus as the demographic composition shifts towards an aging population. Government statistics from the UK indicate a rise in the proportion of the population aged 65 + from 16.4 % in 2011 to 18.6 % in 2021. Notably, 97.3 % of these elderly individuals reside in private residences, with 30.1 % living independently as reported in [9]. Compared to their counterparts in public institutions like hospitals and nursing homes, elderly individuals living in their homes generally receive less routine and emergency medical care. Although elderly cohabitants may benefit from mutual support and assistance, those living alone face greater challenges in accessing timely medical aid. While the National Health Service (NHS) endeavours to provide comprehensive medical support and care to the elderly at no cost, and various private care services are available for a fee, these measures remain insufficient for the escalating needs of the aging population. In response, the UK government announced a strategic initiative in 2019 to invest significantly in the development of care robots, aiming to engineer devices that are safe, reliable, and respectful

* Correspondence to: Warwick Manufacturing Group, University of Warwick: Coventry, CV4 7AL, UK.
** Corresponding author at: School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield, UK;
E-mail addresses: ruomeng.zhang.1@warwick.ac.uk (R. Zhang), jims.marchang@shu.ac.uk (J. Marchang).

https://doi.org/10.1016/j.ish.2025.06.003
Received 6 January 2025; Received in revised form 25 June 2025; Accepted 30 June 2025
Available online xxxx

of user privacy [15]. These robots are designed to perform essential functions such as mobility assistance, task completion, continuous health monitoring, and medication management, thereby enhancing the quality of life for the elderly. Furthermore, these robotic aids are envisioned to support other vulnerable groups, including individuals with disabilities and children, significantly alleviating the burden on medical and nursing professionals. This technological advancement not only promises to fulfil the increasing care demands of the elderly population but also aims to reduce the financial strains associated with elder care. However, the challenge in deploying such technology extends beyond the robotic execution of commands—it critically encompasses ensuring high safety standards. Care robots must be robust against a spectrum of cyber threats while safeguarding user data and privacy. Achieving this necessitates the implementation of advanced network security defences and stringent authentication and access control protocols.

In the current robotics sector, the identity authentication methods employed by robots include password, biometric, token, hardware, and behavioural authentication. However, these mainstream methods often lack convenience for care robots. They either fail to provide continuous authentication, impose interactive requirements on users, or necessitate the wearing of devices, which can be cumbersome. There are different methods for continuous authentication as highlighted in [5] and it is critical in ensuring continuous verification of the users. Given that the primary users are often elderly or vulnerable groups, care robots require an authentication scheme that does not rely on user cognition, demand user involvement, or impede user actions. Additionally, due to high security demands, this method must enable ongoing verification of user identity, guarantee the immediate termination of services upon user departure to ensure that user privacy remains uncompromised. Given the requirements, applying image recognition techniques for user identity verification emerges as a highly suitable option, as explored by [44]. This approach necessitates minimal user interaction with the camera and eliminates the need for users to remember passwords or wear any devices. While image recognition has been successfully utilised in other fields for biometric and behaviour-based recognition, its application in robotics remains unexplored. The objective of this study is to leverage existing image recognition techniques and assess their feasibility for continuous authentication in care robots. If proven viable, the solution would allow the robot to automatically and continuously verify the user's identity, requiring only the user's presence within the robot's field of view. However, the study is aiming at using MoveNet approach, so that the actual face or video is not recorded, but only the data points are collected in a non-intrusive way for user identification. This would obviate the need for users to consciously perform any authentication-specific actions, thereby ensuring that the care robot could effectively serve a broad range of vulnerable individuals, including elderly users with varying needs. The assumption of this study is that the users can at least walk and not bound in a wheelchair.

1.1. Research aim and objectives

The primary research aim of this study is to explore the effective application of MoveNet for the continuous authentication of care robots. This will be achieved by employing MoveNet models and other machine learning algorithms to accurately and consistently identify users within a ROS (Robot Operating System) environment. To accomplish this aim, the study has established the following specific objectives:

1. Design experiments that replicate real-world conditions and collect joint position data from different users while walking. This data will be used to train and validate using various machine-learning models. In the process, additional data features from the existing dataset will be extracted to improve the effectiveness of the models.
2. Assess the performance of various MoveNet models in detecting users' joint positions to ensure the selection of the most suitable model for this application scenario and analyse the efficiency and accuracy of different machine learning algorithms in identifying a user based on the processed data and select the most appropriate algorithm.
3. Implement the programme within a ROS environment to validate the feasibility of using MoveNet and machine learning algorithms for continuous authentication by care robots.

1.2. Background study

Artificial intelligence (AI) has become an integral part of the healthcare sector, offering powerful computational and learning capabilities that enhance clinical diagnosis and pathological analysis. For example, [37] developed an AI model utilizing convolutional neural networks to diagnose COVID-19 through X-ray imaging. Similarly, [38] applied deep neural networks to train a model for detecting blood samples indicative of COVID-19 cases. In the care sector, closely related to healthcare, AI-powered care robots are designed to assist individuals requiring daily support due to illness or aging. The UK’s [35] defines "care" as the daily assistance and supervision that enable individuals to live independently. High-quality care must meet several standards, including the timely provision of medical services, ensuring user safety, accommodating user preferences, and avoiding discrimination. Care robots aim to supplement, not replace, human caregivers, extending the range and quality of care services. They fall into two main categories: service robots and social robots [39]. Service robots assist with daily tasks such as household chores, mobility, personal hygiene, feeding, fall prevention, medication reminders, and health monitoring. They provide continuous oversight, often acting as substitutes for healthcare professionals. Social robots, in contrast, focus on emotional support and companionship. They help users maintain mental well-being by engaging in conversation, entertainment, and activities like cognitive training. Some robots integrate both functionalities. Sawik et al., [39] reviewed 21 representative care robots, most designed for elderly users. Notable examples include the ASTRO robot [19], which supports indoor walking and simple exercises, and the Bandit [17] and Gymmy [27] robots, designed for physical exercises and cognitive training. The PR2 collaborative robot [8] can retrieve objects using its dual-arm design, while the IRMA robot [48] assists in locating misplaced items using natural language commands. Equipped with a grasping function, IRMA can deliver items directly to users.

Social robots like Pepper [49] enhance interaction, engaging in conversations and providing reminders about item locations, to-do lists, and weather updates. Designed with a humanoid appearance and a display screen, Pepper uses facial expressions and body language to create a sense of familiarity. The Healthbots conversational robot [43] integrates vital sign monitoring, home safety features, medication reminders, fall detection, and assistance with phone calls and emails. Robots with single functionalities often struggle to meet comprehensive caregiving needs. Multi-functional robots, such as the Hobbit [20] and RAMCIP [16], integrate key caregiving features. These robots assist in retrieving and organizing items, monitoring health, providing medication reminders, and engaging users in games or conversations [19,32]. One of the most advanced robots in the field is Sophia, developed by [23]. Although not specifically designed for caregiving, Sophia’s lifelike human appearance and advanced interaction capabilities position her for potential caregiving roles. Sophia’s detailed facial features, dexterous hands, emotion recognition technology, and learning abilities allow her to perform complex tasks, interact naturally with users, and foster emotional connections. If adapted for caregiving, Sophia could monitor both physical and mental health, assist with intricate tasks, and provide companionship akin to that of human caregivers.

Most care robots are currently designed for indoor use, primarily in homes or private spaces. Outdoor applications, while less common, involve scenarios where robots remain close to users and may interact with multiple individuals. This highlights the primary user group, i.e., elderly individuals, and the typical application scenario: standard-sized indoor environments. This study focuses on these contexts, aiming to optimize the functionality and deployment of care robots to effectively meet user needs.

1.3. Continuous authentication

Authentication verifies a user's identity before granting access to resources or services [12]. In contrast, continuous authentication enforces verification throughout the user's session, from initial access until termination, making it suitable in managing highly sensitive and high-security service delivery and interaction scenarios where a user's presence may dynamically change. There are different types of authentication methods, and it can be classified into four categories: knowledge-based (e.g., passwords, PINs), possession-based (e.g., tokens, keys), attribute-based (e.g., biometrics), and location-based (e.g., IP address, GPS location) [2]. However, most of the authentication methods will not be adoptable in a care robot deployment settings because of the physical, psychological and mental condition of the elderly users. Token based often involve physical devices like keys or smart cards that users must possess to gain access, while certificate-based authentication relies on digital certificates that uses encryption, decryption keys and a certifying authority's digital signature to verify the identity. Password or PIN based sounds easy, but for the elderly it will be a challenging method and moreover, it is a one-time authentication for a session, so it is not possible for a continuous authentication. Two additional categories are physiological and behavioral biometrics, both of which involve unique biological characteristics. Physiological biometrics utilize physical traits such as fingerprints, palm prints, irises, facial features, voice patterns, or heartbeats [31]. Behavioral biometrics, on the other hand, rely on patterns in user behavior, such as typing speed, finger pressure, mouse movements, signature dynamics, or gait. Other approach like wo-factor authentication, is more secure in nature because it involves multiple methods to verify a user's identity. Among all, biometric authentication approaches using gait are non-intrusive and are seamless in nature. Moreover, the robot collecting the gait data rather than supplying it from user's smart devices should be the best approach because it doesn't involve additional communicating device that the user needs to carry.

Biometrics methods have a dominance in the area of delivering continuous authentication applications. The work of [6] reviewed the accuracy of these technologies, finding that voice recognition outperformed facial recognition, while ECG surpassed eye movement tracking, which in turn was more reliable than EEG. For behavioral biometrics, touch dynamics, stylometry, and keystroke dynamics achieved accuracy rates over 90 %, whereas gait recognition showed variability based on sensor data, and environmental sensing (e.g., IP address, devices) ranged between 80 % and 90 % accuracy. The paper of [2] explored continuous authentication within Internet of Things (IoT) environments, highlighting the integration of environmental and device data. It considers factors like device power consumption, hardware specifications, wireless signals, and GPS location contribute to authentication in IoT environment. Continuous authentication is regarded as a highly secure method for identity verification, it is crucial for care robots that handle sensitive user data. These robots operate in personal settings, accessing user privacy to deliver tailored care services. To ensure optimal care, they must understand physical conditions, preferences, and cultural contexts, minimizing stress or negative reactions. Consequently, cybersecurity is paramount to protect user data and prevent potential manipulation or harm. Given the healthcare sector's vulnerability to cyberattacks [33], robust authentication methods that minimize user involvement are essential, particularly for elderly users who may lack cybersecurity expertise. The paper explores and proposes identity authentication methods for care robots, emphasizing usability for elderly individuals. They found wearable devices, facial recognition, and gait recognition to be user-friendly, with gait recognition and dual-factor authentication offering the highest security levels. They proposed a dual-factor system combining facial recognition and smartwatch detection, achieving 99.39 % facial recognition accuracy and detecting smartwatches within 20 ms. While effective, limitations such as partial camera coverage and unsuitability for users unable to wear devices were noted. The work of [3] advanced this approach by developing a non-intrusive multimodal user recognition system for caregiving robots. Using facial features, voice characteristics, and skeletal data from walking patterns, their model achieved 100 % accuracy. Similarly, [7] used facial features, height, clothing colour, and leg posture for user recognition, demonstrating the utility of multimodal biometrics. The research work of [11] combined heart rate, gait, and respiratory rate data from wearable devices, achieving commendable authentication accuracy. However, their gait recognition system required user motion, posing limitations. The work of [28] identified challenges with facial recognition, such as delays and errors caused by hardware and motion, proposing voice recognition as a complementary method to enhance reliability. However, [1] focused on continuous authentication for collaborative robots, using hand movement recognition to prevent unauthorized access and ensure user safety. Moreover, [14] suggested RFID-based authentication for hospital care robots, integrating continuous health monitoring to minimize user interaction while maintaining secure authentication. Many studies emphasize gait recognition's effectiveness and reliability as a key component of continuous authentication for care robots. Combining gait recognition with other methods offers robust security, providing valuable insights for future research in this area.

1.4. MoveNet

MoveNet is a neural network model developed by Google, based on the TensorFlow framework, designed for human pose estimation [41]. It can detect the positions of seventeen key points on the human body by analysing video footage or real-time camera feeds. MoveNet is characterized by its lightweight nature, high accuracy, and speed. It is capable of running on a far range of devices, including computers and mobile devices. MoveNet offers six different models, ranging from relatively lightweight to more complex options, to cater to various needs [29]. The more complex models provide higher accuracy but operate at slower speeds and require more powerful devices. Nevertheless, all models can run on both CPU and GPU, although the processing speed is faster on GPS. They all can process video streams at over 30 FPS. The input requirements for video streams are minimal, with the ability to handle resolutions as low as 360×270 pixels, using only a standard RGB camera. Additionally, MoveNet performs well in low-light conditions and can detect key points when the person is side-facing or back-facing the camera, unless obstructed by other objects or outside the camera's view. It also supports multi-person detection, allowing it to identify and track key points for multiple users simultaneously when they appear within the frame. The TensorFlow framework, which MoveNet relies on, was also developed by Google. It is mainly used for data computation and machine learning and is currently one of the most popular frameworks in the field of machine learning.

MoveNet can detect a total of 17 key points, which include the eyes, ears, nose, shoulders, elbows, wrists, hips, knees, and ankles. It supports two basic output formats: one directly outputs the name of each key point, along with its absolute XY coordinates and confidence score for each frame of the video or single picture as described in Fig. 1 [45]. The other format visualises the key point positions by drawing them as points connected by lines on an image or video as shown in Fig. 2, the data is generated during the training of the proposed system in this paper. MoveNet has already been widely adopted across various fields requiring pose recognition, such as sport, fitness, wellness, and healthcare [45]. The work of [42] explored the application of MoveNet in the fitness domain, specifically developing a real-time yoga pose

```
console.log(poses[0].keypoints);
// Outputs:
// [
//    {x: 230, y: 220, score: 0.9, name: "nose"},
//    {x: 212, y: 190, score: 0.8, name: "left_eye"},
//    ...
// ]
```

Fig. 1. MoveNet keypoints extraction [45].

estimation method. This system was designed to provide guidance to yoga practitioners and correct improper poses, thereby preventing injury due to incorrect posture. They first utilised MoveNet to extract the key points of the human pose, and then employed a customised TensorFlow machine learning algorithm to interpret these key points. Using an artificial neural network.

(ANN) for training, they achieved a training accuracy exceeding 90%, although the accuracy during testing fell slightly below 90%. Interestingly, they found that certain poses, despite having fewer training samples, achieved accuracy rates close to 100%. Similarly, [36] conducted research aiming to use MoveNet to customise yoga training programmes for users. They employed MoveNet Thunder to obtain key points and used a convolutional neural network (CNN) model for training and testing, ultimately achieving a testing accuracy of 98%. The research of [4] applied MoveNet to estimate movements in Silat, a martial art. They also used MoveNet Thunder for data preprocessing, followed by ANN for training and classification, achieving an impressive accuracy of 97%. However, their dataset was limited to only eight types of images. The paper of [26] utilised MoveNet to detect instances of campus bullying. They processed 400 min of personal behaviour videos and 20 min of physical abuse videos with MoveNet, categorised the data into 13 behaviour types, and used ANN for training and prediction, resulting in an accuracy of 98%. In the health sector, [18] focused on using smartphones to detect human motion as part of a solution for identifying stroke patients. They experimented with three pose estimation models—MoveNet, PoseNet, and BlazePose. Although BlazePose could recognise the most key points, it was the slowest, while PoseNet was as fast as MoveNet but produced lower-quality data. MoveNet emerged as the most suitable pose estimation model for smartphone videos, achieving a classification accuracy of 96% for the upper body of stroke patients. However, the accuracy for lower body detection was less satisfactory, possibly due to insufficient data. The work of [47] also conducted a comparison of pose estimation models. They tested MoveNet Lightning, MoveNet Thunder, OpenPose, and DeepLabCut, and similarly concluded that MoveNet Thunder delivered the best performance. The findings from these studies indicate that MoveNet is one of the best-performing models in pose estimation, demonstrating its reliability. Among the various models, MoveNet Thunder consistently emerges as the most effective. Additionally, most studies that utilise MoveNet have opted to pair it with neural network algorithms, providing clear guidance for the selection of machine learning algorithms in this research. These studies pave ways to use similar MoveNet data for user identification by enabling the assistive or robotic systems to directly gather gait data without the need of users supplying any secret verification code and use the gait data for unique user identification.

1.5. Gait and distance recognition

Gait recognition is a technology that identifies individuals by analysing the way they walk. Each person has a distinct walking pattern, characterised by factors such as step length, speed, and the changing angles of their legs and arms. These features are unique and difficult to replicate, making gait recognition highly suitable for identity authentication. It is a non-invasive behavioural biometric method, frequently used for continuous authentication. Beyond identity verification, gait recognition also has applications in medical care, where it can assist in detecting walking-related illnesses, support patient recovery, and in physical exercise, where it can help individuals assess and correct their posture.

The concept of gait recognition was first introduced in 1994, and early research primarily focused on methods for recognising gait through video analysis (Wan et, al, 2018). As technology advanced, the use of depth cameras to construct 2D or 3D models for analysing gait characteristics gradually emerged. Concurrently, sensor-based methods for data collection gained popularity. These include approaches that utilise body-worn sensors to capture acceleration, floor sensors to measure pressure, and radar systems to track gait positions. In recent years, the focus of gait recognition research has shifted towards more routine and lightweight data collection methods, alongside efforts to achieve higher recognition accuracy. Leveraging sensors from everyday electronic devices—such as computer and mobile phone cameras, smart devices, and wearable technology—has made gait data collection

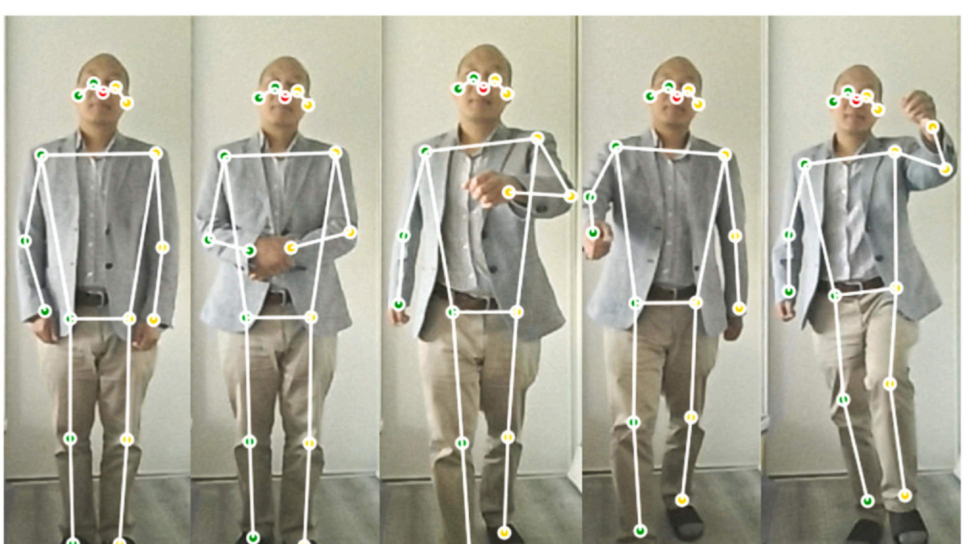

Fig. 2. MoveNet Skeleton Extraction
(Source: Authors Data).

easier, more accessible, and cost-effective. Given the relatively low usability and high cost associated with specialised sensors, video-based data collection remains a more mainstream research direction. In a gait recognition survey, [40] evaluated 15 widely used public video gait databases spanning from 2001 to 2020. The number of participants in these databases varied significantly, with the smallest including only 20 individuals and the largest comprising over 60,000. However, the number of sequences per individual typically ranged between 10 and 20 or more. Most of the databases focused on recording silhouettes, while the more recent ones included skeleton data. The majority of the data collection environments were indoors. The databases also differed in the number of viewpoints recorded, with some capturing gait from only one viewpoint and others from as many as 25 [40]. Additionally, they included varying degrees of other factors, such as balanced gender ratios and even age distributions among participants. Some databases recorded gait under different conditions, such as when participants were empty-handed, carrying a bag, or wearing a coat. Others captured a variety of clothing types, walking surfaces, and footwear. Some databases also recorded gait at different walking speeds, and a few utilised infrared captures to account for nighttime conditions. By summarising the characteristics of existing public gait databases, it is possible to analyse the expected features of the dataset in this study and identify key factors that need to be considered during its design.

Once the video dataset is obtained, the next step is to extract features from it. Gait recognition through video can be categorised into two main approaches: contour-based and skeleton-based [40]. The contour-based approach involves segmenting the human silhouette from the video and then extracting features such as shape, variation, and gait cycle. This method does not require labelling the human body before feature extraction, but it is highly susceptible to external factors such as clothing shape and lighting conditions. In contrast, the skeleton-based approach begins by labelling the skeleton of the human body, after which the relative positions of the joints, the angles between them, and the changes in these joints during movement are extracted [40]. This method is less affected by external factors and yields more accurate data. However, it requires greater computational power and involves more complex data processing steps. Currently, with computational power no longer being a primary constraint and pose detection techniques achieved good accuracy. The focus has shifted to skeleton-based gait feature extraction. In their gait recognition study of [21], selected 18 joint points on the body and used their x and y coordinates along with confidence scores as features. The work of [30] directly utilised the 18 2D body joints extracted by OpenPose, from which they estimated 14 3D joint positions as features. They further enhanced their model by designing three additional features: joint angles, limb lengths, and the dynamics of joints during motion. The research of [10] took a different approach to enhance the model by calculating the average 3D coordinates of all the joints, the Euclidean distances from each joint to the average coordinates, and the rotations around the y-axis relative to the camera, derived from the coordinates of the shoulder and hip joints. To improve tracking, they implemented a sliding window technique to accumulate data and track the trajectory of the joints over time. Nguyen et al., [34] concentrated on feature extraction specifically from the legs to more efficiently identify abnormal gait patterns associated with diseases. They calculated the angles between the leg joints, the lengths of the skeletal segments, the lateral distance between the joints of the left and right legs, and the angle between the left and right feet. From the literature, it is evident that skeleton-based gait recognition features primarily focus on the key joints of the human body and their interrelationships, such as distances, angles, and dynamic changes. This aligns closely with the keypoint information provided by MoveNet, highlighting the potential of using MoveNet for gait recognition. The final step in gait recognition is classification. In their survey on gait recognition, [46] introduced five primary classification methods. The most straightforward approach is the distance method, which involves setting a threshold and then calculating the distance between the collected gait data and the data in the database to measure similarity. Classification is then performed based on the threshold. The second method is correlation calculation, with common metrics such as Pearson correlation coefficient. The absolute value of the correlation coefficient indicates the degree of correlation or dependency between datasets, with values closer to 1 indicating a stronger correlation. The third method, and currently the most prevalent, is the use of machine learning. Gait recognition often employs supervised learning algorithms such as support vector machines (SVM), decision trees, and neural networks. The fourth method is the Hidden Markov Model (HMM), which is also widely used, particularly for handling temporal relationships within gait data. The fifth and final method is the Bayesian model, which calculates the probability that a data vector matches the data in the database and classifies the data accordingly. In the feature selection studies mentioned earlier, most researchers employed supervised machine learning algorithms to classify their data, with CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), and LSTM (Long Short-Term Memory) being the most frequently cited algorithms. The findings from these studies consistently indicate that CNNs achieved higher accuracy compared to the other algorithms.

2. Research methodology

My ontological position asserts that the classification of users' gait data can be objectively measured and analysed as a tangible entity. By employing MoveNet and other machine learning tools as scientific methods for analysing this data, I aim to achieve the objective of user identity verification. In this research, I adopt a positivist approach. I contend that through the quantitative collection and analysis of data, it is possible to obtain objective knowledge about the effectiveness of using MoveNet for continuous authentication in care robots. By designing experiments to collect data and developing programmes to analyse the results, I can ascertain the objective feasibility of my research aims. Based on this paradigm, the study has been designed with a comprehensive experimental process to collect data, analyse it, and compare model performance. Initially, gait data from different users in specific environments will be collected through experiments. MoveNet will then be used to extract key features from this data, generating a dataset. Various machine learning algorithms will be applied to classify these features, with output scores across multiple dimensions to evaluate and compare the accuracy of each algorithm. Finally, these techniques will be integrated into a ROS environment, where the performance of the solution will be tested using a separate test dataset. By assessing the final accuracy and comparing it with existing research solutions, the usability and effectiveness of the proposed solution can be determined.

This study employed a single-subject analysis, focusing exclusively on evaluating the application of MoveNet in continuous authentication for care robots. The research concentrated solely on MoveNet's gait recognition capabilities and did not explore other biometric technologies or multifactor authentication methods.

2.1. Data collection

This study employed quantitative methods for data collection, focusing on the authentication of user gait data processed by MoveNet. The collection of video data of users' gait through quantitative means is essential to achieving this objective. This approach ensures an objective and controlled research process, aligning with a positivist methodology. Quantitative data collection allows for precise control of variables, minimising the influence of extraneous factors and reducing errors in the comparison of results, thereby yielding more accurate outcomes. Additionally, accurately controlling variables during the data collection process enhances the repeatability of the experiment. This ensures that the data collection process can be easily replicated in future studies if necessary. A single method of data collection was employed in this study, utilising a mobile phone camera to record videos of participants' gait at a specific location. In situations where data need to be collected under highly consistent conditions, using a single data collection method is more likely to ensure uniformity in the data. Moreover, this study focuses on achieving continuous authentication using just one technique and device, making the use of multiple methods unnecessary. Drawing on the analysis of mainstream public gait databases in the literature review, gait data were collected from 10 participants, with an equal gender balance, due to the small scale of the study. In line with the sample sizes used in public gait databases and considering the limited number of participants, 40 gait video samples were collected from each participant. To further enhance the dataset, a sliding window technique was applied during the data processing stage, expanding the total dataset size and ensuring that the study's data were more comprehensive.

Data collection for this study took place between 25 July and 5 August 2024, following ethical approval to ensure that the experiment was conducted by ethical guidelines. The data collection period was intentionally kept brief and used a cross-sectional experimental design. A consistent time of day i.e. afternoon was selected for data collection. This approach minimised the influence of external factors, such as weather, and maintained consistent light levels. Additionally, collecting data in the afternoon ensured that participants were awake and energised, reducing the likelihood of unconventional gait patterns resulting from fatigue.

2.2. Data analysis

The test site for this study was set up in a relatively quiet, flat area within a typical building. This location was chosen to ensure participant safety and to minimise disturbances or potential injuries from unexpected events during the experiment. Moreover, an indoor site was selected to simulate real-world application scenarios, as care robots typically operate in hospitals or users' homes, making the test environment a good match for their intended working conditions. However, the data collection process in this study involved user privacy, which posed certain challenges in recruiting volunteers. Ideally, a participant pool of 20 individuals would have been optimal, but due to privacy concerns, there was insufficient willingness to participate. To address these concerns and protect user privacy, the data collection process did not capture the lower half of the participants' faces. Masks were provided to safeguard their identities. Ultimately, ten participants were successfully recruited, which was sufficient to meet the data requirements of this study. All participants in this study were non-participatory, meaning they did not actively influence the experimental process. This approach was adopted to minimise any potential subjective impact from the participants, thereby ensuring the objectivity and consistency of the data. The decision to use quantitative methods for data analysis in this study is grounded in the same rationale with data collection. Quantitative analysis effectively minimises errors in the analytical process. Specifically, well-established and widely accepted scoring criteria exist for evaluating the performance of machine learning algorithms. By calculating the same performance metrics for different models, the differences between each model can be more accurately and objectively assessed. This approach also facilitates easier comparison of this study's results with those of another research. Furthermore, the use of a standardised quantitative analysis method allows for the accurate reproduction of research results, enabling precise control of variables when optimising the research process to identify more effective optimisation strategies. Deductive reasoning was applied in this study. Based on a review of relevant literature, the hypothesis was formed that gait data processed using MoveNet could be used to identify users, thus enabling its application in the continuous authentication of care robots. The data collection and analysis steps were designed around this hypothesis to test its feasibility.

2.3. Method of data collection

The data for this project was collected through experimental methods, with the primary objective of obtaining a sufficient number of video recordings of the participants' gait. The data collecting process achieved the first research objective. The experiments were typically conducted between 1:00 and 4:00 PM, a time when the light intensity at the experimental site was adequate, and the participants were in optimal condition. The experimental process began with meeting the participants at the site and signing relevant documents, including the informed consent form. Participants were then instructed to start at a designated point A and walk in a natural and comfortable manner to the end point B. A video camera was positioned at video point C, directly aligned with the straight path from point A to point B. Each participant made 40 round trips between points A and B, maintaining a consistent walking pattern throughout. The camera recorded only the participant's complete walk from point A to point B, capturing frontal gait footage as the participant faced the camera directly. Any clips where the participant glanced to the side or made extraneous hand and leg movements were excluded from the final dataset. This process resulted in 40 frontal gait videos per participant, yielding a total of 400 videos for the subsequent training and evaluation of the machine learning algorithms. In this study, the speed of walking of the users are not captured, but all the participants walk in their usual and normal walking style on the flat surface of the environment described in Fig. 3.

Additionally, three of the ten participants were selected to record an extra 10 videos each, following the same method, for testing purposes. The researcher also recorded 10 test gait videos to be used for unauthorised person identification testing. The test site was set up in a relatively quiet, windowed hall within an ordinary building. This setup ensured normal lighting conditions and minimal disturbances, allowing the experiment to proceed without interference from others. The hall was approximately 70 square metres in size, unobstructed, and carpeted for noise reduction, providing ample space for the experiment while ensuring the participants' safety by preventing obstructions or falls. The straight-line distance between the start points A and the end point B, used in the experiment, was 6 m, reflecting the typical application scenario for care robot deployment in a room as shown in Fig. 3. In an

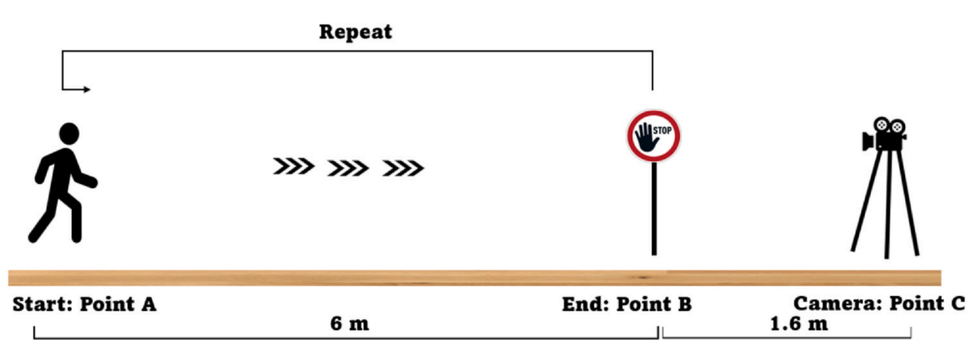

Fig. 3. Data Collection Experiment Environment.

average home, 6 m generally represents the maximum unsegregated distance within a room, which aligns with the care robot's requirement to recognise users within this range. Additionally, the distance from end point B to camera position point C was 1.6 m, ensuring that the participant was close enough to remain fully visible within the video frame when stopping at point B, without any body parts extending beyond the frame.

A standard Android mobile phone was used as the camera for this study. This choice was made because MoveNet is designed to run on a wide range of devices, and the study aimed to minimise the hardware requirements for the care robot. By using a mobile phone camera with a commonly available and lower-end hardware configuration, the study ensured that the solution would be more accessible and widely applicable. The video format was set to MP4, the most common format, with a resolution of 1080p, 30 fps, and a 16:9 aspect ratio. These settings were configured and recorded using the phone's default camera software. This approach ensured sufficient clarity while keeping the video quality and file size manageable, enhancing the scalability and broader applicability of the research solution. During filming, a nine-panel auxiliary grid was used within the camera frame, aligned with wall and floor markers, to maintain a consistent background and ensure that participants were positioned identically in each recording. Moreover, the camera was set on a fixed hight (1.4 m) tripod to ensure stability. Immediately after recording, all video files were uploaded to the university’s cloud storage, where they were encrypted for security. The local files were deleted, with temporary downloads only occurring as needed for analysis.

Participants were recruited from individuals known to the researcher, with efforts made to ensure an equal number of participants from different physiological sexes. The study did not include participants of varying ages or those with physical impairments, because inviting vulnerable groups would not have met the ethical requirements of this study. All 10 participants were aged between 20 and 30 years, in good physical and mental health, with no apparent illnesses or disabilities. There were no specific dress code or hairstyle requirements for the participants, as the gait videos collected were intended for a skeleton-based gait recognition study. Additionally, to validate MoveNet's single-user posture tracking capabilities, the experiment did not prevent others from passing through the test site. However, if someone walked in front of a participant during video recording, that segment was excluded and re-recorded to ensure the data continuity. As of the submission of this paper, no participant has requested to withdraw from the study.

2.4. Method of data analysis

The data analysis method in this study was divided into four steps: feature extraction and creation using MoveNet, data preprocessing, training and evaluation of different machine learning algorithms, and model validation in a ROS environment using test data as shown in Fig. 4. All steps of the data analysis were executed through Python 3, as it is the most widely used programming language in the field of machine learning and is also the language recommended for MoveNet's official data preprocessing methods. The data analysis code utilised several common and essential Python libraries, including TensorFlow, scikit-learn, pandas, OS, matplotlib, and numpy.

The dataset generation steps from the input video source are described in Fig. 5. During the feature dataset generation phase, the collected video files were batch-imported into the program and processed individually. Each video generated a separate dataset in CSV format. For each video, the frames were extracted as static images and processed sequentially. The MoveNet model was then invoked to determine the coordinates of the shoulder and hip keypoints within the image. The excess portions of the image were cropped to ensure that the skeleton size of the person in each frame remained consistent, thereby minimising coordinate discrepancies caused by changes in the person's distance from the camera. MoveNet subsequently identified and outputted the X and Y coordinates of 17 keypoints in the current frame, along with their confidence scores. These keypoints were then used to calculate and generate new feature values. In this study, a new feature, focusing primarily on the distances between keypoints are considered. Considering the use of a 2D skeleton model, speed data during walking was not included due to its potential inaccuracy. These additional features enhanced the discriminative power of the user's gait data, improving the model's ability to accurately identify individuals. Once all features were generated, the data was recorded in a CSV file, with each row representing a frame and the user ID label placed in the first column. The data was then checked to exclude rows with low confidence scores or zero-valued features.

All gait video files were processed by six different MoveNet models, generating separate CSV datasets for each model. These datasets were subsequently used for the next step: machine learning training and evaluation. After identifying the most performant model, it was retrained and evaluated using the datasets generated by the six different MoveNet models, with the scores compared to select the most suitable MoveNet model for this study. This approach ensures a comprehensive evaluation of model performance and identifies the selection of the best MoveNet variant for continuous authentication in care robots. Also, this approach achieved the second research objective. During the machine learning training and evaluation phase, this study selected eight commonly used machine learning algorithms based on the findings from the literature review on gait recognition and the application of MoveNet. These algorithms were compared to assess their performance, with the highest-scoring algorithm being selected for further use. The eight algorithms are Support Vector Machine (SVM), Random Forest, k-Nearest Neighbour (k-NN), Naive Bayes, Neural Networks (NN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM). More details structure of the structure and hyperparameters of NN, CNN, RNN and LSTM are given in the Fig. 6, Fig. 7, Fig. 8 and Fig. 9 respectively. The hyperparameters used for optimising NN: Adam (learning rate = 0.001), Loss function: categorical cross-entropy, Batch size: 32, Epochs: upto 50 epochs are

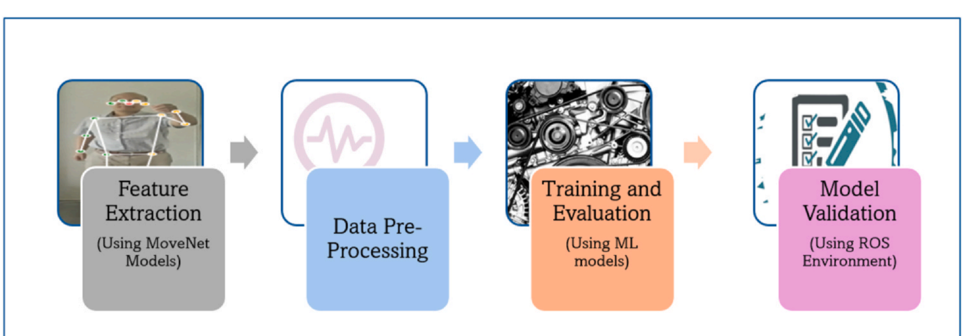

Fig. 4. Block Diagram of Feature extraction, Data Pre-Processing, Training and Validation Process.

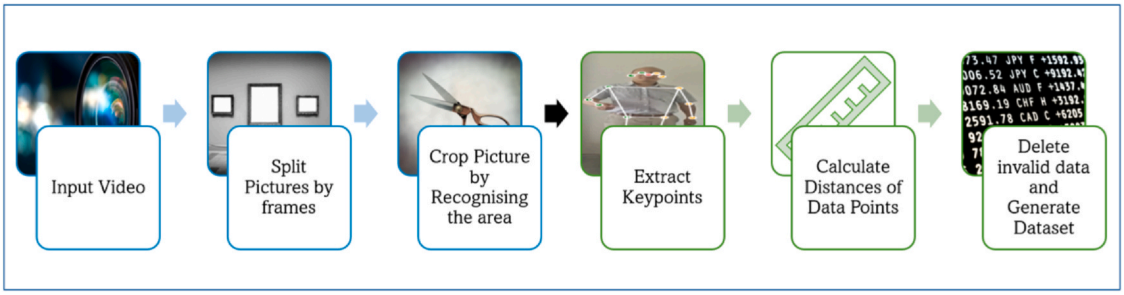

Fig. 5. Proposed MoveNet Dataset Generation Flow Diagram.

experimentally validated. The CNN reached saturation with an optimiser of: Adam (learning rate = 0.0001), Loss function: categorical cross-entropy, Batch size: 32 and Epochs: upto 50 epochs as per experimental validation. The best fit of learning rate for RNN and LSTM is 0.001, however the dropout rate for LSTM is 0.20 (to prevent overfitting) while the rest of the other network parameters are the same as that of NN and CNN. In regard to the random forest as shown in Fig. 10, randomised search technique using RandomisedSearchCV with n_iter = 50, cv = 5, verbose = 2, n_jobs = −1, scoring = 'f1_weighted', and random_state = 42 are considered for hyperparameters.

Following the recommendations from the literature review, the dataset generated by MoveNet Thunder was firstly utilised for comparing these machine learning algorithms. At the beginning, the entire dataset was imported into the application, and a standard sliding window with a window size of 10 was applied to further augment the feature set. The choice of a window size of 10 was made to capture temporal patterns in the gait data, which helps in better understanding the dynamics of movement over a short sequence of frames. Next, the data was pre-processed using encoding and scaling techniques before being split into training and testing sets. Each of the eight machine learning algorithms was then used to train models on the prepared data, followed by an evaluation of their performance. All generated models were evaluated using standard machine learning metrics: accuracy, precision, recall, F1 score, and the confusion matrix. These metrics are widely recognised for assessing the performance of machine learning models. In addition to these metrics, the time taken for prediction of each model was also considered to assess the processing speed of the models. This comprehensive evaluation approach ensured that both the accuracy and efficiency of the models were thoroughly analysed.

Finally, the test programme was created and executed as a node in the ROS environment on Linux. First, a data processing node was implemented using code consistent with the data preprocessing and dataset generation steps, employing the highest-scoring MoveNet model to generate datasets from the test video files while omitting the label column. This process simulates how a care robot would extract data from the footage captured by its camera. Next, an identity recognition node was activated, which called the dataset file and applied the highest-scoring machine learning model. The same data preprocessing steps used during the training phase were employed to predict and output the identity and associated probabilities for each data entry. This mimicked the process of a care robot identifying a user. The accuracy and potential issues of the model were then assessed by comparing the predicted results and their associated probabilities. This evaluation was crucial for determining the feasibility and effectiveness of the proposed solution in real-world applications.

3. Results and analysis

The gait dataset in this study comprised 10 adults between the ages of 20 and 30, with an equal distribution of five biological females and five biological males. The gait videos varied in duration, with the fastest participant taking approximately four seconds to walk from point A to point B, and the slowest participant taking around eight seconds. Using the sliding window method, these videos generated a total of 135,873 data points, averaging 13,587 data points per participant and the data distribution of the participants are shown in Fig. 11. The participant with the least amount of data contributed 10,356 data points, while the one with the most contributed 17,956.

MoveNet offers six models for joint point data extraction for a single person: MoveNet lightning (MoveNet lightning int8, MoveNet lightning

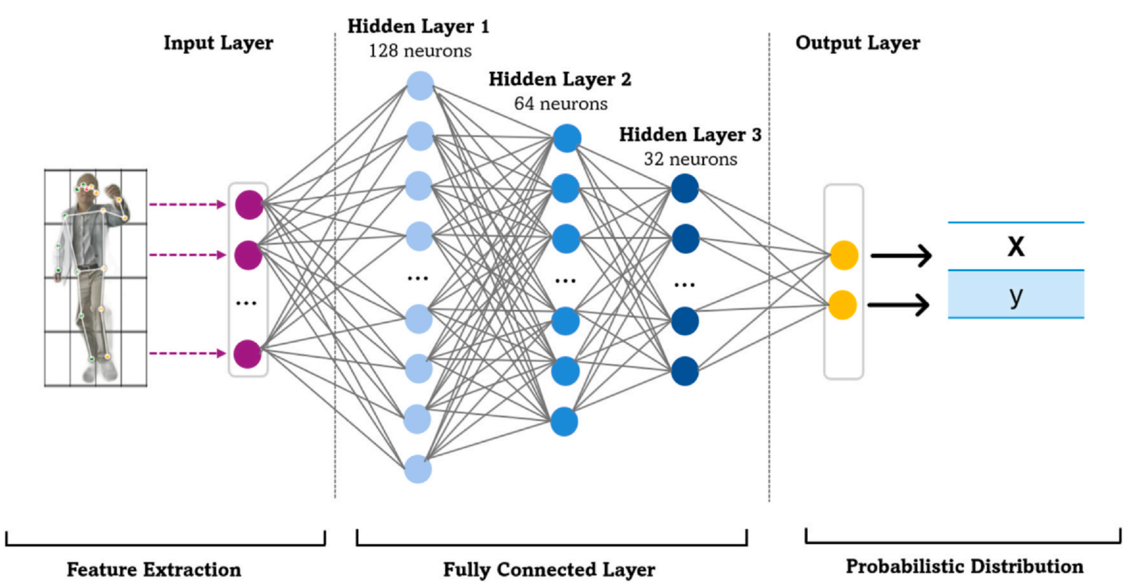

Fig. 6. The NN Framework used for the Proposed System.

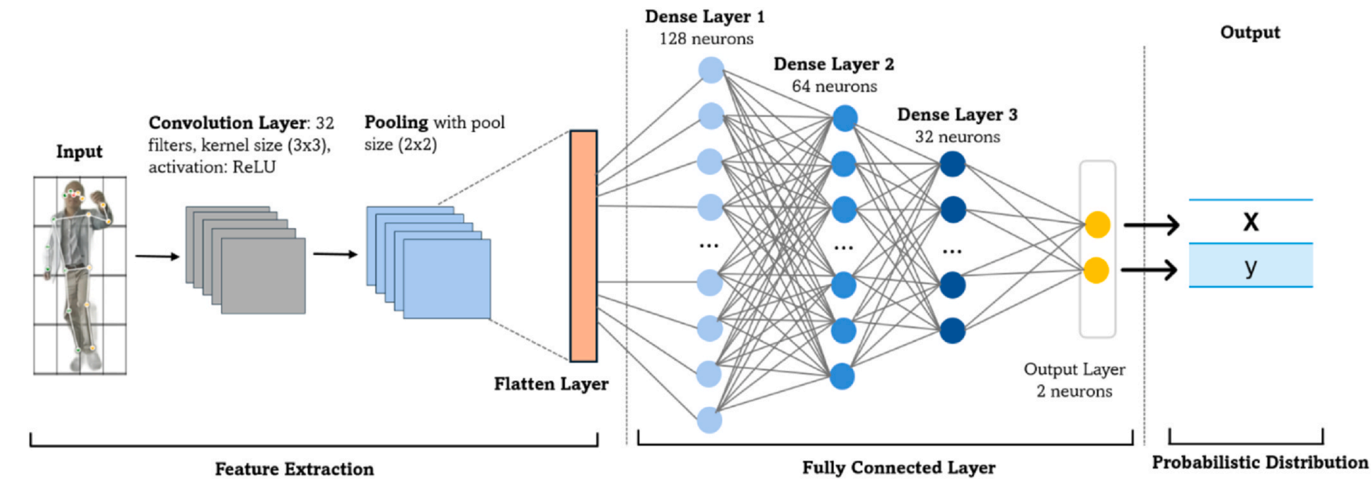

Fig. 7. The CNN Framework used for the Proposed System.

f16), MoveNet thunder (MoveNet thunder int8, and MoveNet thunder f16). According to MoveNet's official performance table, of the four models listed in Table 1, the Lightning int8 model is the smallest in size and operates the fastest, while the Thunder f16 model is the largest in size and operates the slowest. Although official performance values for the Lightning Standard and Thunder Standard models are not provided, it is reasonable to assume that the Lightning Standard model is the smallest and fastest overall, while the Thunder Standard model int8 has a size larger than the Lightning f16 but smaller than the Thunder f16, with a speed slower than the Lightning f16 but faster than the Thunder f16.

To evaluate the performance of datasets generated by different MoveNet models, the same Neural Network model is used for training and assessing each dataset with 50 epochs. The results from the neural network model revealed that in terms of evaluation time, the Lightning model had the longest prediction time at 1.06 s as shown in Table 2. The Lightning f16 model, with a prediction time of 0.96 s, was the fastest among the Lightning series. The Lightning int8 model was only 0.04 s faster than the Lightning model, while the remaining three models had similar prediction times. When examining the metrics of accuracy, precision, recall, and F1 score, all models yielded identical results when rounded to four decimal places. Even the unrounded metrics showed only slight differences. Considering these four metrics as a whole, the Lightning f16 and Thunder models demonstrated the highest performance, both achieving scores of 99.9%. Following closely were Lightning, Thunder int8, and Thunder f16 models, each with metrics hovering around 99.8%, with minimal differences between them. The Lightning int8 model, however, had the lowest scores across all metrics, at 99.64%, showing a 0.36% difference compared to the highest-performing models.

Additionally, confusion matrix heatmaps were generated for the evaluation results of the neural network models trained on datasets produced by different MoveNet models. Due to the large number of heatmaps, only the confusion matrix heatmaps for the highest-scoring

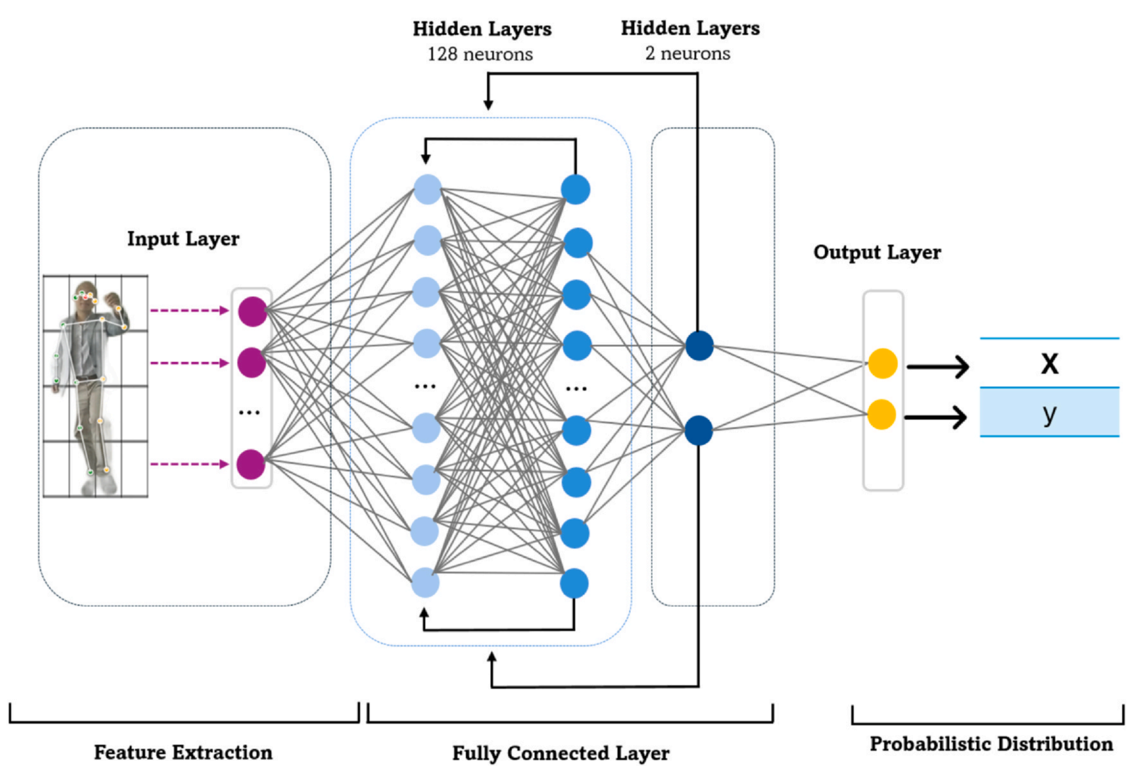

Fig. 8. The RNN Framework used for the Proposed System.

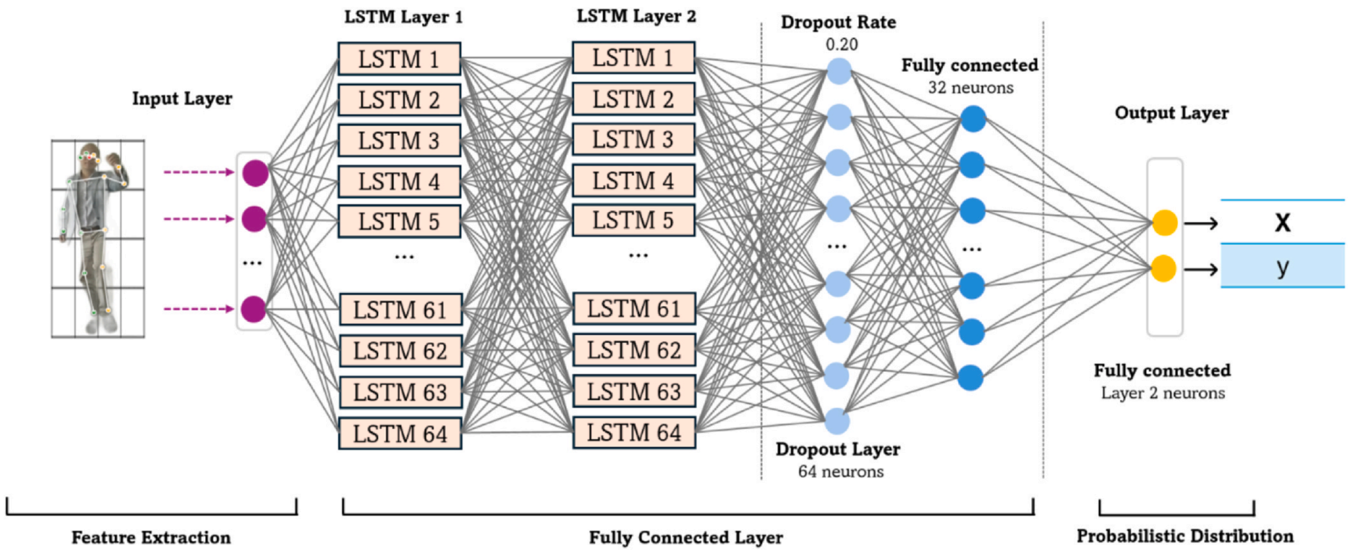

Fig. 9. The LSTM Framework used for the Proposed System.

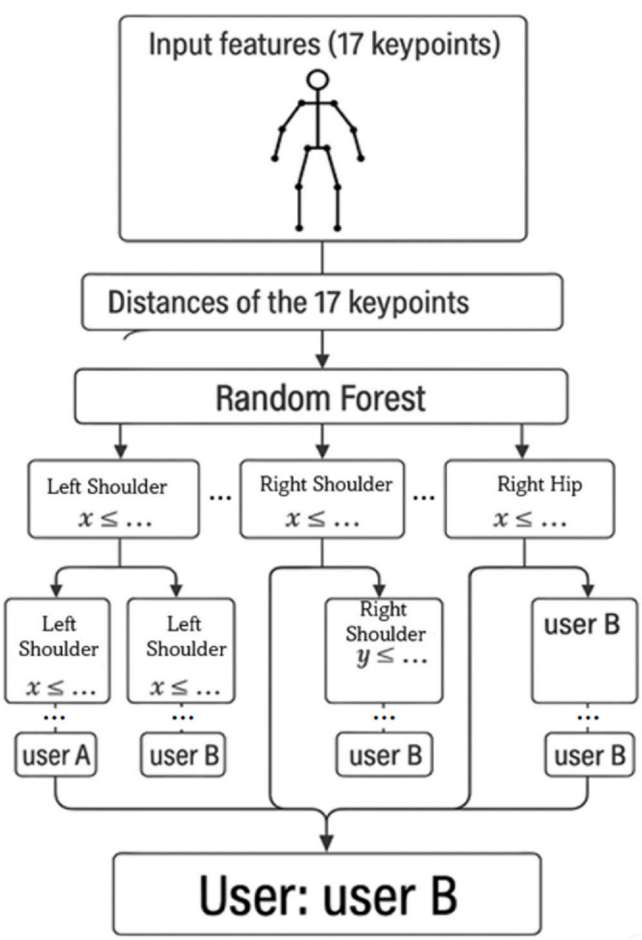

Fig. 10. The Random Forest Architecture Framework used for the Proposed System.

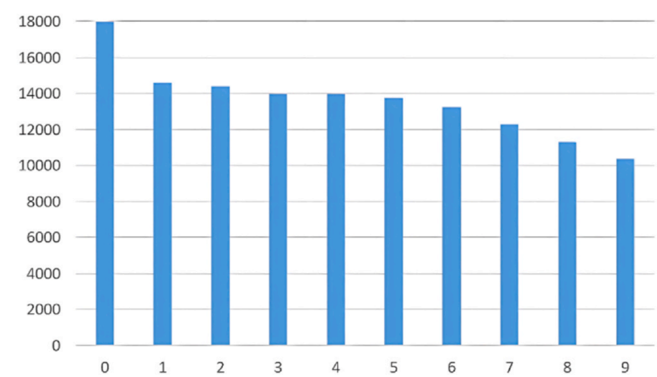

Fig. 11. The Users' data volume Distributions.

models, Lightning f16 and Thunder, are presented in Fig. 12 and Fig. 13 respectively. These two models show some differences in accuracy when classifying different users. Firstly, the recognition accuracy for all users is very close for both the Lightning f16 and Thunder models, with both achieving over 99.68% accuracy. However, the accuracy variance among different users is more pronounced with the Thunder model, where the lowest accuracy is 99.68% and the highest is 100%. In contrast, the Lightning f16 model shows a smaller range between the lowest and highest accuracies, with a difference of 0.29%. Although Thunder has a slightly higher average accuracy at 99.9%, compared to 99.895% for Lightning f16, the two models exhibit nearly identical accuracy rates when classifying the same users, with differences typically below 0.1%. Only two users showed larger discrepancies, with accuracy differences of 0.13% and 0.32%. In the first case, Thunder achieved higher accuracy, while in the second case, Lightning f16 outperformed Thunder.

3.1. Machine learning models

Ss In the machine learning model evaluation phase, the dataset generated using MoveNet Thunder was employed to train and evaluate all the machine learning models. The evaluation process involved calculating prediction time, accuracy, precision, recall, F1 score, and plotting the confusion matrix. In this part, all the neural network models were trained with 10 epochs. Among all the machine learning models, the shortest prediction times were achieved by Random Forest, SVM, Naive Bayes, and NN, all of which completed the predictions in under 1.5 s as shown in Table 3. KNN, CNN, and RNN also performed the predictions in less than 30 s, while LSTM was the slowest, taking around 68 s. In terms of accuracy, precision, recall, and F1 score, the differences of them across most models were relatively small, generally within a 2% range. However, SVM stood out as the model with the lowest performance, with an F1 score of only 13.23%. It also exhibited the largest disparity among the four metrics. The four best-performing models—Random Forest, KNN, NN, and CNN—achieved nearly

Table 1
MoveNet models performance [29].

Model	Size	mAP	Latency (Second)		
			Pixel 5 – CPU 4 threads	Pixel 5 - GPU	Raspberry Pi 4 – CPU 4 threads
MoveNet.Thunder (f16 Quantised)	12.6MB	72.0	0.155	0.045	0.594
MoveNet.Thunder (int8 Quantised)	7.1MB	68.9	0.100	0.052	0.251
MoveNet.Lightning (f16 Quantised)	4.8MB	63.0	0.060	0.025	0.186
MoveNet.Lightning (int8 Quantised)	2.9MB	57.4	0.052	0.028	0.095

identical scores across all four categories. Random Forest emerged as the top performer with a highest F1 score of 99.93%, followed by CNN and NN with scores of 99.89% and 99.86%, respectively. Despite being the most complex model, LSTM yielded only moderate performance, with an F1 score of 89.41% and an accuracy of 89.67%.

The Random Forest model accurately identified 4 users with 100% precision, with an additional 4 users achieving accuracy rates exceeding 99.9%, and the lowest accuracy recorded at 99.78% and the confusion matrix of the features are shown in Fig. 14. In contrast, the KNN model did not achieve 100% accuracy for any user; its recognition accuracy generally ranged between 99.7% and 99.8%, with only one user exceeding 99.9%. KNN's confusion matrix is highlighted in Fig. 15. While Fig. 16 and Fig. 17 represents the confusion matrixes of NN and CNN respectively. The NN model, while achieving accuracy rates of 99.9% or higher for five users, did not reach 100% accuracy for any user. Aside from one user with a minimum accuracy of 99.54%, all other users were recognised with accuracy rates of 99.75% or above. The CNN model, on the other hand, accurately identified two users with 100% precision and three users with accuracy rates above 99.9%, with the lowest accuracy still reaching 99.72%.

For providing a more intuitive comparison, this study conducted further training on the two best-performing neural network models, namely the NN and CNN models as shown in Fig. 18 (NN model with up to 50 epochs), Fig. 19 (CNN model with up to 50 epoch), Fig. 20 (NN with no sliding window) and Fig. 21 (CNN without sliding window with only 10 epoch due to quick saturation) and evaluating them after 50 epochs of training. Additionally, the learning curve of the NN model was closely examined. The performance of the NN model aligned with the data provided in the MoveNet evaluation section. Compared to its performance at 10 epochs, the NN model's prediction time increased by only 0.02 s after 50 epochs. The learning curve demonstrated that the NN model's training accuracy rapidly improved during the initial 0–10 epochs, after which it gradually increased from 99.86% to 99.9% between epochs 10 and 50, with minimal growth observed between epochs 32 and 50.

In contrast, after 50 epochs of training, the accuracy of the CNN model decreases to 99.68%, falling short of the accuracy achieved by the KNN model. Analysis of the learning curve reveals a pattern similar to that observed in the training accuracy curve of the NN model, where both models exhibit a rapid increase in accuracy during the initial 0–10 epochs, followed by a deceleration in the rate of improvement. However, the validation accuracy of the CNN model demonstrates greater instability and consistently remains lower than its training accuracy.

Subsequently, the three highest-performing machine learning models—Random Forest, NN, and CNN—were retrained and re-evaluated without the inclusion of sliding windows. During this process, the NN model was trained using the optimal 50 epochs, while the CNN model was trained using the most effective 10 epochs.

For the Random Forest model, the exclusion of sliding windows resulted in a reduction of 0.2 s in prediction time. However, its accuracy decreased to 99.86%. After the exclusion of sliding windows, the accuracy and other performance metrics of the NN model experienced a slight decline, decreasing to 99.82%. Nevertheless, the validation accuracy depicted in the learning curve exhibited greater stability. Furthermore, in comparison to the version that utilised sliding windows, the validation accuracy was relatively closer to the training accuracy.

Following the exclusion of sliding windows, the accuracy and other performance metrics of the CNN model declined to 99.53%. However, its learning curve also showed improvement, with the validation accuracy demonstrating a generally stable upward trend. The performance ranking of these three models underwent some changes following the exclusion of sliding windows: Random Forest remained the top performer, followed by NN, with CNN ranking last.

In the final ROS environment testing, two MoveNet models—MoveNet Lightning f16 and MoveNet Thunder—were utilised, alongside three machine learning models: Neural Network, CNN, and Random Forest. Each of these machine learning models was tested in two versions: one using a sliding window and one without. Each model was tested in combination with both MoveNet models, resulting in a total of 12 tested models. The testing dataset comprised data from four authenticated users and one unauthenticated user, with the amount of data across users being roughly equal.

The Random Forest models were the first to be eliminated. Although all Random Forest models correctly identified the unauthenticated user in 19% – 22% of the datasets, they only succeeded in recognising one of the authenticated users. Among the remaining eight models, the fastest prediction time was achieved by the CNN model using the MoveNet Thunder dataset, completing the prediction for all data in just 0.71 s. The next fastest was the CNN model using the MoveNet Thunder dataset with an added sliding window, which took 0.74 s. This was followed by two CNN models—one using Thunder and the other using Lightning f16 without applying a sliding window—both of which completed their predictions in under 1 s. In terms of accuracy, the best-performing model was the Neural Network (NN) without a sliding window using the Thunder dataset. This model correctly classified 19.59% of unauthenticated users and identified a higher percentage of

Table 2
MoveNet models performance with Neural network (NN).

Model (MoveNet)	Prediction Time (Second)	Accuracy	Precision	Recall	F1 Score
Lightning	1.060	99.88%	99.88%	99.88%	99.88%
Lightning int8	1.020	99.64%	99.64%	99.64%	99.64%
Lightning f16	0.960	99.90%	99.90%	99.90%	99.90%
Thunder	0.980	99.90%	99.90%	99.90%	99.90%
Thunder int8	0.980	99.80%	99.80%	99.80%	99.80%
Thunder f16	0.990	99.89%	99.89%	99.89%	99.89%

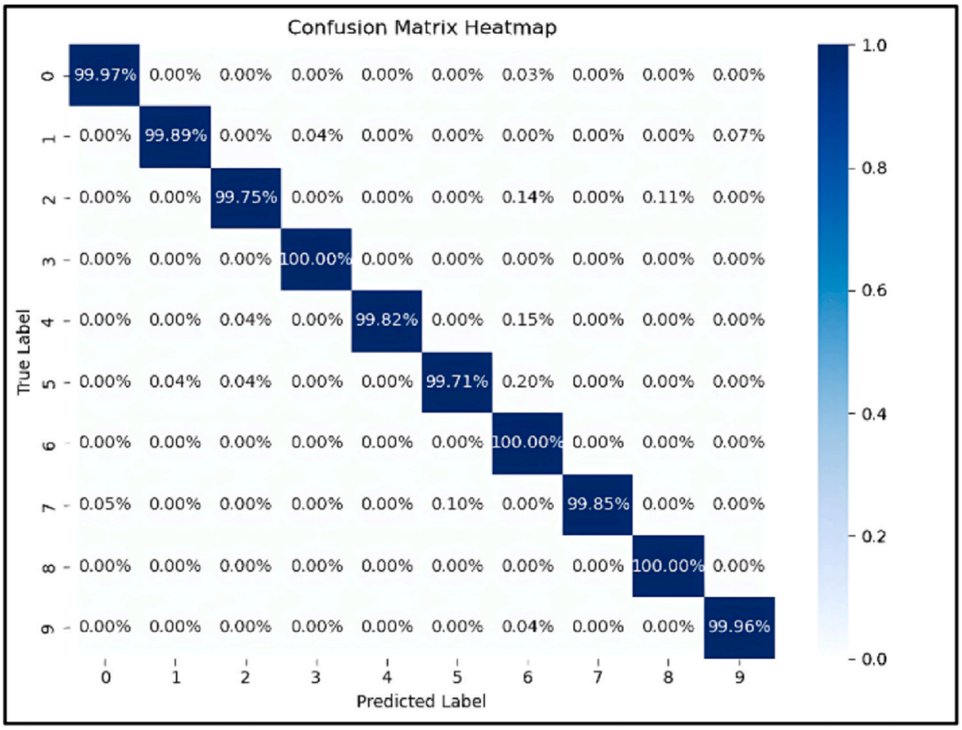

Fig. 12. MoveNet Lightning F16 model confusion matrix heatmap using NN.

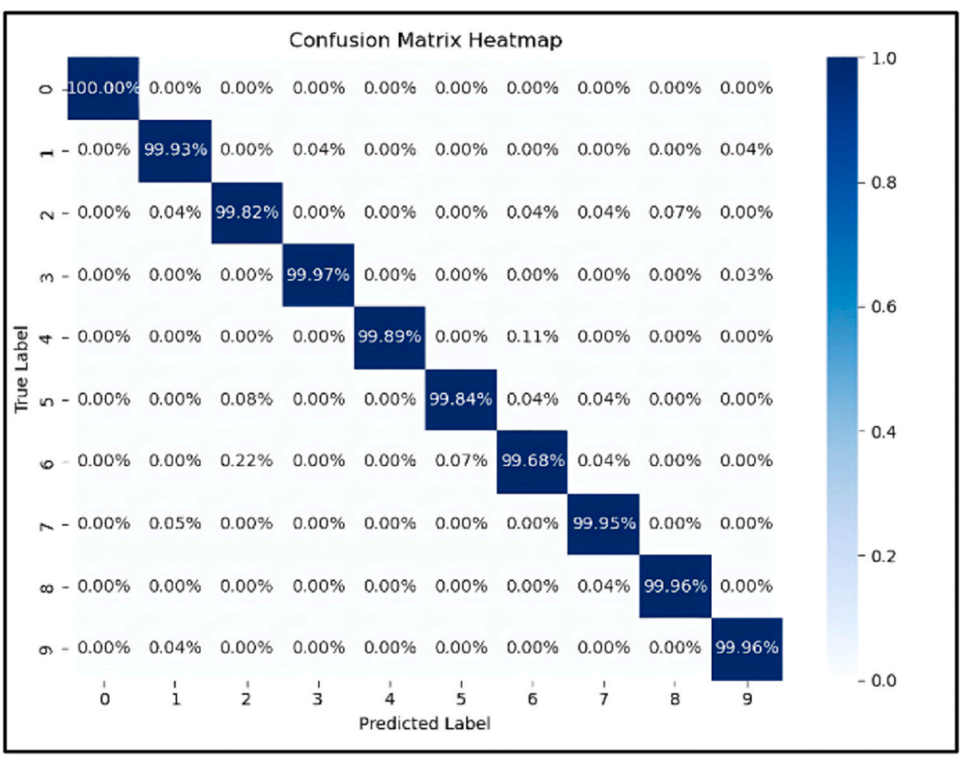

Fig. 13. MoveNet Thunder model confusion matrix heatmap using NN.

the four authenticated users, ranging between 14% and 19%. The next best was the CNN model without a sliding window using the Lightning f16 dataset. Although it identified 21.69% of unauthenticated users, it also recognised five authenticated users, with four of these users having relatively high identification percentages, ranging from 14% to 22%. The CNN model using Thunder, despite classifying 42.69% of users as unauthenticated, identified four authenticated users with significantly higher proportions, ranging from 15% to 19%. Other models that successfully identified four users with similarly high proportions were the CNN model without a sliding window using Thunder and the two CNN models using Lightning f16.

4. Discussion

In this study, the training and testing data are collected from young and healthy group of people and not from the elderly section of the

Table 3
Machine learning models performance with Thunder model.

Model	Prediction Time (Second)	Accuracy	Precision	Recall	F1 Score
SVM	00.790	21.73%	16.29%	21.73%	13.23%
Random Forest	00.800	99.93%	99.93%	99.93%	99.93%
KNN	15.380	99.74%	99.74%	99.74%	99.74%
Naïve Bayes	01.420	77.95%	79.83%	77.95%	77.24%
NN	00.950	99.86%	99.86%	99.86%	99.86%
CNN	25.000	99.89%	99.89%	99.89%	99.89%
RNN	30.030	93.18%	93.25%	93.18%	93.16%
LSTM	67.870	89.60%	89.60%	89.67%	89.41%

society. The aim is to study if our body movement data can be used to verify user's identity. So, it does not matter if the movement data was generated from young adults or elderly people for the study. The study has proven that movement data of people are unique, and they can be used as a method for user verification. This study paved a way for non-intrusive way of data collection for user verification. In the field of biometric verification process, most of the time, the user had to directly provide the biometric data to the reader (e.g. fingerprint) or follow guidelines to scan biometric data (e.g. scan the retina or face in front of the camera) or transmit the biometric data (e.g. heart rate variability data or gait data through smart wearable) etc. However, in this study, the person must simply walk and the camera in front captures the movement data and analyse the 17 data points of the body along with the distances of each data points to verify the user identity. It means that, you can freely walk in a normal pace without any restrictions. The 17 data points of the MoveNet are flexible in terms of collection and it doesn't matter if the user walk towards the camera by facing the camera or walk away from the camera by facing the face away from the camera e.g. the eyes, nose, ears and other data points of the body are estimated at a right positions of the body by the MoveNet irrespective of the person facing towards camera or away from the camera as shown in Fig. 22. Such a verification system is robust, flexible and easy to adopt. It doesn't require remembering (e.g. like password and passcode) or carrying (e.g. swipe card) or following any strict guidelines during biometric data provision process (e.g. scanning fingerprint or looking into the camera for face detection and recognition) etc. Thus, such a verification system will of great use for the elderly people because it doesn't involve remembering passcodes or carrying IDs or cards for verification, it is non-invasive, doesn't need wearables or sensors, or active user input. Therefore, such a verification system will be of great value not only for user verification but for continuous monitoring of the elderly users e.g. in care homes, hospitals, homes, etc.

In this study, data is collected with a fixed camera as participants walk toward it. However, in real-world settings, varying camera angles can cause inconsistencies in capturing the 17 key body movement data points. These variations may reduce data accuracy and reliability, potentially affecting the effectiveness of identity verification. Unstable or shifting viewpoints could lead to errors in recognising movement patterns essential for verifying users. Therefore, maintaining a consistent camera angle is crucial for accurate data collection and model performance. This highlights a key challenge in translating controlled experimental setups to practical, real-life applications in user verification systems. The following section of the discussion focuses on the performance of the models used in the movement analysis for user verification.

The performance evaluation of the different MoveNet models in this study generally aligns with the official data. The MoveNet Thunder series models demonstrated higher overall accuracy compared to the MoveNet Lightning series. However, the Lightning series exhibited longer prediction times than the Thunder series. This discrepancy may be attributed to the lower precision of the datasets generated by the Lightning series, which could have necessitated additional prediction time for the models. MoveNet is primarily designed for pose estimation and classification, without specific adaptation for identity verification. This may explain why the performance of the Lightning series models fell below the expected benchmarks. The MoveNet Thunder model's performance aligns with the conclusions drawn from the literature. In most literatures, the researchers addressed that it is consistently recognised as the best-performing model, particularly well-suited for integration with machine learning algorithms. This study confirms the MoveNet Thunder model's superiority in generating high-precision datasets. However, contrary to some previous research, the MoveNet Lightning f16 model demonstrated comparable performance with the added advantage of a shorter prediction time compared to the Thunder series. Although this finding deviates from conventional expectations, it may be due to the Lightning f16 model's optimisation for computational efficiency in lower-resolution environments. While the MoveNet Thunder series typically excels in accuracy, the other two Thunder models underperformed relative to the Lightning f16 when processing lower-quality video data. This inconsistency might stem from the Thunder models being optimised for high-quality video inputs, whereas the Lightning series models are better equipped to handle lower-resolution or noisier inputs. The Lightning f16 model, in particular, proved to be well-suited to the scale and accuracy requirements of this study's application scenario. Consequently, both the lightweight MoveNet Thunder from the Thunder series and the highly accurate MoveNet Lightning f16 from the Lightning series demonstrated strong performance and efficiency.

In terms of prediction time, the performance of the machine learning models generally corresponded with their algorithmic complexity. Simpler models like SVM, Naive Bayes, and NN had the shortest prediction times, while the most complex model, LSTM, took the

Table 4
Machine learning models performance in testing stage.

Model	MoveNet	Time (Second)	Unauthenticated in %										
NN	Lightning f16	3.657	8.49	10.77	2.22	3.00	15.13	1.36	19.53	4.11	17.72	20.64	5.53
NN - NoSW	Lightning f16	1.415	21.69	11.53	1.09	6.65	22.71	0.41	14.67	6.69	15.24	15.10	5.91
CNN	Lightning f16	1.697	0.49	22.92	4.15	6.04	6.95	2.61	18.27	2.90	15.05	9.09	12.02
CNN NoSW	Lightning f16	0.887	51.50	24.93	0.03	0.03	5.48	1.46	20.76	4.97	15.82	9.60	10.98
NN	Thunder	0.815	6.70	6.47	2.57	2.45	13.83	10.32	25.26	3.91	14.61	8.82	11.76
NN - NoSW	Thunder	0.891	19.59	15.43	0.78	8.31	14.38	7.96	19.12	3.43	17.83	5.79	6.98
CNN	Thunder	0.736	42.69	7.24	2.16	4.91	15.19	4.17	19.14	4.97	15.83	19.65	8.74
CNN NoSW	Thunder	0.709	51.87	11.39	2.15	4.55	7.49	14.59	23.12	3.32	17.38	7.24	8.76

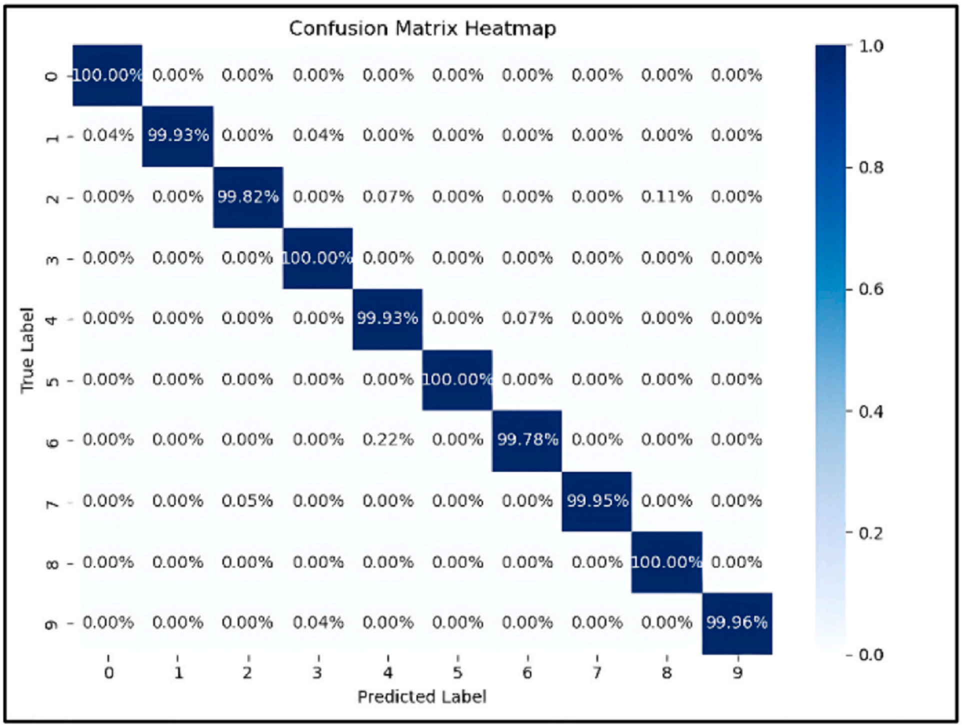

Fig. 14. Random forest model confusion matrix heatmap with thunder.

longest. Interestingly, the prediction time for the simpler KNN model was relatively long, whereas the more complex Random Forest model had the second-shortest prediction time. This difference is likely due to the inherent workings of these models. KNN requires minimal computation during the training phase, as it primarily involves storing the training data, but during the evaluation phase, it needs to compute the distance between all test samples and the training data, which becomes time-consuming with high-dimensional datasets [22]. In contrast, Random Forest builds its decision tree structures during training, allowing it to make predictions quickly by parallel processing the decision trees during evaluation [25]. Similarly, the simpler SVM and Naive Bayes models also exhibited lower performance in terms of accuracy, while the more complex models generally achieved higher scores. However, KNN, despite its simplicity, demonstrated relatively good accuracy. This aligns with findings in the literature review, where KNN was frequently used in gait recognition research. KNN's method of classifying by calculating the distance between prediction data and training data might be particularly well-suited to gait data, where each

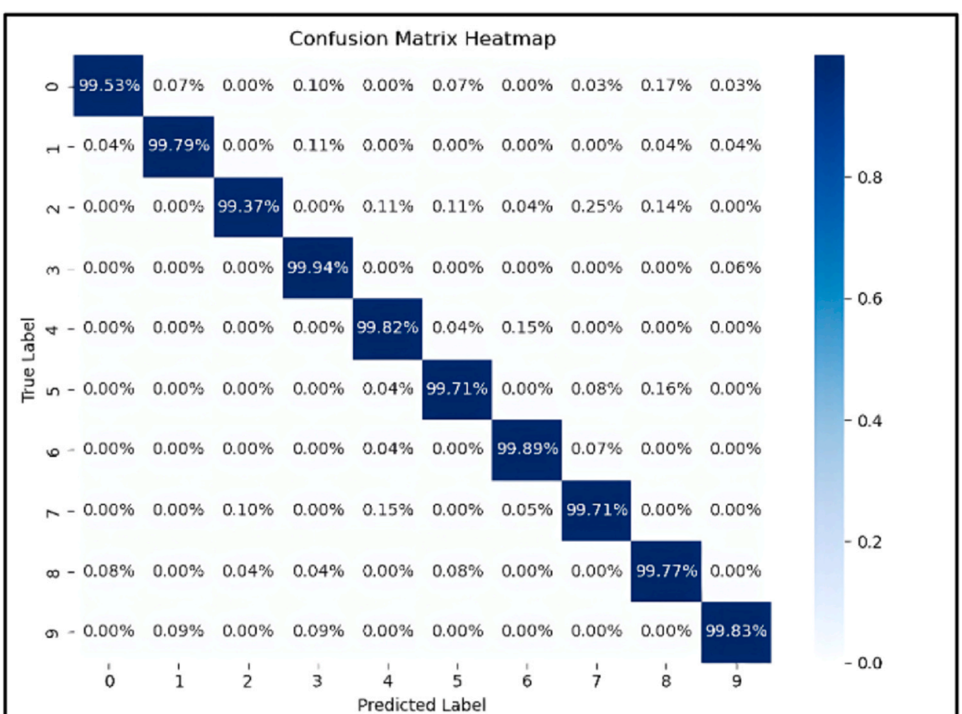

Fig. 15. KNN model confusion matrix heatmap with thunder.

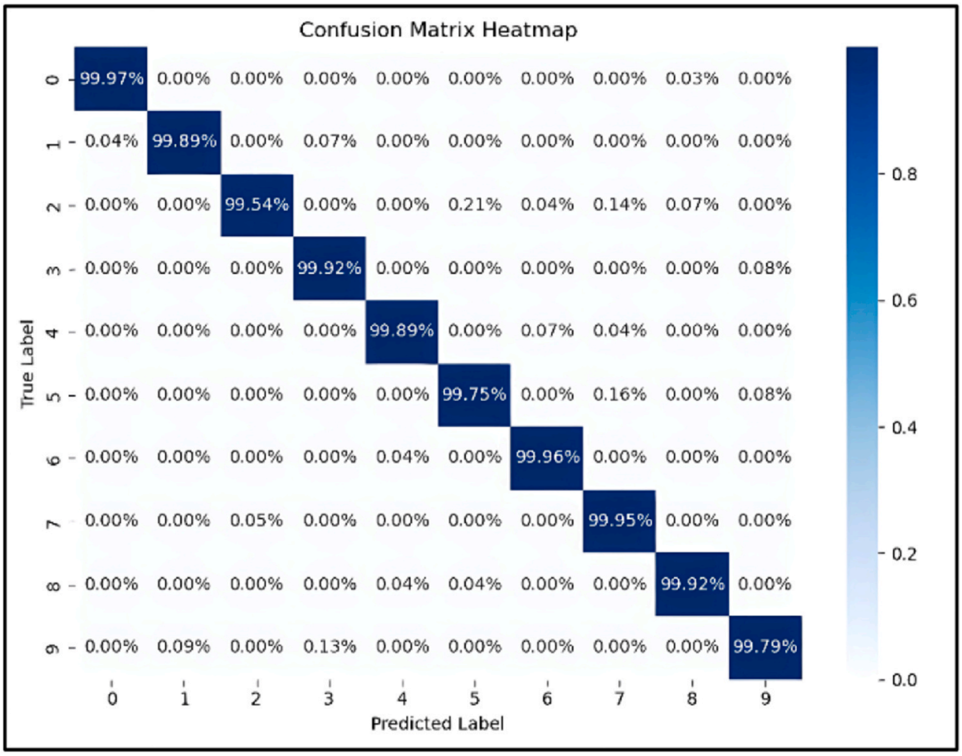

Fig. 16. NN model confusion matrix heatmap with thunder.

individual's gait and body distance data are unique, making the differences in distances more pronounced and leading to better performance compared to some complex models.

Interestingly, Random Forest, which was not commonly used in the literature reviewed, outperformed other models in terms of accuracy during both training and evaluation phases. This may be due to the model's capacity to process high-dimensional data effectively and its resilience in dealing with noise, which could account for its high accuracy [25]. However, since Random Forest is relatively a "black box" algorithm, further research would be necessary to fully understand why it achieved the highest accuracy. Following Random Forest, the CNN and NN models performed well, which corroborates their frequent use in related studies, highlighting their suitability for applications like gait recognition. In this study, the gait and body distance data were input as a series of static frames, with the dataset emphasising spatial features, making NN and CNN particularly applicable. On the other hand, more

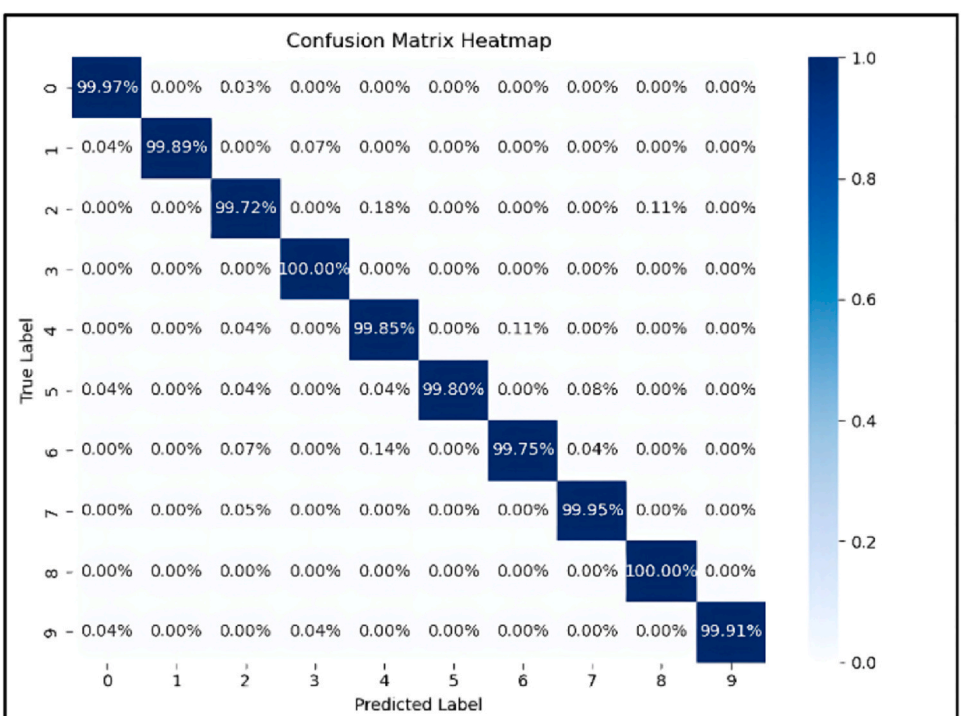

Fig. 17. CNN model confusion matrix heatmap with thunder.

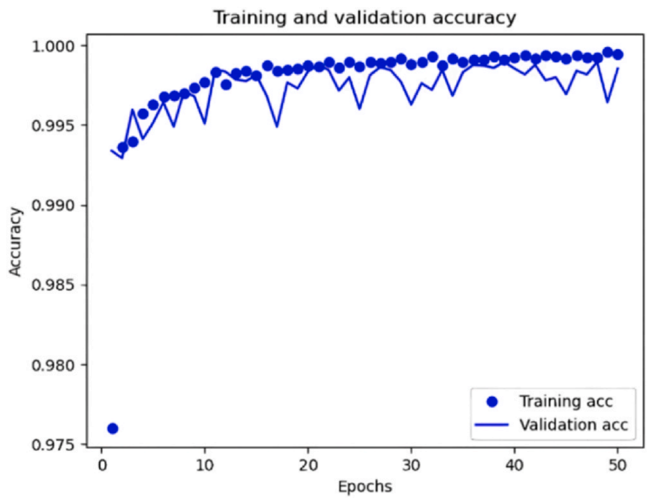

Fig. 18. NN model 50 epochs learning curve with thunder.

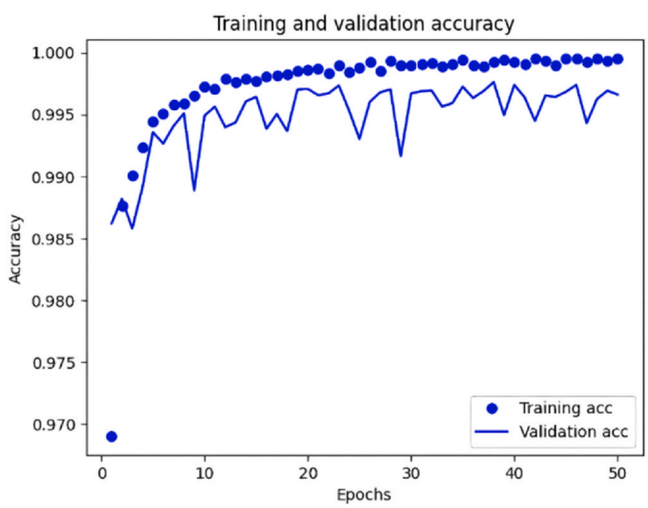

Fig. 19. CNN model 50 epochs learning curve with thunder.

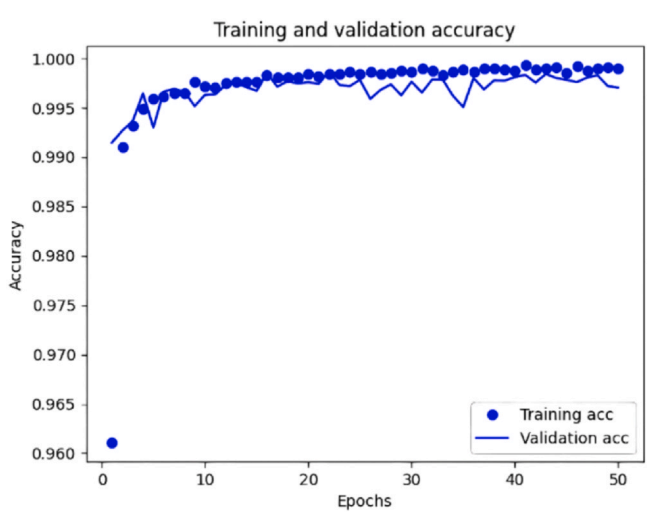

Fig. 20. NN model 50 epochs without sliding window learning curve with thunder.

complex models like RNN and LSTM exhibited lower accuracy, possibly because these models are designed to capture temporal features, making them more effective for datasets with dynamic temporal characteristics. If the dataset had emphasised the temporal variations in gait

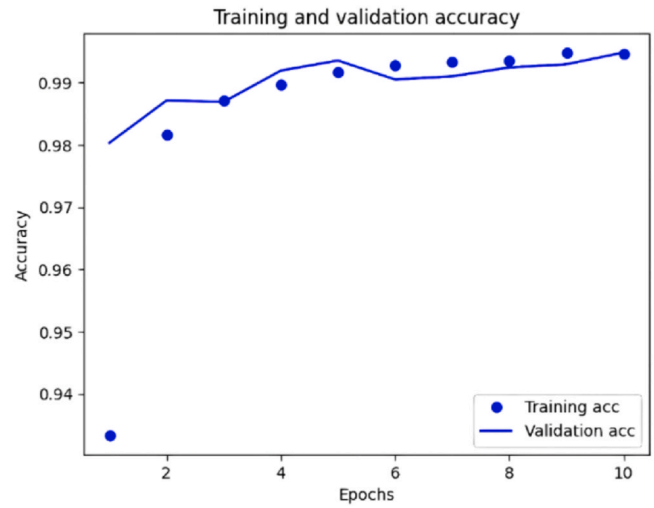

Fig. 21. CNN model 10 epochs without sliding window learning curve with thunder.

and body distance data, these models might have performed better. After increasing the number of training epochs, the accuracy of the NN model improved, while the accuracy of the CNN model decreased. This could be due to the CNN's more complex structure, which might have overfitted to the relatively simple dataset, leading to a decline in accuracy after too many training cycles [24]. In contrast, the simpler NN model benefited from additional training, becoming more adept at identifying features and improving its accuracy. The learning curves for both models revealed significant fluctuations in validation accuracy, with less alignment with the training accuracy, suggesting potential overfitting. The dataset in this study had relatively few labels but many features, and the application of a sliding window may have exacerbated the discrepancy between the number of labels and features [13]. When the sliding window was removed, the accuracy of all three models decreased, likely due to the significant reduction in sample size. However, the validation accuracy curves in the learning curves of the NN and CNN models became noticeably smoother and more closely aligned with the training accuracy curves. This suggests that removing the sliding window might have effectively mitigated the issue of data overfitting.

During the validation phase, an unexpected issue arose: while the Random Forest model was reasonably accurate in identifying unauthorised users, it mistakenly classified all authorised users as the same individual. This suggests that the Random Forest model may not be well-suited for gait recognition, although a more detailed analysis is required to confirm the underlying reasons for this outcome. The best performance was achieved by the NN model without a sliding window, using the Thunder dataset, which is consistent with earlier analyses. The Thunder-based model outperformed the Lightning f16 version in terms of accuracy and prediction speed. This might be attributed to Thunder providing a more precise dataset, resulting in faster prediction times. On average, models using the Thunder dataset predicted faster than those using Lightning f16, solidifying MoveNet Thunder as the most suitable MoveNet model for this study. Furthermore, the NN model, which avoided overfitting and had a simpler structure, proved more adept at gait and body distance recognition, offering higher prediction speeds and the highest accuracy rates. The NN model without a sliding window, using Lightning f16, yielded the second most accurate results, reinforcing NN as the most suitable machine learning algorithm for this study. However, the CNN model with a sliding window generally outperformed the NN model with a sliding window, and the CNN model without a sliding window performed better than its sliding window counterpart. This indirectly highlights CNN's potential for application in this study. With a larger dataset and more optimised CNN

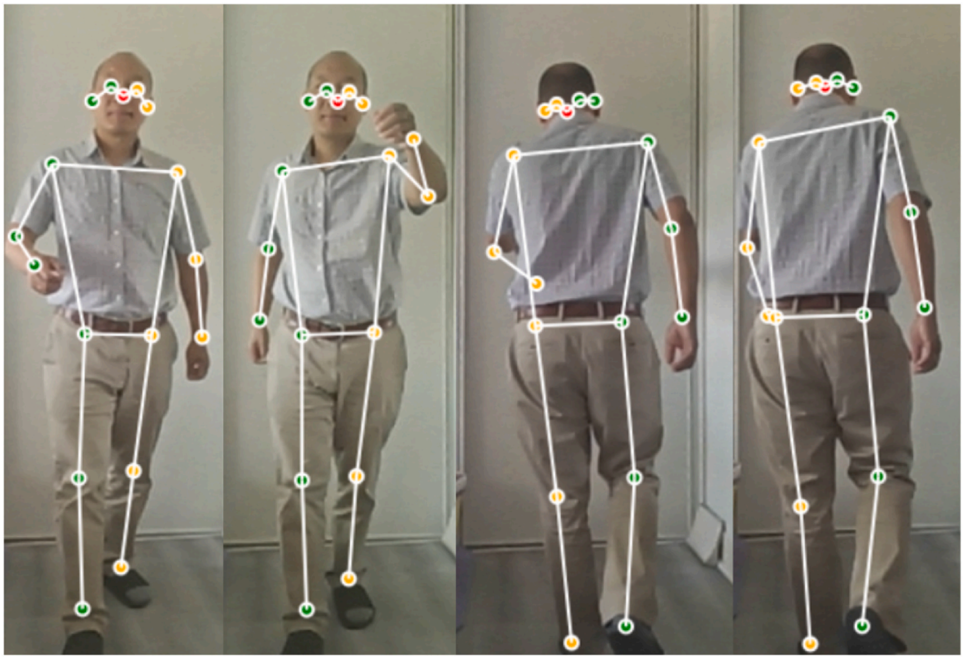

Fig. 22. MoveNet's 17 Data Points of the Body Parts Irrespective of Face Position.

model construction, CNN could potentially demonstrate even higher performance. Overall, the accuracy demonstrated by the Thunder-based NN model without a sliding window in identity recognition is sufficient to confirm that MoveNet can indeed be utilised for gait and distance recognition as a viable continuous authentication solution for care robots.

5. Future direction and conclusion

The study highlights the potential of using MoveNet for gait recognition in continuous authentication for care robots and outlines several future research directions:

1. Multi-Factor Authentication: Combining gait recognition with other biometric technologies (e.g., facial and voice recognition) could enhance accuracy and security in user identification for care robots.
2. Exploring Other Pose Estimation Models: Testing alternatives like OpenPose, PoseNet, and BlazePose may lead to more effective identity verification solutions. This could also drive the development of healthcare-specific pose estimation models.
3. Enhancing Auxiliary Functions: Expanding the system to recognize attributes like gender or age could improve user categorization, ensure privacy, and enable personalized care plans.
4. Broader Applications: Leveraging pose estimation models for tasks beyond user identification—such as posture recognition, emergency monitoring, and real-time assistance—could improve daily care and safety. Integration with existing algorithms could reduce costs and enhance robot usability.

This study explores the feasibility of using MoveNet, a pose estimation tool, for continuous user authentication in care robots through gait and body points distances recognition. Key achievements include evaluating different MoveNet models, collecting gait data, applying machine learning algorithms, and testing in a ROS (Robot Operating System) environment. Key Findings:

a) MoveNet Models: The Thunder and Lightning f16 models performed best in generating datasets for user identification.
b) Machine Learning: Neural networks, particularly CNNs, showed potential in recognizing gait and body distance features for authentication.
c) Impact: This system can enhance care robots' security and usability, especially for elderly and vulnerable users.

The study demonstrates MoveNet's capability for non-intrusive, continuous user verification. Addresses a gap in research on using pose estimation for gait recognition. Highlights the integration potential of MoveNet into broader care robot functionalities. However, it has limitations on the testing Environment (Conducted under controlled conditions) and algorithm Scope (Limited real-time video stream and advanced ML development). So, in the future a further work to include and incorporate diverse participants and realistic environments, expand to multi-factor authentication, experiment with other pose estimation models and further integrate authentication with care robots' overall functionality. Thus, the study underscores the promise of MoveNet for secure, seamless interaction in care robots, contributing to advancements in robotics and other domains requiring continuous, non-intrusive user verification.

CRediT authorship contribution statement

Raymond Mawanda: Data curation, Writing – review & editing. **Alessandro Di Nuovo:** Methodology, Resources, Supervision, Writing – review & editing. **Ruomeng Zhang:** Writing – original draft, Methodology, Investigation, Conceptualization, Data curation, Formal analysis, Software, Validation, Visualization. **Solan Keishing:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **Jims Marchang:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization, Data curation, Formal analysis. **Ning Wang:** Writing – review & editing.

Data availability

The experimental data and its findings are reported in this paper. If more detailed raw data are required then it will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the Warwick University and the Sheffield Hallam University for providing the hardware resources needed for this project.

References

[1] S. Almohamade, J. Clark, J. Law, Continuous user authentication for human-robot collaboration, Proc. 16th Int. Conf. Availab. Reliab. Secur. (2021) 1–9, https://doi.org/10.1145/3465481.3470025 Accessed 31 August 2024).
[2] F.H. Al-Naji, R. Zagrouba, A survey on continuous authentication methods in Internet of Things environment (Available from:), Comput. Commun. 163 (2020) 109–133, https://doi.org/10.1016/j.comcom.2020.09.006 Accessed 31 August 2024).
[3] M.K. Al-Qaderi, A.B. Rad, A multi-modal person recognition system for social robots (Available from:), Appl. Sci. 8 (3) (2018) 387, https://doi.org/10.3390/app8030387 Accessed 31 August 2024).
[4] N.L. Atqia, E. Suryani, Utilization of Body Posture Feature Extraction with Movenet for Silat Movement Classification Using Artificial Neural Network (ANN) (Available from:), 6th Int. Conf. Inf. Commun. Technol. 2023 (2023) 372–377, https://doi.org/10.1109/ICOIACT59844.2023.10455800 Accessed 31 August 2024).
[5] S. Ayeswarya, J. Norman, A survey on different continuous authentication systems, Int. J. Biom. 11 (1) (2019) 67–99.
[6] A.F. Baig, S. Eskeland, Security, privacy, and usability in continuous authentication: A survey (Available from:), Sensors 21 (17) (2021) 5967, https://doi.org/10.3390/s21175967.
[7] N. Bellotto, H. Hu, A bank of unscented kalman filters for multimodal human perception with mobile service robots (Available from:), Int. J. Soc. Robot. 2 (2010) 121–136, https://doi.org/10.1007/s12369-010-0047-x Accessed 31 August 2024).
[8] Bernier, C. (2013) Collaborative Robot Series: PR2 from Willow Garage. [online] Available from: https://blog.robotiq.com/bid/65419/Collaborative-Robot-Series-PR2-from-Willow-Garage (Accessed 31 August 2024).
[9] Centre for Ageing Better . (2023) *Our Ageing Population | The State of Ageing 2023–24*. Available from: https://ageing-better.org.uk/our-ageing-population-state-ageing-2023-4#:~:text=2.01%20Number%20of%20people%20(in,Census%2C%20England%2C%201981%20to%202021&text=This%20chart%20shows%20that%3A,6.8%20million%20people%20(47%25 (Accessed 31 August 2024).
[10] A.A. Chaaraoui, J.R. Padilla-López, F. Flórez-Revuelta, Abnormal gait detection with RGB-D devices using joint motion history features (Available from:), 2015 11th IEEE Int. Conf. Workshops Autom. face gesture Recognit. 7 (2015) 1–6, https://doi.org/10.1109/FG.2015.7284881 (Accessed 31 August 2024).
[11] W. Cheung, S. Vhaduri, Continuous authentication of wearable device users from heart rate, gait, and breathing data, 2020 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (2020) 587–592. Available from: https://doi.org/10.1109/BioRob49111.2020.9224356 (Accessed 31 August 2024).
[12] Citrix. (2024) *What is continuous authentication?* [online] Available from: https://www.citrix.com/glossary/what-is-continuous-authentication.html#:~:text=Continuous%20authentication%20works%20by%20assessing,posture%2C%20and%20other%20behavioral%20data (Accessed 31 August 2024).
[13] Data Overload. (2022) *Sliding Window Technique — reduce the complexity of your algorithm.* [online] Available from: 〈https://medium.com/〉@data-overload/sliding-window-technique-reduce-the-complexity-of-your-algorithm-5badb2cf432f〉 (Accessed 31 August 2024).
[14] J. Deepthi, N.P. Dharani, S. Aswini, D.G. Priya, K.M. Harsha, Y.S.S.S. Bhargav, Care Guard: Holistic Health Monitoring and Autonomous Care System 2024 3rd International Conference on Applied Artificial Intelligence and Computing, 2024, pp. 1325–1331. Available from: https://doi.org/10.1109/ICAAIC60222.2024.10574934 (Accessed 31 August 2024).
[15] Department for Business, Energy & Industrial Strategy, Care robots could revolutionise UK care system and provide staff extra support, UK Research and Innovation and The Rt Hon Chris Skidmore, 2019 Available from: https://www.gov.uk/government/news/care-robots-could-revolutionise-uk-care-system-and-provide-staff-extra-support (Accessed 31 August 2024).
[16] European Commission. (2018) Periodic Reporting for period 3 - RAMCIP (Robotic Assistant for MCI patients at home*)*. [online] Available from: 〈https://cordis.europa.eu/project/id/643433/reporting〉 (Accessed 31 August 2024).
[17] J. Fasola, M.J. Mataric, Using socially assistive human–robot interaction to motivate physical exercise for older adults (Available from:), Proc. IEEE 100 (8) (2012) 2512–2526, https://doi.org/10.1109/JPROC.2012.2200539 (Accessed 31 August 2024.
[18] Z.B. Feliandra, S. Khadijah, M.F. Rachmadi, D. Chahyati, Classification of stroke and Non-Stroke patients from human body movements using smartphone videos and Deep Neural Networks, Int. Conf. Adv. Comput. Sci. Inf. Syst. (2022) 187–192. Available from: https://doi.org/10.1109/ICACSIS56558.2022.9923501 (Accessed 31 August 2024).
[19] L. Fiorini, K. Tabeau, G. D'Onofrio, L. Coviello, M. De Mul, D. Sancarlo, I. Fabbricotti, F. Cavallo, Co-creation of an assistive robot for independent living: lessons learned on robot design, Int. J. Interact. Des. Manuf. 14 (2019) 491–502. Available from: https://doi.org/10.1007/s12008-019-00641-z (Accessed 31 August 2024).
[20] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, M. Vincze, Hobbit, a care robot supporting independent living at home: First prototype and lessons learned, Robot. Auton. Syst. 75 (2016) 60–78. Available from: https://doi.org/10.1016/j.robot.2014.09.029 (Accessed 31 August 2024).
[21] S. Gao, J. Yun, Y. Zhao, L. Liu, Gait-D: skeleton-based gait feature decomposition for gait recognition, IET Comput. Vis. 16 (2) (2022) 111–125. Available from: https://doi.org/10.1049/cvi2.12070 (Accessed 31 August 2024).
[22] GeeksforGeeks. (2024) *K-Nearest Neighbor(KNN) Algorithm.* [online] Available from: 〈https://www.geeksforgeeks.org/k-nearest-neighbours/〉 (Accessed 31 August 2024).
[23] Hanson Robotics. (2024) *Sophia*. [online] Available from: https://www.hansonrobotics.com/sophia/ (Accessed 31 August 2024).
[24] IBM. (2024) *What is overfitting?* [online] Available from: https://www.ibm.com/topics/overfitting#:~:text=IBM-,What%20is%20overfitting%3F,of%20the%20machine%20learning%20model (Accessed 31 August 2024).
[25] IBM. (2024) *What is random forest?* [online] Available from: https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,Decision%20trees (Accessed 31 August 2024).
[26] Z. Kozhamkulova, B. Kirgizbayeva, G. Sembina, U. Smailova, M. Suleimenova, A. Keneskanova, Z. Baizakova, MoveNET Enabled Neural Network for Fast Detection of Physical Bullying in Educational Institutions, Int. J. Adv. Comput. Sci. Appl. 14 (5) (2023) Available from: https://doi.org/10.14569/IJACSA.2023.0140578 (Accessed 31 August 2024).
[27] M. Krakovski, S. Kumar, S. Givati, M. Bardea, O. Zafrani, G. Nimrod, S. Bar-Haim, Y. Edan, Gymmy": designing and testing a robot for physical and cognitive training of older adults, Appl. Sci. 11 (14) (2021) 6431. Available from: https://doi.org/10.3390/app11146431 (Accessed 31 August 2024).
[28] M. Kudo, Robot-assisted healthcare support for an aging society, Annu. SRII Glob. Conf. 2012 (2012) 258–266. Available from: https://doi.org/10.1109/SRII.2012.36 (Accessed 31 August 2024).
[29] LeViet, K., Chen, Y.H. (2021) *Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite.* [online] Available from: https://blog.tensorflow.org/2021/08/pose-estimation-and-classification-on-edge-devices-with-MoveNet-and-TensorFlow-Lite.html (Accessed 31 August 2024).
[30] R. Liao, S. Yu, W. An, Y. Huang, A model-based gait recognition method with body pose and human prior knowledge, Pattern Recognit. 98 (2020) 107069. Available from: https://doi.org/10.1016/j.patcog.2019.107069 (Accessed 31 August 2024).
[31] J. Marchang, A. Di Nuovo, Assistive multimodal robotic system (AMRSys): security and privacy issues, challenges, and possible solutions, Appl. Sci. 12 (4) (2022) 2174. Available from: https://doi.org/10.3390/app12042174 (Accessed 31 August 2024).
[32] E. Martinez-Martin, F. Escalona, M. Cazorla, Socially assistive robots for older adults and people with autism: An overview, Electronics 9 (2) (2020) 367. Available from: https://doi.org/10.3390/electronics9020367 (Accessed 31 August 2024).
[33] R. Mawanda, S. Keishing, J. Wang, J. Marchang, Secure by Design Smart Authentication for Care Robots to Support the Elderly, IEEE Int. Conf. Ind. Technol. 2024 (2024) 1–6. Available from: https://doi.org/10.1109/ICIT58233.2024.10541033 (Accessed 31 August 2024).
[34] T.N. Nguyen, H.H. Huynh, J. Meunier, Skeleton-based abnormal gait detection, Sensors 16 (11) (2016) 1792. Available from: https://doi.org/10.3390/s16111792 (Accessed 31 August 2024).
[35] Nurseline. (2023) *What Is Personal Care and Why Is It Important?* [online] Available from: https://nurselinecs.co.uk/personal-care/what-is-personal-care-and-why-is-it-important/ (Accessed 31 August 2024).
[36] A. Parle, R. Shinde, R. Chougule, S. Agrawal, YogaWise: Enhancing Yoga with Intelligent Real Time Tracking using TensorFlow MoveNet, Int. Conf. Cogn. Robot. Intell. Syst. 2024 (2024) 498–505. Available from: https://doi.org/10.1109/ICC-ROBINS60238.2024.10534004 (Accessed 31 August 2024).
[37] A. Rangarajan, H. Ramachandran, A preliminary analysis of AI based smartphone application for diagnosis of COVID-19 using chest X-ray images, Expert Syst. Appl. 183 (2021) 115401Available from: https://doi.org/10.1016/j.eswa.2021.115401 (Accessed 31 August 2024).
[38] S.B. Rikan, A.S. Azar, A. Ghafari, J.B. Mohasefi, H. Pirnejad, COVID-19 diagnosis from routine blood tests using artificial intelligence techniques, Biomed. Signal Process. Control 72 (2022) 103263Available from: https://doi.org/10.1016/j.bspc.2021.103263 (Accessed 31 August 2024).
[39] B. Sawik, S. Tobis, E. Baum, A. Suwalska, S. Kropińska, K. Stachnik, E. Pérez-Bernabeu, M. Cildoz, A. Agustin, K. Wieczorowska-Tobis, Robots for Elderly Care: Review, Multi-Criteria Optimization Model and Qualitative Case Study, Healthcare 11 (9) (2023) 1286. Available from: https://doi.org/10.3390/healthcare11091286 (Accessed 31 August 2024).
[40] A. Sepas-Moghaddam, A. Etemad, Deep gait recognition: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (2022) 264–284. Available from: https://doi.org/10.1109/TPAMI.2022.3151865 (Accessed 31 August 2024).
[41] TensorFlow. (2024) *MoveNet: Ultra fast and accurate pose detection model.* [online] Available from: https://www.tensorflow.org/hub/tutorials/movenet (Accessed 31 August 2024).
[42] T. Tomy, D. Narang, P. Ahuja, R. Rani, A. Sharma, A. Dev, Yoga Pose Estimation Using MoveNet Preprocessor and Deep Learning Technique, Int. Conf. Cogn. Comput. Cyber Phys. Syst. (2023) 509–552. Available from: https://doi.org/10.1007/978-981-97-2550-2_37 (Accessed 31 August 2024).

[43] University of Auckland. (2024) *HEALTHBOTS*. [online] Available from: https://cares.blogs.auckland.ac.nz/research/healthcare-assistive-technologies/healthbots/ (Accessed 31 August 2024).
[44] M. Vasanthi, K. Seetharaman, Facial image recognition for biometric authentication systems using a combination of geometrical feature points and low-level visual features, J. King Saud. Univ. Comput. Inf. Sci. 34 (7) (2022) 4109–4121.
[45] Votel, R., Li, N. (2021) Next-Generation Pose Detection with MoveNet and TensorFlow.js. [online] Available from: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html (Accessed 31 August 2024).
[46] Wan, C., Wang, L. and Phoha, V.V. eds. (2018) A survey on gait recognition. *ACM Computing Surveys (*CSUR), 51 (5): 1–35. Available from: https://doi.org/10.1145/3230633 (Accessed 31 August 2024).
[47] E.P. Washabaugh, T.A. Shanmugam, R. Ranganathan, C. Krishnan, Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics, Gait Posture 97 (2022) 188–195. Available from: https://doi.org/10.1016/j.gaitpost.2022.08.008 (Accessed 31 August 2024).
[48] I. Wieser, S. Toprak, A. Grenzing, T. Hinz, S. Auddy, E.C. Karaoğuz, A. Chandran, M. Remmels, A. El Shinawi, J. Josifovski, L.C. Vankadara, A robotic home assistant with memory aid functionality, KI 2016 Adv. Artif. Intell. 39 (2016) 102–115. Available from: https://doi.org/10.1007/978-3-319-46073-4_8 (Accessed 31 August 2024).
[49] C.Y. Yang, M.J. Lu, S.H. Tseng, L.C. Fu, September. A companion robot for daily care of elders based on homeostasis, 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan, 2017, pp. 1401–1406. Available from: https://doi.org/10.23919/SICE.2017.8105748 (Accessed 31 August 2024).