e-ISSN : 0976-5166
p-ISSN : 2231-3850

Multi-Modality Driven Sparse Inertial Feature Representation for Gait-Based Scalable Person Authentication System

Ambika K
Assistant Professor, BMS College of Engineering, VTU, Bull temple road, Basavanagudi, Bengaluru, Karnataka 560019, India
ambikak.tce@bmsce.ac.in

Dr.Radhika K R
Professor, BMS College of Engineering, VTU, Bull temple road, Basavanagudi, Bengaluru, Karnataka 560019, India
rkr.ise@bmsce.ac.in

Abstract : In the last few years, gait-based person-authentication systems have gained widespread attention, because of its touchless nature and non-replicable behavioral characteristics. However, the classical gait-based person authentication systems, especially designed with human's static body structure, and dynamic movement patterns are limited due to high-reliance onto local conditions and exhaustive computation. Moreover, scalability of such at-hand solutions often remain suspicious. On the contrary, in the last few years the high-pace rise in sensor technologies, smartphones, wearables etc. has broadened the horizon for automatic gait-feature acquisition or inertial measurements, like acceleration and gyroscope data. Such continuous gait-sequences can be applied for (feature)-learning and classification to serve person authentication task; however, gigantic continuous data processing with indefinite gait-feature localization within small gait-cycle imposes high equal error rate or false-positive. Nevertheless, it confines suitability towards resource constrained devices like Smartphone, Inertial Measuring Units(IMU), sensors etc. Considering such limitations, in this paper a first of its kind solution is proposed that addresses almost all known and quantifiable challenges towards sensor driven gait-based person authentication. More specifically, in this paper, stacked denoising auto-encoder driven sparse multi-modality feature representation model is developed for gait-based scalable person authentication. Being multi-modal, the proposed model exploits acceleration and gyroscope data, where the inputs are at first processed for fixed window-based segmentation, followed by stationary wavelet transformation driven stacked denoising auto-encoder (SWT-SDAE) for sparse data representation. The sparse reconstructed data, which can also be stated as a compressed continuous gait-stream is processed with a modified Long and Short-Term Memory (LSTM) model for feature extraction and learning. The MATLAB based simulation has revealed that the proposed person authentication model yields accuracy of 99.43%, precision 99.02%, recall 98.86%, F-Score 0.9893 and equal error rate of 0.57%, signifying robustness of the proposed model towards real-time scalable infrastructure.

***Keywords*: Person Authentication, Multi-Modality Gait-Authentication, Inertial Measurement Features, Sparse Feature Learning, Scalable Authentication System, LSTM-RNN.**

1.Introduction

The world has witnessed high-pace rise in advanced software computing, wireless communication, and low-cost miniatured hardware to meet up-surging demands. Internet technologies in sync with aforesaid innovations has broadened the horizon for real-time proactive decision systems. With exponentially upsurge in aforesaid technologies too has broadened the horizon for the applications serving socio-economic enterprises and industries in different manner. However, being an exceedingly dynamic computing environment ensuring process-control, process security and allied surveillance has become inevitable [1]. Whether it is system access or authentication, resource access control or personalized access control purpose, guaranteeing authorized access to the relevant user is of paramount significance. Moreover, there are a large number of applications including decentralized surveillance and tracking, human-machine-interface (HMI) etc., [1] where person identification and authentication turn out to be inevitable. Person authentication has emerged as a vital prerequisite serving numerous HMIs or allied applications to serve security, privacy and convenience [2]. Person authentication represent the set of functional components with the ability to process input example, create corresponding identification tag and validate user's authenticity when demanded in run time by matching test sample with pre-stored tags in the database [3]. It makes person authentication solution capable to serve smart home [5], intrusion detection [6], access control [7], etc. Though, authentication systems root from the classical use of cryptography; however, computational complexities, overheads, delay and vulnerability and intrusion related breaches confine its scalability [4]. To alleviate such limitations towards person authentication, biometrics-based methods have gained widespread attention. The minimal replication probability makes it robust towards person authentication. In sync with its robustness and scalability, numerous biometric modalities including fingerprint [7], Iris [8], voice [9], vital signs [10][11], and gait [12] have been employed for person authentication. Despite significant efforts majority of the existing biometrics are confined due to high computational overheads, complexities, dependency on the local conditions (i.e., illumination, viewing angle, data suitability, sensor conditions etc.). On the other hand, COVID-19 pandemic has questioned the optimality of touch-based biometrics due to the acknowledged likelihood of infection, primarily caused due to single point of contacts [13]. Noticeably, gait-based authentication systems apply behavioral characteristics rather body-part's specific cue to perform identification and authentication. Amongst other existing solutions, gait functions as a vital and user-friendly signature which doesn't need any continuous input by user for verification. Moreover, being touchless it turns out to be more effective scalable for authentication. Different modalities including sensing cameras [14], inertial sensors, wearable devices [15] and wireless sensing devices [12] are developed to retrieve gait-related features to enable person identification and authentication.

Gait-based authentication systems have evolved significantly in the form of model-based, model free, behavioral cue-based, where varied source of inputs have been applied like camera, wearable sensors, etc. [16]. These approaches exploit person's unique behavioral characteristics or movement pattern to perform identification and authentication [16]. Though, numerous efforts have been made towards vision-based methods, the allied challenges like local lighting conditions, viewing angle, shape, size etc. impacts generalizability as well as scalability of the solution. Silhouette image analysis [17]-based mechanisms undergo significant ambiguities, especially over large non-linear samples and hence high equal error rate (EER). To address such issues; though large number of spatio-temporal features have been extracted, it resulted high complexity, especially over real-time ecosystem [18]. Despite claimed high accuracy, these approaches often fall short in terms of computational efficiency, delay and scalability. Typically, gait-based authentication systems are divided into multiple types, like model-based, model-free and inertial measurement driven systems. Model-based method apply static features like body structure, movement patterns, key-points or landmark information (i.e., limb lengths, body height, body width [19] and dynamic parameters like angular velocities and walking speed) to perform person-authentication. On the contrary, model-free methods are free from any three-dimensional representation of human posture or walking behavior [19]. It mainly exploits statistical properties from the received gait-specific data to perform person authentication. In sync with model-based solutions, numerous efforts have applied by applying shape feature for authentication [17][19]; yet, its generalizability over large inputs seems questionable. Authors recently found that the use of multi-modal gait features can make person authentication more efficient; though, its scalability remained challenge [20-23]. It is mainly because of continuous data nature and respective cue-segmentation, feature extraction and learning [24-28].

Gait-based biometrics have been applied for both identification as well as authentication purpose [29]. Consequently, it demands industries to develop scalable solution with higher veracity. Unlike identification, which intends to identify the relationship of an unknown sample with the known gait-database; authentication requires validating the gait-sample by comparing enrolled gait-data to validate user. Thus, authentication measures require processing huge input samples and validate requested sample towards person authentication. Under such complex input scenarios, the classical approaches can be limited, and can be used merely for a definite sized ecosystem [30][31]. Methods like gravitational force measurement and smart-mats [32], force plates Derlatka and Bogdan [33] and floor vibration measurement [34] too are limited and suitable for indoor ecosystems. IMU-based methods have gained more attention because of cost-efficient and wearable solutions like smartwatches, wearable sensors etc. [29][35-41]. The ease of quantifiable latent information and its acquisition like accelerometer or gyrometer measurements has become easier due to different wearables and sensor availability. These effort-less gait-specific modalities can make person authentication scalable [37][42]. However, issues like latency, veracity, robustness and reliability remain challenge for industries [36]. None of the known approaches has addressed the issue of large data processing in real-time that can impose significant delay. Moreover, training over a huge non-linear input patterns might force machine learning model to undergo local minima and convergence and hence cause false positive performance [36]. Despite their claim to have achieved negligible or very small EER, authors failed in revealing that their proposed model was trained over merely 22 subjects [43] or smaller samples that raises question on its scalability. The likelihood of EER increases over increasing sample size [36][37], and therefore generalizing a solution based on merely two-digit data cannot be justifiable. Though, the use of smartphone or allied wearables can help in increasing a central-data driven authentication; however, training a model over original non-linear long-term pattern is a tedious task [44]. Moreover, a few researches have indicated that the efficacy of gait-behavioral driven models primarily depends on the position of sensor [29][44] and speed [39] that limits their suitability in contemporary world. To alleviate it, training a model with large heterogenous non-linear pattern seems to be a viable solution. The use of multi-modality or multiple features can make gait-based authentication more accurate [42]. Authors have also applied accelerometer and face data as multi-modal solution [42]; however, it seems to be irrelevant [45][46]. In the last few years, to exploit maximum possible spatio features, authors [47-50] applied deep learning methods like convolutional neural network (CNN); however, it lacked temporal cues to make classification better. Moreover, these approaches could not address the issue of scalability over the large subject presence. Authors have also used Long Short-Term Memory (LSTM) network [51] to exploit possible temporal features from the gait-sequence that resulted superior performance towards person identification [52]. Yet, the gait-cycle estimation imposed significantly high computation. Moreover, it failed in addressing large volume and allied expanses in real-time applications.

The emergence in micro-electronic mechanical sensors (MEMS) and inertial measurement units (IMU) and sensors have gained significant attention for gait data acquisition. Different tools like wearable sensors, smartphone, etc. [53-57] have been applied to retrieve person's gait patterns for further analysis. It has helped different (say, multi-modal) features like accelerometer (acceleration), and gyroscope (rotation), or even pace sensors (walking related force estimation) to record physical movement patterns, which are used to identify the person [53-57]. Existing inertial model-driven solutions use acceleration sequence retrieved from the different body-parts and walking patterns to identify person [53-57]. However, very minute quantitative (signal) presentation and local noise conditions limit efficacy of these approaches. Non-linear human behaviour, especially for large samples raises question over the scalability. IMU gait behavior-based authentication systems demand large continuous data and allied processing that makes it exhaustive and hence reduces longevity of the solution. The classical methods applying wavelet analysis [58] over continuous IMU makes it highly exhaustive, delayed and false-prone. Pattern matching over multi-modality features (like wearable driven acceleration or gyroscope outputs) can make gait-based authentication superior [53-57]; however, guaranteeing system scalability remains a challenge, which is not yet studied in academia. Noticeably, the term scalable signifies "the ability of a system to accommodate millions of the users simultaneously to serve personalized authentication". As the aforesaid term "scalability" defined, the demand of hour and future is to design a highly robust computationally efficient and scalable gait-driven solution which could provide accurate person authentication services with minimum exhaustion and high veracity. Though, to improve accuracy, authors have applied the different machine learning method [59] to train the input gait features; yet questionable to serve scalable system demands.

Considering above stated key challenges and allied scopes, this research hypothesizes that the use of sparse presentation over the input multi-modality features can make person authentication more scalable, reliable and enterprise-centric. In this paper, a highly robust sparse representation driven multi-modality feature learning environment for gait-based person authentication system is developed. As the name indicates, in this paper we applied acceleration and gyroscope signals obtained from the significantly large users to perform authentication. Here, the aforesaid multi-modality signals are processed for initial pre-processing and fixed windowing-based gait cycle estimation. Subsequently, a highly robust stationary wavelet transforms (SWT) driven stacked denoising auto-encoder (SDAE) model is developed that processes over the input multi-modal signals to generate corresponding sparse signal representation. Noticeably, the use of sparse signal makes computation more efficient by reducing the cost of original data processing. Thus, these approaches contributed a robust processing environment where the gigantically large continuous signals from large number of users can be converted into sparse data, and hence learning sparse data for gait-based authentication can make system more reliable, accurate and efficient. In our proposed model, once extracting the sparse data from the original multi-modal inputs (i.e., acceleration and gyroscope), we applied modified LTSM model that exploit spatio-temporal features to perform person authentication. Here, we modified LSTM in such manner that it retains maximum possible temporal features for further learning and classification. The proposed SWT-SDAE driven multi-modality Gait-features learning over LSTM has resulted superior accuracy (99.43%), precision (99.02%), recall (98.86%) and F-score with minimal known EER in target domain. It shows that the use of sparse inputs not only improves computational efficiency but also minimizes computational cost and energy exhaustion. It can greatly improve the longevity and lifecycle of power constrained devices for authentication. Moreover, SWT-SDAE can help accommodating large number of users for person authentication and hence can yield a scalable solution.

The remaining sections of this paper are divided as follows: Section 2 discusses some of the key literatures discussing related work, while the problem formulation and research questions are discussed in Section 3. Section 4 presents the overall proposed model and implementation, while the research conclusion and inferences are discussed in Section 5.

1. Related Work

Though, in the past different modalities including model-based, model free and IMU driven methods have been proposed towards gait-based authentication; however, the recent literatures have clearly indicated that the use of acceleration and gyroscope altogether can yield superior performance. Though, a number of efforts have been made towards vision-based gait-analysis for personal authentication [60-64]; however, the use of inertial sensor data have enabled more efficient and scalable solution [65][66-70][43][49]. In sync with this hypothesis, in this section the literatures primarily employing aforesaid gait-modalities are discussed for corresponding strengths as well as limitations. Different recent literatures [71][72] have revealed that smartphone or wearable gadgets driven systems can provide continuous movement and behavioral pattern of an individual that consequently can make pattern learning more efficient towards accurate identification and authentication. Recently, authors [71][72] claimed that the use of gyroscope can provide more intrinsic feature component for authentication. Though, the efforts made in [73][74] revealed that the smart-phone attached sensors can be deployed anywhere in the form of connected pouch or inside the pocket to provide continuous user's behavior for further authentication tasks. Interestingly, authors stated that they require measuring 20 to 50 samples per second to enable accurate learning, which can be computationally exhaustive for a large real-time authentication infrastructure. Undeniably, authors failed in addressing mammoth data and allied computing cost [75]. An effort by Osaka university [76] provided gait-data, where authors acquired movement pattern by using three sensors placed on subject's waist. Though, they didn't address data sparsity issues and its significance; yet, stated that the strategic use of accelerometer and gyroscope data can make authentication more scalable and accurate. Despite this fact, authors considered input session of 60 seconds, which can be insufficient towards accurate classification. Additionally, authors [76] failed in employing gyroscope data for analysis. In sync with feature extraction and learning, numerous researches have questioned the segment-based methods and cycle-based approaches [21][47][49]. Interestingly, in the cycle-based approaches, the retrieved activity data used to be a periodic signal where each cycle starts as soon as foot touches the ground and terminates when the same foot touches the ground for the subsequent time [77]. In this case, estimating cycle-length becomes complex, especially over a large user with non-linear walking patterns. Unlike cycle-based methods, segment-based approaches split input signal into fixed-length windows like 10 seconds, 30 seconds etc. However, in reality the gait-pattern of a user can be periodic and non-linear and therefore retaining merely 10 seconds of periodic input for analysis can miss significant features for further analysis, and hence can cause high EER. On the other hand, the activity streams pertaining to the period when one is sitting or sleeping does not make any sense towards gait- based authentication. Therefore, selecting effective features with sufficient intrinsic cue is vital towards learning and classification. In some recent studies, authors [78-80] have applied Dynamic Time Warping (DTW) method as distance measure for person authentication. In these methods, input gait sequences are split into gait-cycles by applying DTW, which was later used for comparison for authentication. Though, authors [41] applied Hidden Markov Models (HMM) to compare gait-cycle feature for classification. A few methods like cyclic rotation metric (CRM) [81] have been applied as the substitution of DTW to analyze gait-cycle for further authentication. Unlike these static approaches the signal analysis in time-domain and frequency domain can be of vital significance [42]. Moreover, the use of window-based segment [70][82] with calibrated size can also be vital to provide sufficient feature for further learning and classification [42]. Interestingly, no significant effort has been made towards time-frequency domain analysis with both acceleration and gyroscope data for gait-based authentication [83]. Authors [70] employed time domain features encompassing average time in between consecutive peaks to perform user identification. Researches indicates that the gait-cycle must be sufficiently large to accommodate at least one complete cycle [1][18][42]. The spectral relationships amongst the movement patterns of the different body parts [84] were applied to perform gait-based authentication. In [84], authors intended to exploit time-frequency representation representing the time-frequency signals simultaneously to perform person-authentication. Later, researches [36] found that unlike classical feature vectors the use of high-dimensional (minimum second-order) features can yield superior accuracy towards gait-based authentication. Moreover, the use of multi-modality feature fusion, especially over the different gait-inputs can yield reliable authentication solution [54][55]. Unlike decision-level fusion methods [85], feature-fusion can yield superior solution [42]. Numerous at hand methods have focused on applying more efficient feature learning methods to improve accuracy. In [86] different gait-signatures were applied to learn over gait-information using k-NN and radial basis function-based neuro-computing [87] classifier. In the last few years, deep learning methods including CNN and LSTMs have been applied to exploit spatio-temporal and temporal features, respectively to perform IMU driven gait-based authentication [1][42]. Unlike classical approaches, the use of deep learning methods helps in automatic feature extraction and learning over the input IMU data [68]. A few recent works at first transformed inertial data into an image, which was later processed with CNN for feature learning and classification towards person authentication. However, allied computational complexity might limit their efficacy towards real-time scalable system demands. Though, majority of the existing approaches indicate that the strategic amalgamation of multi-modality features encompassing acceleration and gyroscope data with discriminative features can yield superior performance. The inclusion of space representation can further improve the overall efficiency [15][27][47].

2. Research Questions

The last few years have witnessed exponential rise in advanced computing, low-cost hardware and allied application demands. Despite great innovations and evolutions, guaranteeing security to the systems, data or infrastructure has remained a challenge for industries. There are numerous application ecosystems like smart-home, resource-access, system-access, data security, infrastructure security etc., where person authentication turns out to be an inevitable need to ensure seamless and secure processing. In sync with such vital security demands, different approaches including cryptography variants, biometrics etc. have been proposed; however, the allied computational complexities, exhaustion, vulnerability and latency confine majority of the at-hand solutions to meet real-time authentication purpose(s). Moreover, with aforesaid exhaustive and vulnerable and low-efficacy solutions, guaranteeing of the solution is difficult. Though, biometric-based approaches yield superior performance, their computational complexities, and operating conditions cap scalability towards real-time purposes. Undeniably, fingerprint has been applied for person identification and authentication for a large ecosystem serving, smart-home, corporate identification and authentication, HMIs, IoT systems etc.; however, the recent COVID-19 pandemic has put question on its suitability, primarily because of central or common touch-based solution. To alleviate such problems, touchless biometrics can be of great significance. In sync with this demand, unlike classical biometrics gait-based approaches have gained widespread attention because of its touchless nature behavior-driven identification or authentication policy. Undeniably, numerous efforts have been made towards gait-based person authentication; however, the classical approaches employing model-based (static body shape, size and posture details) or even model-free (dynamic behavior) concepts undergo significantly huge computational overheads and high reliance onto the local conditions like lighting or illumination, noise, external outlook of a person etc. This as a result gives rise to the significant false positive or equal error rate. It confines scalability of these at hand solution. Unlike aforesaid gait-based modalities, recently smartwatches or IMU driven gait-based authentication systems have developed. Despite robustness over the classical gait-based authentication systems inertial measurement or sensor-based approach undergo numerous challenges including gait-cycle estimation, continuous large data acquisition and processing, low EER, noise-impact etc. However, despite such challenges, being easier to implement or carry, low-cost, wearables driven solution inertial map or sensor data-based solutions can be scalable. In other words, the use of sensor's data obtained from each user (through wearables, sensors, smartwatches etc.) can be applied to perform authentication, even for millions of users by applying certain centralized authentication unit; provided the model alleviates the key issue of time-efficiency, high veracity or reliability etc. In the past, though numerous efforts have been made by applying inertial measurements characterizing gait features to make person authentication; however, majority of the existing solutions have certain limitations like inefficient gait-cycle estimation, single modality-based training (i.e., either accelerometer output or gyroscope data), very small training data and hence lack of generalizability, and most importantly lack of scalability to cope up with real-time demands. Some of the approaches have exploited either temporal features or spatial cues to learn and classify. Though, a few methods have tried to use both accelerometer as well as gyroscope outputs for gait-based authentication; it failed in addressing the scalability problem and high EER. The depth assessment of the potential literatures reveal that the use of multi-modality features driven approaches can yield superior performance than the standalone feature-based solutions. Moreover, the use of segmented windowing concept too can help retaining intrinsic gait-related features for better learning and higher accuracy. Additionally, the most important inferences observed is that the use of sparse continuous signal can help accomplishing scalable gait-based authentication system. Sparse continuous data representation can help the authentication model to retain large window features (spatio-temporal) and hence the threat of important gait -feature loss can be avoided. Moreover, sparse presentation of the continuous inertial measurement can reduce computational exhaustion and hence a large number of users can be authenticated simultaneously without undergoing delay or exhaustion. These key scopes have been considered as the key driving forces behind this study.

In sync with above stated research goals and allied scopes, in this work a highly robust person Stacked Denoising Auto-Encoder Driven Sparse Multi-Modality Feature Representation model is developed for Gait-Based Scalable Person Authentication. As the name indicates, the proposed model applied SDAE auto-encoder for sparse continuous (multi-modality) data representation, which has been followed by feature extraction and classification using modified LSTM-RNN network. Being a multi-modality solution, in this work both accelerometer and gyroscopes data inputs are considered from a large user, which are initially processed for static segmentation or static window based continuous gait data retrieval. Once obtaining the windowed input signals from both accelerometer as well as gyroscope, at first SDAE model has been applied that applies stationary wavelet transform algorithm to retrieve detailed and approximated coefficients from each input gait-modality. In other words, once implementing SWT over each input gait-sequences it obtains approximated coefficient (over accelerometer and gyroscope segmented data), which is subsequently processed with SDAE model to retrieve sparse data representation of the input(s). In this manner, SDAE acts as a compressive sensing model that reduces data volume significantly without losing any intrinsic feature and allied significance towards person authentication. In function, SDAE applied SWT with a threshold method and inverse SWT (ISWT) to reconstruct the sparse inertial data representation. The sparse data over each input sequence is then processed with a modified LSTM that exploits maximum possible temporal as well as spatial cues to perform classification towards personal authentication. To assess efficacy of the proposed model, in this work multi-modality inputs encompassing both accelerometer and gyroscope output sequences were obtained from a benchmark dataset encompassing large users with non-linear patterns. The efficiency of the proposed model is examined in terms of accuracy, precision, recall, F-Score and EER. In sync with the overall proposed model, as discussed above, this research defines certain research questions. These questions are as follows:

RQ1: Can the use of Static Windowing and Stationary Wavelet Transform (SWT) driven SDAE sparse representation yield intrinsically feature enriched information to perform Person Authentication?

RQ2: Can the use of SDAE sparse representation over multi-modality Gait (say, feature) sequences (i.e., gyroscope data and accelerometer output) provide sufficiently large feature space for learning and classification to guarantee Person Authentication solution with high veracity and reliability?

RQ3: Can the use of multi-Modality features encompassing (sensor-driven) accelerometer and gyroscope data) yield highly accurate performance towards gait-based personal authentication?

RQ4: Can the strategic implementation of S2DAE with Multi-Modality Gait feature sequences enable scalable and highly efficient (and scalable) person authentication solution?

The overall research intends to achieve optimal answers for these key research questions.

3. System Model

The graphical depiction of the overall proposed model is given in Fig. 1.

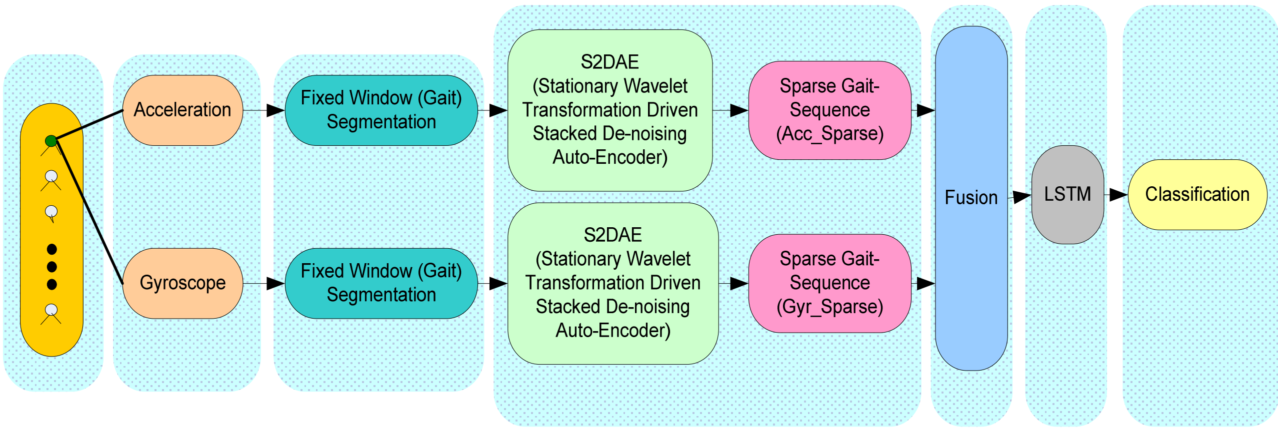

Fig. 1. Proposed Multi-Modality Sparse Gait-Sequence Driven Person Authentication System

The detailed discussion of the overall proposed model is given in the subsequent sections.

3.1. *Data Acquisition*

As discussed in the previous sections, sensory data or the inertial measurement units driven gait-sequences can yield superior results in comparison to the model-based or classical model-free approaches. Moreover, sensory details and allied acquisition is relatively easier where the different gadgets like wearables, sensory units, smartphone etc. can be employed. These data sources can provide continuous gait-data stream and hence deploying a scalable person authentication solution on to those data can become easier. Such systems can work in the same manner as Aadhar system employed in India, provided each user or person is assigned a unique ID for their corresponding gait-behavior. The use of aforesaid sensory inertial measurement data can be more effective and easier in comparison to the vision-based methods which require addressing mammoth challenges including the impact of orientation, local conditions, wearing, segmentation efficacy etc. Considering these key facts, in this paper we considered aforesaid sensory inertial measurement data or sensory data acquired by means of accelerometer and gyroscope. Noticeably, accelerometer and gyroscopes represent MEMS driven sensory units where accelerometer measures body speed or movement speed, while gyroscope measures body's rotation or angular movement. Thus, this research hypothesizes that the strategic amalgamation of linear two-dimensional movement with angular movement can provide sufficiently large gait-features to authenticate a person. Though, in the past authors have applied any of these two features as standalone gait-feature for learning and classification, we considered these together so as to improve the efficacy or veracity. The inclusion of these two gait-parameters (i.e., acceleration and gyroscope data) in sync with an authentication server help making optimal feature learning and hence scalable person authentication solution. In sync with this hypothesis, to retain a large feature instance for further learning, we considered OU-ISIR Dataset [88]. The intellectual rights of the Institute of Scientific and Industrial Research (ISIR), OSAKA University (Japan). To collect the data, the proprietary has applied three

IMUZ sensors, each encompassing a triaxial accelerometer and triaxial gyroscope. Noticeably, these key sensors were available in their considered smartphone, and not almost major smartphones carry these sensors. It can help generalizing this solution as scalable model. The data were collected for a large number of users, where the IMU measurements were done over different locations, ground conditions, sensor types, etc. Here, the key purpose was to retain maximum diversity, which is common in scalable system or allied demands. In sync with the scalable solution demands, we considered the multi-modal gait-sequence dataset from 744 subjects, encompassing 389 males and 355 females.

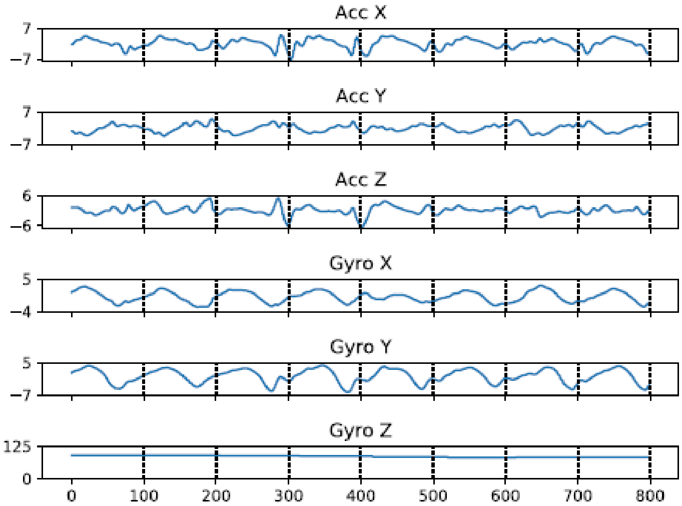

Fig. 2. Representarion of accelerometer and gyroscope data

Noticeably, the age of the subjects was in the range of two to 78 years. In the considered datasets, each subject contributed two different level-walk sequence which helped in assuring that the pattern with little changes can have the minute difference and hence a system in demand must be sensitive enough to classify or identify the user. Noticeably, the considered IMU data (i.e., accelerometer and gyroscope data) were obtained in three different dimensions (x, y, z). A symbolic representation of the collected gait data is given in Fig.2. Now, Once collecting a large set of input IMU data, we processed for gait-cycle estimation.

3.2. *Fixed Window Segmentation*

This is the matter of fact that human walking behave represents a continuous quasi-periodic activity, which originates from the first contact of the left foot with the ground to the alternation of the right foot till the left foot touches the ground again. In sync with our considered IMU gait-driven authentication demands, where multi-modal gait-sequences are considered for analysis, we require performing segmentation of the input time series data. In other words, we need to segment the input time-series gait-data, where each sample or segment can be further used for training and classification. Noticeably, here, each segment signifies the chunk of gait-sequence characterizing the movement behavior of subject. Typically, data segmentation is performed based on two methods, first, the period-based segmentation and second sliding window-based segmentation. Though, the selection of these segmentation methods depends on distinct (subject's) movement patterns and subject specific characteristics. For instance, the gait pattern for a child can different from the elderly. Typically, in period-based segmentation the input gait-sequence data is split into multiple chunks of the different length; however, in our considered scalable authentication solution this approach cannot be suitable. Unlike period-based segmentation, sliding window size-based approach split input sequence into fixed data segment. In this work, to ensure that the data or each segment retains sufficiently large continuous signal with intrinsic pattern, we applied sliding time window-based decision. We fixed the size of sequence in the gait-cycle as 4000 instances on time scale, also called gait-sequence length. In other words, on time-scale input presentation, we split input gait-sequence into regular blocks of gait-cycle size 4000 (say, length of sequence).

3.3. *Multi-Modal S2DAE Sparse Gait-Representation*

Once segmenting the input gait-sequences, to cope up with scalable solution demands, unlike classical approaches where authors [1] have applied merely gait-segment of length 120-140 (on time axis), we have considered large window size so as to guarantee that no significant gait or allied feature might miss-out. However, retaining gait-cycle of larger size or even complete input sequence might make a sensory system exhausting and even can limit the performance. On the contrary, over the large number of users, performing computation across input gait sequence (of large window size) can impact delay and scalability issue. Considering this fact, in this paper, we converted input gait-cycle segments into sparse representation. Noticeably, sparse representation enables data compression and volumetric reduction while retaining the same feature significance or allied intrinsic feature and quality. Sparse representation of the gait-sequence can not only help in reducing computational exhaustion but can also support infrastructure to accommodate large number of users and time-efficient authentication. We designed a stationary wavelet transform driven Stacked Denoising Auto-Encoder (SWT-SDAE) for gait-sparse representation, where the resulting outputs are further processed for LSTM for feature extraction, learning and classification. The detailed discussion of the overall proposed SWT-SDAE (here onwards called S2DAE) model is given in following section. The graphical depiction of the proposed S2DAE sparse (gait) representation model is given in Fig. 2. Noticeably, in gait-based personal authentication problem, the considered IMU sensory gait-sequences (especially over a large input from the different subjects) can have non-linear feature distribution. Therefore sparse gait-representation looks like a non-linear signal reconstruction problem. Considering this fact, before discussing S2DAE sparse representation, non-linear Sparse Gait-sequence reconstruction problem is discussed as follows:

3.3.1 Non-Linear Gait-Sequence Reconstruction Problem

In sparse signal representation and allied reconstruction problem, the key motive remains centered on solving a linear inverse problem, which is mathematically derived as per Eq. (1).

(1)

Though, for a predefined system the solution can be linear; however, under-determined systems consider the minimum energy solution as linear. Interestingly, in real-world gait-based authentication problem the data can be non-linear and hence corresponding signal sparse representation or reconstruction problem is non-linear. In this reference, the non-linearity involved in aforesaid non-linearity gait-signal can be defined as per Eq.(2). In realistic world, to achieve sparse representation of the continuous gait-sequence, we require solving l_0-minimization problem, as defined in Eq. (2).

$$\min_{x} \|x\|_0 \ \ \textit{subject to } y = \Phi x \tag{2}$$

Noticeably, the l_0-minimization problem Eq. (2) represents an NP-hard problem and hence require certain greedy concepts like orthogonal matching pursuit (OMP) [109] to achieve the corresponding sparse gait-sequence. These heuristic or greedy method identify the non-zero position in input sequence x, which is also referred as the support value at certain time. Consequently, it estimates the value of the detected non-zero element. In general, aforesaid greedy method for typical sparse representation problem follows the instruction as provided in Table 1.

Pseudo- OMP Greedy Model driven Sparse Gait Representation or Reconstruction

Input: *y, $\boldsymbol{\Phi}$, k (support)*
Initialize: $\boldsymbol{r} = \boldsymbol{y}$, $\boldsymbol{\Omega} = \emptyset$ *(set of support values)*
Iterate for k iterations
Estimate Correlation: $\boldsymbol{c} = \boldsymbol{abs}(\boldsymbol{\Phi^T r})$
Identify the value of Support: $\boldsymbol{l} = \underset{i}{\boldsymbol{argmax}}\, \boldsymbol{c_i}$
Tune the Support Value: $\boldsymbol{\Omega} = \boldsymbol{\Omega} \cup \boldsymbol{l}$
Calculate Support value at $\boldsymbol{\Omega}$*:* $\boldsymbol{x_\Omega} = \underset{x}{\boldsymbol{min}} \|\boldsymbol{y} - \boldsymbol{A_\Omega x_\Omega}\|_2^2$
Measure Residual value: $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{\Phi_\Omega x_\Omega}$
End

Table 1 Pseudo- OMP Greedy Model driven Sparse Gait Representation or Reconstruction

In the pseudo discussed above, Ω states the set of support values in x. The other variable A_Ω contains the specific attributes or columns (in A), which also have presence in the support vector or set of support values. Over each iteration, it results a solution containing values at the non-zero positions. To be noted, aforesaid OMP being a non-linear function, requires iterative calculation of the maximum value once detecting the support value and hence turns into a highly complex an iterative non-linear process-loop. To solve such problems without undergoing exhaustive calculations, we have considered convex-optimization measure as a viable solution for which l_0-norm is taken into consideration. In this gait-sequence sparse representation problem, the proposed convex optimization measure relaxes NP-hard l_0-norm to the adjoining or neighboring convex surrogate, l_1-norm. Numerous literatures find this approach robust towards non-linear continuous (data) sparse representation and hence turns out to be viable towards at hand gait-data's sparse reconstruction. Usually, this approach applied Eq. (3) to get sparse presentation.

$$\min_{x} \|y - \Phi_{\Omega} x_{\Omega}\|_2^2 + \lambda \|x\|_1 \qquad (3)$$

Though, literatures indicate that the easier approach to achieve Eq. (2) can be the use of Iterative Soft Thresholding Algorithm (ISTA) that embodies double-phase mechanism in each iteration (k in Eq. (4)). At first, it applied Landweber Iteration as presented in Eq. (4), while in subsequent step it applies a soft thresholding measure as defined in Eq. (5).

$$b = x_{k-1} + \sigma \Phi^T (y - A x_{k-1}) \qquad (4)$$

$$x_k = \mathrm{sign}(b) \max\left(0, |b| - \frac{\lambda\sigma}{2}\right) \qquad (5)$$

In Eq. (5), the parameter σ, signifying the step-size represents the inverse of the maximum Eigenvalue of $A^T A$. Though, the solution of Eq. (3) can apply any conventional gradient descent measure in the form of linear function; however, soft-thresholding measure as defined in Eq. (4) requires certain thresholding measure demanding ability to deal with non-linear sparsity. In reference to the gait-sequence data (IMU data), we considered soft-thresholding based approach for sparse data reconstruction over inputs.

To reduce computational exhaustion, especially caused because of non-linear inversion functions over iterative calculation, in this paper a data-driven learning environment was designed for sparse gait-sequence reconstruction. Neuro-computing methods have been applied as function approximator. These approaches, applies non-linear activation functions over the input training data to learn instance and generate intended data-sequence. Any continuous function signifying multiple variables can be approximated by performing superposition of the continuous functions of one instance. In sync with this inference, the proposed method applies the global functional approximation by applying deep neuro-computing method. The neuro-computing paradigm 'learns' inversion function by using SDAE algorithm. In other words, SDAE model intends to achieve an approximated inverse of a linear system, as defined in Eq. (6).

$$x' y = \Phi^T \Phi x \qquad (6)$$

Recalling the fact that the original gait-sequences and allied inversion to be employed for approximating x' can be noisy version of the realistic sparse solution x, and removing such noise component is inevitable. To alleviate such noise element soft-thresholding concept can be employed [110]. Despite numerous innovations, deep learning methods have exhibited superior over existing measures, as it can learn high-level hierarchy, despite by using low-level features. It can not only reduce computational overheads but also improves accuracy of sparse reconstruction. Considering low computational exhaustion, minimum hyper parameter tuning and ability to suppress local-minima and compression makes SDAE a potential solution for at hand gait-sparse representation. In this paper, a soft-thresholding driven auto-encoder model was developed by using stacked AEs to denoise the input gait-sequence. In other words, the proposed model applies numerous noisy versions of the gait-input sequence (say, segmented gait-sequence) which are fed as input to the auto-encoders to output a clean output with sparse signal representation. Here, auto-encoder is executed to learn map-information across input gait-sequences to generate expected output. As depicted in Fig. 3, in our proposed work, we retrieve the approximated coefficients from each gait-modality (i.e., accelerometer and gyroscope data) (say, x') by performing SWT and feed it to SDAE to generate corresponding sparse gait-representation. Unlike classical auto-encoder solutions, to refine resulting outputs (say, sparse data), we applied stacked AEs, also called Stacked AEs (SDAE).

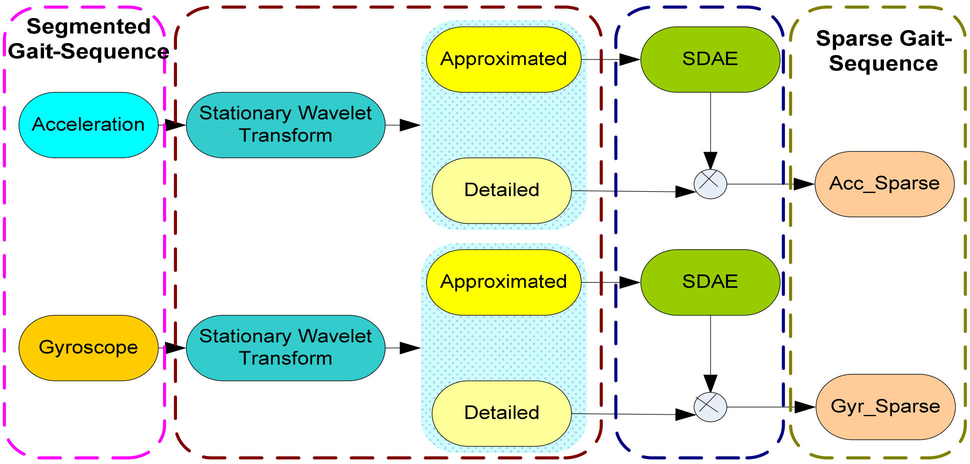

Fig 3. Sparse Gait-Sequence generation

The overall presentation of the proposed S2DAE sparse gait-representation model is depicted in Fig. 3. As depicted, in the proposed S2DAE model, each gait-sequence or modality is given as distinct input. Here, over each gait-sequence (note, here we pass segmented gait-sequence as input) is processed for SWT analysis which splits input gait-data into two sets of coefficients; approximated coefficient and detailed coefficient. Here, we select approximated coefficients (also called horizontal coefficients) from each modality and feed to the corresponding (distinct) SDAE, that converts each input gait-sequence into allied sparse output. The brief of SWT model and SDAE method used in this work is given in the subsequent sub-sections.

3.3.2 SWT Coefficients

To improve error performance and computational efficiency, in this work the original gait-sequences or segmented gait-samples were processed for SWT analysis, say wavelet analysis. Noticeably, unlike classical wavelet methods like discrete wavelet analysis or continuous wavelet analysis, the use of SWT enables low-dimensional feature generation that can further reduce computational overheads and redundant computation. The linear, shift-invariant characteristics of SWT enable better approximation of the results. Let, $x(a)$ be the segmented gait-sequence, then the proposed SWT model generates two coefficients (Fig. 4) named approximated coefficient $v_{i,k}$ and the detailed coefficient $w_{i,k}$, by applying two different filters; low-pass filters $\mathrm{H_i}$ and high-pass filters $\mathrm{G_i}$. In this work, these filters (i.e., H_1 and G_1) were retrieved by executing up-sampling of the filters over the descending step (i.e., H_{j-1} and G_{j-1}). In general, the detailed coefficient value $w_{i,k}$ remains same as the output of HPF. In the same manner, the output of approximated coefficient $v_{i,k}$ remains same as the output of LPF filter. To retain depth features, in this work we obtained Level-2 approximated coefficient (v_{2k}). To further improve intrinsic feature significance towards at hand gait sparse-representation, we employed a threshold on the basis of the percentage of R-peak. Here, we assigned threshold value as 0.99 with (level of significance 0.01). The approximated coefficients with R-value more than 0.99 were retained for further SDAE learning and allied sparse representation. To reconstruct the signal, we applied inverse SWT (ISWT) in sync with the detailed coefficients $w_{i,k}$ and SDAE sparse output that eventually generates the sparse representation for the input gait-sequence. In this work, we applied S2DAE over accelerometer and gyroscope data and thus, we obtain the sparse representation for both gait-modalities. A brief of the auto-encoder model applied in this work is given as follows:

3.3.3 Auto-Encoder

The approximated coefficients from the acceleration and gyroscope data, representing the under sampled measurement $\mathrm{Y} \in \mathrm{R}^{\mathrm{N}\times\mathrm{L}}$is applied to reconstruct the original signal $\mathrm{X} \in \mathrm{R}^{\mathrm{N}\times\mathrm{L}}$. It is achieved by applying matrix representation $\Phi \in \mathrm{R}^{\mathrm{m}\times\mathrm{N}}$, where $\mathrm{m} \ll \mathrm{N}$. It can also be rewritten as per Eq. (7).

$$Y = \Phi X + V \tag{7}$$

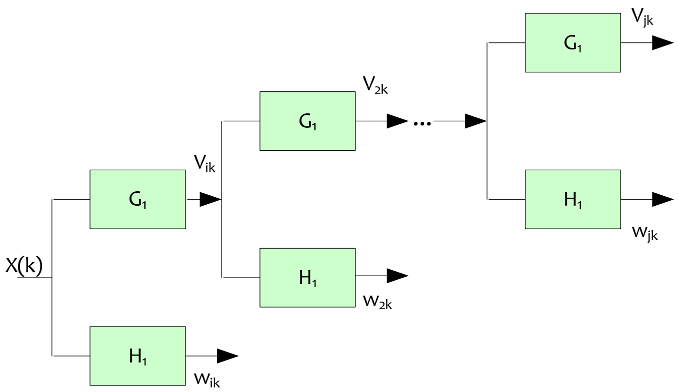

Fig 4 SWT coefficient estimation

In Eq. (7), $\mathrm{X} = [\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{L}] \in \mathrm{R}^{\mathrm{N}\times\mathrm{L}}$ represents the unknown matrix, while $\mathrm{Y}= [\mathrm{y}_1, \mathrm{y}_2, \dots, \mathrm{y}_\mathrm{L}] \in \mathrm{R}^{\mathrm{m}\times\mathrm{L}}$, represents the measurement matrix. Meanwhile, $\Phi \in \mathrm{R}^{\mathrm{m}\times\mathrm{N}}, \mathrm{m} \ll \mathrm{N}$ represents the sensing matrix and $\mathrm{V} \in \mathrm{R}^{\mathrm{m}\times\mathrm{N}}$ be the noise related components or vector. For a unit measurement vector, we assign $\mathrm{L} = 1$. On the contrary, for the multiple measurement vector we consider $\mathrm{L} > 1$. In sync with the at hand gait-sparse generation problem, the inverse problem the sparse data representation can be presented as per (8).

$$\hat{X} = \underset{X}{argmin} \|X\|_0 \ such\ that\ Y = \Phi X \tag{8}$$

Despite aforesaid sparse generation, ensuring source adaptive (here, we need to generate sparse data in reference to the gait-sequences) matrix (i.e., $\Phi \in \mathrm{R}^{\mathrm{m}\times\mathrm{N}}$) is a challenging task, especially in signal reconstruction which cannot be allowed in our proposed gait-driven person authentication system (as it would lead high EER and false positive). To alleviate it, we designed a data-adaptive learning model for sparse feature generation.

Typically, auto-encoder acts as a feature learning tool encompassing three layered neural network architecture, which help in reconstructing the un-labelled input data. In general, auto-encoder inputs input vector $x_i = [x_i(j)] \in R^n$ with $0 \le x_i(j) \le 1$, and encodes it to yield input for the hidden layer, $y_i \in R^m$. This as a result, helps in estimating the source signal, (i.e., $\hat{x}_i \in R^n$). In this process, a bias vector $b \in R^m$ is appended to the hidden layer and output layer so as to generate fine-grained feature or data learning ability. The values at the hidden layer and output layer be Eq. (9), and Eq. (10), respectively.

$$y_i = f\left(W^{(l)}x_i + b^{(l)}\right) \tag{9}$$

$$\hat{x}_i = f\left(W^{(l+1)}y_i + b^{(l+1)}\right) \tag{10}$$

In (Eq.9- Eq.10), $W^{(l)} \in R^{m\times n}$ states the weight factor, while bias values at the hidden layers and the output layer be $b^{(l)} \in R^m$ and $b^{(l+1)} \in R^n$, correspondingly. Though, auto-encoders can be designed as both non-linear as well as linear; however, in our case we designed it as non-linear activation using sigmoid function in Eq. (11). Though, it can be designed as per hyperbolic tangent function as well Eq. (12). Specifically, we applied sigmoid function in Eq. (11) with stochastic gradient descent for better learning and sparser output generation. In Eq. (11) and Eq.(12), $x_i \in R$.

$$f_s(x_i) = \frac{1}{1 + exp(-x_i)} \tag{11}$$

$$f_h(x_i) = \frac{exp(x_i) - exp(-x_i)}{exp(x_i) + exp(-x_i)} \tag{12}$$

In SDAE, for every auto-encoder layer we employed a non-linear hidden layer with linear activation function, often known as the identity activation function to design output layer acting as a linear decoder. Moreover, it applies a synthesis operator, $W^{(l+1)} = W^{(l)^T} \in R^{n\times m}$ to reconstruct source data, $x_i \in R^n$, with n as the block size. The hyper-parameter acting at the hidden layer y_i and output layer $\hat{x}_i$ are defined as per Eq. (13) and Eq. (14), respectively. To tune the allied hyper parameter, it applied a loss-function minimization concept. We used Euclidean distance-based loss function to tune hyper-parameters Eq. (15).

$$\Omega_1 = \left[vec(W^{(l)})^T, vec(b^{(l)})^T\right]^T \tag{13}$$

$$\Omega_2 = \left[vec(W^{(l+1)})^T, vec(b^{(l+1)})^T\right]^T \tag{14}$$

Noticeably, the complete training samples used is defined as k in Eq. (15).

$$L(x_i, \hat{x}_i) = \frac{1}{K}\sum_{i=1}^{K} \|x_i - \hat{x}_i\|_2^2 \tag{15}$$

To optimize Eq. (15), a non-linear differentiable activation function is used which is solved by means of Stochastic Gradient Decent algorithm. Let, the input of auto-encoder be the binary-valued vectors, then the cross-entropy loss function can be used Eq. (16).

$$L(x_i, \hat{x}_i) = x_i^T\, log_2(\hat{x}_i) + (1 - x_i)^T\, log_2(1 - \hat{x}_i) \tag{16}$$

Here, the loss function towards SDAE is reframed in the form of diversity measure, often called as the non-sparsity Eq. (17).

$$L(x_i, y_i) = \frac{1}{K}\sum_{i=1}^{K} \|x_i - \hat{x}_i\|_2^2 + \gamma\|y_i\|_1 \tag{17}$$

In Eq. (17), $\gamma \in R$ be the penalty imposed over the non-sparsity. Noticeably, SDAE employs loss function regularization to avoid over-fitting problem. As already stated in previous section, l_0-norm regularization methods are NP-hard in nature, while l_1 regularization can provide sparse model, thus enabling different hidden activations ($y_i(j) \in R$) as zero, and hence makes the corresponding attributes as insignificant. To alleviate such issues, dropout regularization is employed that drops certain units at varied layers. Moreover, dropout regularization helps in thwarting away the over-fitting issues and hence gains better efficiency.

The proposed SDAE model inculcates auto-encoder with the "denoising-capacity" and hence improved gait-intrinsic features cues to make classification accurate. It was done because while performing data acquisition through smartwatches or other IMUs, the likelihood of noise components can't be denied. Thus, in the proposed SDAE model, let the input gait-sequence is having noisy bits distributed throughout-sequence, $x_i \in R^n$, thus to generate the original gait-sequence x_i over the mixed noisy data $\tilde{x}_i \in R^n$, a loss-minimization concept is applied as per Eq. (18).

$$L(x_i, \tilde{x}_i) = \frac{1}{K}\sum_{i=1}^{K}\{\|x_i - \tilde{x}_i\|_2^2\} \tag{18}$$

Being SDAE, the proposed model applies cascaded sequence of auto-encoders or denoising auto-encoder. In this manner, initially a unit auto-encoder model is trained with the input gait-sequence (say, segmented gait-sequence) $x_i \in R^n$, where it applies unsupervised training in layered manner. Now, the trained hidden layers are employed as input to the next auto-encoder to perform training and sparse generation. Here, the hidden layer neurons of auto-encoder $y_i^{(\cdot)}$ is considered as the sparse generation function or the compression function. Thus, with the initial layer input as $x_i \in R^n$, the proposed model generates the measurement matrix using Eq. (19).

$$y_i^{(1)} = f_s\left(W^{(l)}x_i + b^{(l)}\right) \in R^m \tag{19}$$

Noticeably, the kind of compression depends on the weight matrix, $W^{(1)} \in R^{m\times n}$, Sigmoid activation function f_s (.). In this work, we employed input, $x_i \in R^n$ in sparse domain with a linear projection matrix, $W^{(1)} \in R^{m\times n}$ in sync with the activation function $f(z) = z$ for $\in$ R . In the subsequent layer, $y_i^{(1)}$ is fed as input, and thus the training is performed as per Eq. (20). Here, an unsupervised layer-wise pre-training in conjunction with a supervised (sub-sequent layer) tuning is done by means of back-propagation method. Here, each auto-encoder tunes corresponding hyper-parameters using (15). Additionally, the hyper-parameter to be improved be Eq. (21).

$$\begin{aligned} \hat{x}_i^{(1)} &= f_s\left(W^{(2)}y_i^{(1)} + b^{(2)}\right) \\ \boldsymbol{y_i^{(2)}} &= \boldsymbol{f_s}\left(\boldsymbol{W}^{(3)}\boldsymbol{\hat{x}_i^{(1)}} + \boldsymbol{b}^{(3)}\right) \\ \hat{x}_i &= f_s\left(W^{(4)}y_i^{(2)} + b^{(4)}\right) \end{aligned} \tag{20}$$

$$\Omega = \left[vec(W^{(1)})^T \cdots vec(W^{(l)})^T (b^{(1)})^T \cdots (b^{(l)})^T\right]^T \tag{21}$$

In Eq. (21), some pre-conditions were employed such as, $W^{(1)}, W^{(3)} \in R^{m\times n}$ and $W^{(2)}, W^{(4)} \in R^{n\times m}$. To alleviate above discussed hyper-parameter tuning demands and allied complexities, we applied mini-batch learning using Stochastic Gradient Decent (SGD). It helped in achieving swift convergence and hence alleviated local minima problem as well. This process enabled hyper-parameter update over a large batch size, where its performance remained close to the gradient descent method. Let, the input gait-sequence has n size data, then with μ learning rate (often we keep, 0.0001, or 0.001), the process of mini-batch updates parameters according to Eq. (22).

$$\Omega_k \leftarrow -\mu\frac{\partial}{\partial\Omega_k}\Omega_k \leftarrow \Omega_k - \mu\frac{\partial}{\partial\Omega_k}J\left(\Omega, \left(x_i^{(l)}, y_i^{(l)}\right), \cdots, \left(x_i^{(l+\zeta)}, y_i^{(l+\zeta)}\right)\right)\forall\, kJ\left(\Omega, \left(x_i^{(l)}, y_i^{(l)}\right), \cdots, \left(x_i^{(l+\zeta)}, y_i^{(l+\zeta)}\right)\right)\forall\, k \tag{22}$$

In Eq.(22) function ζ states the batch's size. The process of learning and eventual sparse output generation continues till each input segmented gait-sequences is processed. Now, once obtaining the sparse data output, we applied ISWT to reconstruct the input data (i.e., acceleration and gyroscope data). Thus, once generating the sparse data outputs from both IMU gait-sequence inputs, we feed them to the LSTM for feature extraction learning and classification. The detailed discussion of the LSTM method used is given in the subsequent section.

3.4. *LSTM Feature Extraction and Classification*

Once obtaining the sparse representation of the gait sequences (i.e., accelerometer and gyroscope data), we fed them as input to the LSTM model. Noticeably, unlike CNN which exploits merely spatial information to perform feature learning over continuous data, LSTM performs superior [47] by exploiting temporal features over continuous data stream. Considering continuous gait-sequence and allied frequent transmission (or assessment) nature, we applied LSTM as the deep model for feature extraction and classification. Though, numerous literatures like [47][52] state that recurrent neural networks (RNN) can be applied for continuous time-series data analysis; however, is unable to address the key issue of gradient disappearance. This as a result confine their suitability towards continuous long-range feature extraction and learning [1]. Considering aforesaid key limitations of the deep models, the use of LSTM seems to be a viable and potential solution [47]. In this work, we designed a deep architecture where LSTM acts as a storage element for RNN. The applied LSTM-RNN structure is illustrated in Fig.5. As depicted in Fig.5, the deployed LSTM possesses the ability to add or remove information (sequential sparse gait-inputs) to the cellular states by means of the component called "GATE" (Fig. 5). Here, GATE acts as a function which allows information to pass through and is designed with a sigmoid neural network layer with point-wise multiplication.

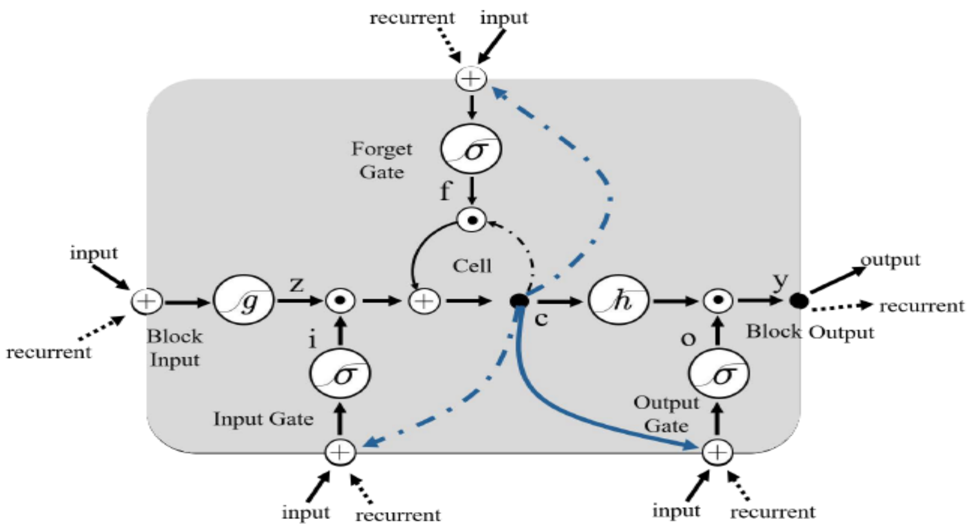

Fig 5. Proposed LSTM feature extraction

As depicted in Fig. 5, the deployed LSTM possesses three distinct gates, input gate, output gate and forget gate. These gates help in protecting and controlling the state of the cell. In this manner, the proposed LSTM-RNN architecture learns over the consecutive maps obtained from the input sparse gait-sequences to perform classification using Softmax layer. We applied LSTM architecture appended with Softmax layer which replaces the classical fully-connected later of CNN to perform feature learning and classification. LSTM architecture (used in this work) encompassed six distinct LSTMs as there are six different inputs from multi-modal gait-sequence. In other words, in sync with acceleration and gyroscope gat-sequences in $x, y, and\ z$ dimensions, we applied single LSTM for each sparse input, and hence a total of six LSTMs were taken into consideration. Here, each LSTM processes a distinct sparse gait-sequence (input) (say, $s_j^{(i)}$) and generates output vector (say, $f_j^{(i)}$) of length H. Subsequently, once each LSTM has resulted corresponding outputs, the extracted features were concatenated to generate a composite feature vector in Eq. (23) .

$$f^{(i)} = [f_1^{(i)}, f_2^{(i)}, f_3^{(i)}, f_4^{(i)}, f_5^{(i)}, f_6^{(i)}] \tag{23}$$

Now, the composite unique vector is passed on to the classifier as input. In the proposed design, with the input feature vector in Eq. (23), the classifier layer is nothing else but the fully connected layer that predicts people identify and authenticates it. In our proposed person authentication problem, for a user base comprising $N_u$users or members, the deployed fully connected layer maps the feature vector $f^{(i)}$ to the resulting (say, output) vector $y^{(i)}$. Here, each component of output $y^{(i)}$is measured from all comprising components of $f^{(i)}$ from Eq. (23). Functionally, it applied Softmax layer as activation function.

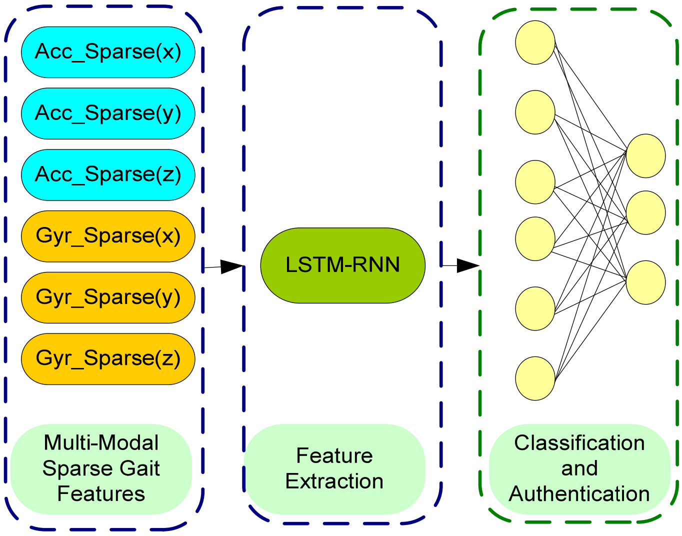

Fig 6. Sparse Gait-sequence driven LSTM feature extraction and Learning

To alleviate any possibility of over-fitting and local minima, we employed dropout with the rate of 0.5, and therefore it drops almost 50% of the feature maps along with corresponding connections. It not only improves computational efficiency, accuracy but also enables better network parameter tuning, learning and hence veracity. Unlike existing approaches like [47], where authors applied group size as 10, number of hidden layers as 40, and number of layers as two, in our proposed model we designed LSTM with the total hidden layers as 70, mini-batch size 27. Interestingly, in majority of the at hand solutions authors applied gait-sequence length as 100 on time-scale, our proposed model applied gait-sequence length of size 3000 (on time scale). However, with the use of sparse data generation and allied LSTM learning and classification, he proposed model yielded superior efficiency. Here, we used ADAM as adaptive learning algorithm.

4. Results and Discussion

In sync with the demand of scalable person authentication system for real-time purposes, this research work emphasized on optimizing both feature as well as allied computation. Here, the key hypothesis considered were that the inclusion of Inertial measurements from different wearables or smartwatches can be employed for person authentication. Moreover, this research hypothesized that the inclusion of sparse gait-representation (i.e., sparse IMU data) can reduce both volumetric complexities as well as allied computation that consequently can enable scalable person authentication solution. Nevertheless, it also considers that the strategic amalgamation of the multi-modal inertial measurement data, especially sparse data with LSTM which is a well-known temporal feature assessment model can yield superior solution towards person authentication. In sync with this hypothesis, in this paper a robust stacked denoising auto-encoder driven sparse multi-modality feature representation model was developed for Gait-Based Scalable Person Authentication. As the name indicates, being multimodal in nature, we exploited accelerometer and gyroscope inertial measurement data [88] for gait-based person authentication. Noticeably, unlike classical efforts [1][47] etc., where authors have merely applied 10s of training data even with 120-140 gait-samples (say, gait sequence length), in this paper we considered a total of 744 subjects pertaining to the different age-groups, gender. The data considered was collected even with the different movement patterns like step-up, step-down, normal walk etc. The key purpose of using such data with heterogeneity was to ensure that the training takes place over diversity of features to achieve higher sensitivity and scalability. Subsequently, unlike classical approaches where authors struggled in gait-segmentation and considered merely 120-140 instances on time scale (as gait sequence length), we considered relatively larger gait-sequence length (i.e., 3000 instance). Though, our key intension was to retain maximum possible gait-sequences and allied cues for better learning, it could have caused computational overhead. To alleviate such limitations, in this work a highly robust sparse reconstruction model was designed that transformed continuous gait-sequence into corresponding sparse feature vector. Noticeably, to retain optimal spatio-temporal feature and allied intrinsic feature distribution we considered both accelerometer as well as gyroscope outputs in x, y, z directions and hence, the total input vector was six. These six input vectors were at first processed for fixed window segmentation. This approach helped in reducing the computational cost of iterative gait-cycle estimation. Once segmenting the input gait-sequences, we processed it for sparse representation. This is the first of its kind solution where the input continuous gait-sequences are processed for sparse representation so that we might process a large sequence even with minimum computation without losing any intrinsic information. The input gait-sequence were obtained and saved in *.CSV, which was later processed for further processing. In this paper, we developed stationary wavelet transform (SWT) driven Stacked Denoising Auto-Encoder (S2DAE) model for sparse data reconstruction. More specifically, to fine tune the feature sanity, we designed a robust threshold adaptive SWT with SDAE for sparse data reconstruction. Here, SWT decomposes the input gait-sequence (which is segmented already) into two coefficients; approximate coefficient and detailed coefficient. Noticeably, we applied HAAR mother wavelet for SWT to decompose the input data and retrieve corresponding coefficients. The approximate coefficient was fed as input to the SDAE that retrieves sparse vector for each input sequence. Subsequently, the sparse outputs were processed with the detailed coefficient to reconstruct sparse gait-sequence output. Noticeably, we executed SDAE in batch-wise processing so as to reduce any likelihood of excessive hyper-parameter tuning. The final sparse outputs were fed as input to the LSTM that exploits temporal cues to learn over the continuous gait-sequence (say, sparse gait sequences). We designed LSTM with multiple convolutional layers, batch-normalization, dropout layer and fully connected layer. In fact, we designed the deep network in such manner that LSTM works as a memory element, while fully connected layer was replaced with Softmax layer to act as RNN. In this manner, our proposed LSTM model (say, Sparse LSTM) functioned as an LSTM-RNN algorithm for feature extraction and learning. Noticeably, we designed proposed LSTM-RNN as "$Input\ Layer > Convolutional - 1 > MaxPool > DropOut > Convolutional - 2 > MaxPool > DropOut > Convolutional - 3 > LSTM > SoftMax$". We assigned number of epochs as 200 to perform learning. Moreover, the learning rate assigned was 0.01, while the batch size was fixed at 8. We simulated our proposed model with non-linear adaptive learning concept named ADAM. Thus, executing above discussed method, the proposed model classifies many-to-one classification. The overall proposed model was developed using MATLAB 2020b, and the simulation was done onto Microsoft Window operating system armored with 8GB RAM, Intel i5 processor, functional at 3.2 GHz frequency. To assess performance by the proposed model, we obtained different statistical performance parameters including accuracy, precision, recall, F-measure, equal error rate (EER) etc. To estimate EER, we considered the standard approach defining (1-Accuracy) [18]. Noticeably, these parameters were derived from the confusion matrix, representing true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The definition of the key performance parameters is given in Table 2.

Parameter	Mathematical Expression
Accuracy	$\frac{(\boldsymbol{TN}+\boldsymbol{TP})}{(\boldsymbol{TN}+\boldsymbol{FN}+\boldsymbol{FP}+\boldsymbol{TP})}$
Precision	$\frac{\boldsymbol{TP}}{(\boldsymbol{TP}+\boldsymbol{FP})}$
Recall	$\frac{\boldsymbol{TP}}{(\boldsymbol{TP}+\boldsymbol{FN})}$
F-measure	$\mathbf{2}.\frac{\boldsymbol{Recall}.\boldsymbol{Precision}}{\boldsymbol{Recall}+\boldsymbol{Precision}}$
EER	$(\mathbf{1}-\boldsymbol{Accuracy})$

Table 2 Performance Parameters

The overall performance analysis and allied characterization was done in terms of intra-model assessment and inter-model assessment. Here, in intra-model assessment the efficiency of the proposed system was assessed in terms of accuracy, precision, recall, F-measure and EER. Since, in this research we hypothesized that the inclusion of sparse multi-modality gait-sequence can help achieving scalable authentication system, we simulated the proposed model with sparse data input as well as with the original gait-sequence. We applied LSTM-RNN over original gait sequence as obtained from the source data [88]. This assessment was considered as reference model. Though, almost all existing inertial measurement-based models have applied the same approach to perform person identification or classification. Unlike classical methods, we proposed S2DAE driven LSTM-RNN, and hence we simulated our proposed "S2DAE driven LSTM-RNN" over input gait-sequences, and compared the performance with original data driven LSTM for intra-model assessment. Here, the key motive was to assess whether the proposed sparse multi-modality driven approach yields superior performance or not. For inter-model assessment, we compared the efficiency of the proposed gait-based person authentication model with other state-of-art methods. The detailed discussion of the results obtained is given in the subsequent sections.

4.1. *Intra-Model Assessment*

As stated above, to assess efficacy of the proposed multi-modality gait-sequence based S2DAE driven LSTM-RNN model for person authentication, we executed the proposed model as well as a classical solution where the original gait-sequences were learnt over LSTM to perform classification. The details of such methods can be found in [1][18][47] etc. In other words, we passed original gait-sequence data (in *.CSV) to the static segmentation where we retained 3000 instances each segment (on time-scale). Subsequently, the proposed sparse gait sequence driven (i.e., S2DAE outputs) model was simulated with LSTM-RNN. Here, the key purpose to implement these two methods was to assess whether the proposed model (with sparse data) yields superior over the classical gait sequence driven LSTMs [1][27]. Since, this research hypothesized that the use of sparse data can yield superior over the complete original data processing, it is important to assess its efficacy. The simulation results obtained are given in Table 3. Noticeably, here onwards we call proposed system as "Sparse LSTM-RNN", while the simulation with original (non-sparse data) is referred as LSTM (Table 3).

Model	Parameters				
	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	ERR (%)
LSTM	93.70	94.41	94.65	94.52	6.3
Sparse LSTM-RNN	99.43	99.02	98.86	98.93	0.57

Table 3 Intra-Model Assessment

Observing the results (Table 3), it can be easily found that the proposed sparse LSTM-RNN model exhibits superior over the classical LSTM based person authentication system. More specifically, the classical approach in which the original gait sequence (post segmentation) is fed as input to the LSTM for temporal feature learning and classification results accuracy of 93.70%, precision 94.41%, recall 94.65% and F-Measure of 94.52%. Additionally, it yields ERR of 6.3%. Interestingly, even the LSTM driven model over original data yields superior results over numerous existing systems like [1][47][68], etc. Here, the key reason can be the use of multiple data together (i.e., accelerometer and gyroscope data together), with larger gait-cycle or gait-sequence length. Though, aforesaid methods have applied merely 120-140 instances on time-scale, while we had applied large window with 3000 instances and hence it enabled more intrinsic feature learning for classification. In contrast to the LSTM-based solution, our proposed sparse LSTM-RNN model was executed over sparse data and hence delivered superlative performance with accuracy of 99.43%, precision 99.02%, recall 98.86%, and F-measure or F-score of 98.93%. Moreover, the equal error rate (ERR) performance of 0.57% shows robustness of the proposed model towards real-time person authentication. Here, the significance of sparse data representation over large sequential inputs (six different gait-sequences from accelerometer and gyroscope sensors) can easily be visualized. The intra-model assessment results are shown in Fig. 5. As stated above, unlike classical approaches where to retain computational efficacy, authors [1][47][68] etc., authors retained smaller gait-window and even used standalone feature (either accelerometer or gyroscope in either of $x, y \, or \, z$ directions), we have applied multi-modal gait-signature in three different dimensions.

Considering this fact, we examined the efficacy of the proposed model over the different gait-sequence length (instances over time scale). Here, we wanted to assess whether the proposed model can yield satisfactory results with smaller window size or gait-sequence length or not. To be noted, despite using large window size, the complexity can increase only when the data is processed in its raw or original format. Since, in this paper we proposed for sparse data processing the computational cost automatically gets reduced significantly. However, to examine the performance we simulated the performance of the proposed model over the different window sizes, and the results are given in Table 4 and Fig. 8. The graphical depiction of the results over varying gait-sequence length or size is given in Fig.7.

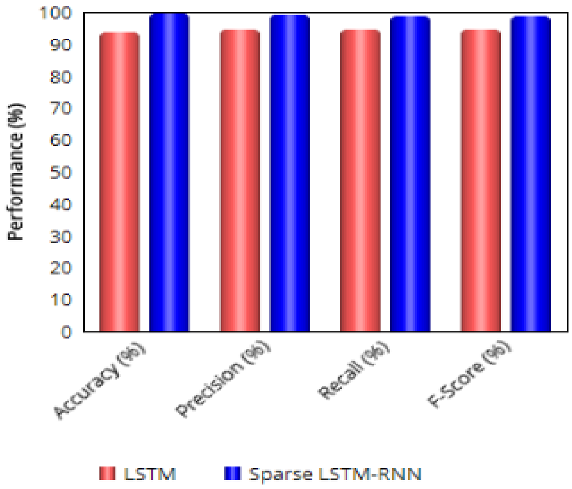

Fig 7. Intra-Model Assessment

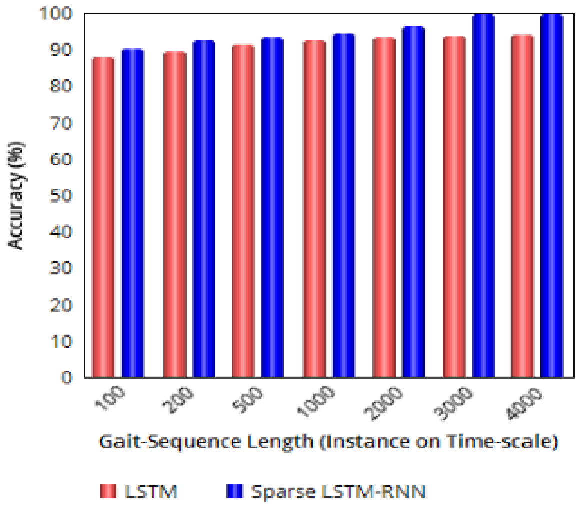

Fig.8 Accuracy (%) over different Gait-sequence length (time-scale)

Technique	**Gait-Sequence Size (on time scale)**													
	100		**200**		**500**		**1000**		**2000**		**3000**		**4000**	
	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)	Accuracy (%)	ERR (%)
LSTM	87.71	12.29	89.46	10.54	91.20	8.80	92.27	7.73	93.31	6.69	93.70	6.3	93.75	6.25
Sparse LSTM-RNN	89.96	10.04	92.31	7.69	92.98	7.02	94.36	5.64	96.01	3.99	99.43	0.57	99.45	0.55

Table4 Performance over different gait-sequence length

Observing the results, it can easily be found that the proposed sparse LSTM-RNN model yields superior accuracy even with the smaller gait-cycle, signifying that it can accommodate larger user request in run-time to yield authentication support without undergoing any delay or allied exhaustion. Interestingly, one can find that the proposed model yield superior EER performance even with very small gait-sequence and these results are quite encouraging than any other known method available.

In sync with the results obtained it can be stated that the proposed Sparse LSTM-RNN model (in sync with fixed window segmentation and S2DAE driven multi-modal feature learning and classification using LSTM-RNN) exhibits superior performance and hence can be capable to accommodate large users for their authentication without undergoing exhaustive computation or delay. This as a result can improve scalability of the inbuilt standalone authentication solution or for a third-party authentication infrastructure. Thus, observing overall results for intra-model assessment, it can be stated that the proposed sparse LSTM-RNN model with S2DAE sparse gait-sequences can be a viable and most effective approach towards real-time person authentication. Now, considering it as inferences, we have compared our proposed Sparse LSTM-RNN model with other state-of-art methods, which is discussed in the subsequent section.

4.2. *Inter-Model Assessment*

This is the matter of fact that in the past different approaches including vision-based, model-based, model-free methods have been developed towards gait-based person identification and authentication; however, since in this research we employed inertial measurement or sensory data for gait-based authentication, only those efforts employing aforesaid data are considered (as state-of-art methods) for relative comparison. To assess efficacy of the proposed model over the existing state-of-art approaches, we have identified accuracy and EER as the common performance measure or variable. The relative EER and Accuracy (in percentile) performance outputs by the different existing approaches are given in Table 5.

Reference	EER (%)	Accuracy (%)
[1]	-	96.9
[4]	6.15	-
[18]	-	91.8
[38]	13.5	-
[47]	3.37	89.79
[57]	6.7	-
[68]	-	94.8
[79]	5.7	-
[89]	-	93.3
[90]	20.1	-
[91]	-	85.48
[92]	-	90
[93]	6.1	-
[94]	10 (SVM), 2.36 (HMM)	-
[95]	6.15	-
[96]	10.1	-
[97]	8.24	-
[98]	29.39	-
[99]	-	91.33
[100]	1.85	-
[101]	-	97.9
[102]	15.08	-
[104]	13	-
[105]	-	98
[106]	-	93
[107]	17.2	-
[108]	-	93.2
Proposed	**0.57**	**99.89**

Table 5 Inter-Model Assessment

Authors in [89] applied accelerometer gait-sequence for person identification. Realizing the need of reduces sample size so as to minimize computational overheads, authors applied principal component analysis (PCA) over input gait-sequence. Despite their feature selection measure, the highest detection and classification accuracy observed was 93.3%, which is significantly lower than our proposed sparse LSTM-RNN model. In [90] as well smartphone driven accelerometer data was used for person authentication; however, the EER observed was 20%, which is significantly higher than our proposed model (0.57%). The accuracy observed in [91] was 85.48%.

Noticeably, authors in [89], [90], and [91] applied gait-sequence from 6, 51 and 6 subjects, respectively. It questions their robustness towards real-time application where the number of users can be in millions or even more. Authors in [93] as well applied merely acceleration data to perform person authentication. However, the EER observed was 6.1%. In [94], authors applied gait inertial data, which was trained over SVM and HMM models distinctly; however, the corresponding EER results were 10% and 2.36%, respectively. Interestingly, the same authors [95] claimed to have achieved EER of 6.15% using aforesaid HMM model. It raises suspicion and unreliability of the solution they proposed. Authors in [96] exploited gait-sequence data from 36 users, where the best EER observed was 10.1%. Authors in [97] applied k-NN classifier over gait inertial measurement, where the best EER observed was 8.24%. The EER performance in [98] was 29.39%, which is significantly high and can give rise to unreliability in authentication system. Cross-device gait-information were applied and trained over SVM and RBF machine learning methods. The highest accuracy observed was 91.33%, which is still lower than our proposed model. Authors in [4] applied HMM over gait-continuous sequences to perform user authentication. The best EER observed by authors was 6.15%, which is significantly higher than our proposed model. Time domain analysis with neuro-computing was done in [100]. Unfortunately, authors trained their model over merely five subjects and the best EER observed was 1.85%. A similar effort was made in [101], where authors applied neural network to train over inertial measurements obtained from I-Phone smartphone with the data collected from 8 subjects. The highest accuracy observed over time-domain feature analysis was 97.9%. A very recent work applied acceleration inertial measurement from smartphone towards person authentication. To assess their system efficacy over large samples, authors considered a total of 44 subjects and obtained the gait-sequence for the different activities. The depth assessment of their model stated that the EER varies as per the movement pattern types; yet, the cumulative performance assessment revealed 15.08% of EER over all activities. The EER observed here seems to be higher than any reliability system demands. A very interesting and effective approach was proposed in [4], named GaitCode. GaitCode was designed as a multi-modal solution where sensors were placed in socks and shoes which retrieved continuous movement patterns (acceleration and gravitational central force) of the user. Similar to our approach, authors applied auto-encoder as feature extraction model, which was followed by PCA-based feature selection. Finally dynamic time warping and SVM were applied to perform classification for authentication. Same as our efforts authors tried to retain high feature intrinsic cues, while retaining lower sample size to perform classification. However, it exhibited different EERs over the different inputs like socks input or shoes inertial inputs or gait-signatures. For instance, for shoes the EER observed was near 1.58%, while for socks it showed EER of 5.58%, which is higher in comparison to our proposed model. Authors in [103] applied accelerometer, gyroscope and GPS inputs altogether to perform gait-based person authentication. Yet, it could not be found justifiable towards real-time scalable authentication demands. The other efforts like [105][106] etc. applied different machine learning methods over inertial measurements derived from smartwatches to perform gait-based person authentication. Yet, the highest accuracy observed in [105] and [106] were 98% and 93%, respectively. Unfortunately, none of these approaches have considered scalability problem as they have merely applied a few dozens of subjects or even lesser to train a model, which can't be generalizable over real-time scalable authentication systems. In [68]., authors applied the same data OU-ISIR for gait-based person authentication, where CNN was applied as a feature learning model. The highest accuracy observed was 94.8%. Authors in [47] applied OU-ISIR data and used a hybrid deep model encompassing LSTM and CNN together to perform feature extraction and learning. Interestingly despite significant effort with aforesaid dataset, they could achieve the highest accuracy of 89.79%, which is significantly lower than our proposed model. Recalling the fact that unlike our 3000-instance selection or window segmentation, authors [47] has applied merely 120-140 instances on time-scale to perform feature extraction and learning. Authors in [1] converted inertial measurement data into equivalent images, which was later processed with CNN-LSTM to exploit spatio-temporal features for person authentication. This process was highly complex and exhaustive; yet delivered the accuracy of approximate 97%, though the subject considered were small and cannot be generalizable over large subject dataset. Authors [108] achieved applied inertial measurement data obtained from five subjects and despite complex processing could achieve the highest accuracy of 93.2%, which is significantly lower than our proposed model. Thus, observing overall results obtained and allied inferences it can be stated that the proposed sparse LSTM-RNN driven model delivers superior performance in terms of veracity as well as low-error likelihood. Consequently, it makes the proposed model robust towards scalable person authentication demand.

Conclusion

In sync with high-pace rise in scalable person authentication demand, this research proposed a highly robust multi-modality driven authentication system. In other words, this work intended to exploit both multi-modal gait-features and enhanced computing environment to ensure reliable, scalable and more importantly accurate person authentication solution. Realizing the at hand limitations of vison-based approaches or even existing model-based as well as model free gait-based authentication measures, in this paper wearable sensor driven authentication solution is designed. More specifically, in this paper inertial measurements including sensory details like accelerometer and gyroscope data were applied altogether to train a deep model for person authentication. Undeniably, the key motive behind the use of multi-modality (i.e., accelerometer and gyroscope data) was to retain maximum possible behavioral cues or features to learn and classify users for accurate person authentication task. Though, it could affirm higher efficacy; yet, its scalability over continuous non-linear gait patterns from thousands or millions of users might limit its realistic significance. Unfortunately, no potential effort could be made towards this neglected problem that might confine any of the existing IMU driven or sensory data driven solutions. Interestingly, most of the existing approaches have applied very small data (merely in tens), and hence generalizing their optimality over scalable demands seems exaggerated. And the key reason behind such limitation is ineffective gait-cycle identification or selection, low sample counts, standalone modality etc. To alleviate these problems, in this paper a robust stationary wavelet transforms (SWT) assisted Stacked Denoising Auto-Encoder (S2DAE) was developed that retained sparse presentation of the input multi-modal gait-data sequences (i.e., accelerometer and gyroscope data). At first, the input multi-modal gait-sequences were processed for static windowing or segmentation, which was processed with SWT-SDAE sparse data representation that guaranteed to retain the intrinsic features for further feature extraction and learning. Finally, we employed LSTM-RNN network to perform feature learning and classification. The simulation results revealed that the proposed model yields higher accuracy of 99.43%, precision 99.02%, recall 98.86%, F-Score 0.9893 and EER of 0.57% in comparison to the other relevant gait-based person authentication solutions. The superior efficacy even over the sparse (say, compressed) gait-sequence with large subjects retaining multi-modality features make the proposed system scalable and robust to meet up-surging person authentication demands.

Conflicts of Interest

The authors declare no conflict of interest.

Reference

[1] X. Zeng, X. Zhang, S. Yang, Z. Shi and C. Chi, "Gait-Based Implicit Authentication Using Edge Computing and Deep Learning for Mobile Devices", Sensors 2021, 21, 4592, pp. 1-23.
[2] Patel, V.M.; Chellappa, R.; Chandra, D.; Barbello, B. Continuous User Authentication on Mobile Devices: Recent progress and remaining challenges. IEEE Signal Process. Mag. 2016, 33, pp. 49–61.
[3] R. Delgado-Escaño, F. M. Castro, J. Ramos Cózar, M. J. Marín-Jiménez, N. Guil, "An End to-End Multi-Task and Fusion CNN for Inertial-Based Gait Recognition", IEEE Access, Vol. 7, 2019.
[4] I. Papavasileiou, Z. Qiao, C. Zhang, W. Zhang, J. Bi, S. Han, "GaitCode: Gait-based continuous authentication using multimodal learning and wearable sensors", Smart Health, Vol. 19, 100162, 20221, pp. 1-18.
[5] Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J. Spanos. 2018. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. Energy and Buildings 174 (2018), 309 – 322.
[6] YuanYuan Li and Lynne E Parker. 2008. Intruder detection using a wireless sensor network with an intelligent mobile robot response. In IEEE SoutheastCon 2008.
[7] Dexter D. Geralde, Milton M. Manaloto, Dionn Elijah D. Loresca, Jacqueline D. Reynoso, Eden T. Gabion, and Gabriel Rodnei M. Geslani. 2017. Microcontroller-based room access control system with professor attendance monitoring using fingerprint biometrics technology with backup keypad access system. In Proceedings of HNICEM
[8] Jutta Hämmerle-Uhl, Karl Raab, and Andreas Uhl. 2011. RobustWatermarking in Iris Recognition: Application Scenarios and Impact on Recognition Performance. SIGAPP Appl. Comput. Rev. 11, 3 (aug 2011), 6–18.
[9] Nazia Hossain and Mahmuda Nazin. 2018. EmoVoice: Finding My Mood from My Voice Signal. In Proceedings of ACM IMWUT.
[10] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In Proceedings of IEEE INFOCOM.
[11] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human Respiration Detection with Commodity Wifi Devices: Do User Location and Body Orientation Matter? In Proceedings of ACM IMWUT.
[12] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In Proceedings of ACM IMWUT.
[13] J. Galbally, L. Beslay, and G. Böstrom, "3D-FLARE: A Touchless Full-3D Fingerprint Recognition System Based on Laser Sensing", IEEE Access, VOLUME 8, 2020, pp. 145513- 145534.
[14] Ju Han and B. Bhanu. 2004. Statistical feature fusion for gait-based human recognition. In Proceedings of IEEE CVPR.
[15] Yuting Zhang, Gang Pan, Kui Jia, Minlong Lu, Yueming Wang, and Zhaohui Wu. 2015. Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points with Clusters. IEEE Transactions on Cybernetics 45, 2015.
[16] VM Zatsiorky, SL Werner, and MA Kaimin. 1994. Basic kinematics of walking. Step length and step frequency. A review. The Journal of sports medicine and physical fitness 34, 2 (June 1994).
[17] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 6, pp. 863–876, 2006.
[18] P. Musale, D. Baek, N. Werellagama, S. S. Woo and B. J. Choi, "You Walk, We Authenticate: Lightweight Seamless Authentication Based on Gait in Wearable IoT Systems," in IEEE Access, vol. 7, pp. 37883-37895, 2019.
[19] A. Ramesh, T. Karthi, and C. Thamilarasi, "Gait Analysis for Human Authentication using Depth Sensor", International Journal of Research and Advanced Development (IJRAD), Data Science & Big Data Analytics for Sustainability (ICDSBD2020), 2020, pp. 1-7.
[20] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in Proceeding IEEE Conference Computer Vision and Pattern Recognition, Kauai, HI, Dec. 2001, vol. I, pp. 439–476.
[21] A. Kale, A.K. RoyChowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in Proceeding IEEE International Conference Acoustics, Speech, and Signal Processing, Montreal, Canada, May 2004, vol. 5, pp. 901–904.
[22] J. E. Boyd, "Synchronization of oscillations for machine perception of gaits," Computer Vision and Image Understanding, vol. 96, no. 1, pp. 35–59, 2004.
[23] M. S. Nixon and J. N. Carter, "Advances in automatic gait recognition," in IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2004, pp. 139–144.
[24] M. Nixon, J. Carter, D. Cunado, P. Huang, and S. Stevenage, "Automatic gait recognition," in Biometrics. Springer, 1996, pp. 231–249.

[25] J. Han and B. Bhanu, “Individual recognition using gait energy image,” IEEE Trans. on pattern analysis and machine intelligence, vol. 28, no. 2, pp. 316–322, 2006.
[26] Z. Liu and S. Sarkar, “Improved gait recognition by gait dynamics normalization,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 6, pp. 863–876, 2006.
[27] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, “The humanID gait challenge problem: data sets, performance, and analysis,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, pp. 162–177, 2005.
[28] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, “Gait recognition: a challenging signal processing technology for biometric identification,” IEEE Signal Processing Magazine, vol. 22, no. 6, pp. 78–90, 2005.
[29] Zhong, Y., & Deng, Y. (2014). Sensor orientation invariant mobile gait biometrics. In Proc. IEEE IJCB conf (pp. 1–8).
[30] Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 316–322.
[31] Chen, C., Liang, J., Zhao, H., Hu, H., & Tian, J. (2009). Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters, 30, 977–984.
[32] Pataky, T. C., Mu, T., Bosch, K., Rosenbaum, D., & Goulermas, J. Y. (2012). Gait recognition: Highly unique dynamic plantar pressure patterns among 104 individuals. Journal of The Royal Society Interface, 9, 790–800.
[33] Derlatka, M., & Bogdan, M. (2015). Ensemble kNN classifiers for human gait recognition based on ground reaction forces. In Proc. IEEE HSI conf. (pp. 88–93).
[34] Pan, S., Yu, T., Mirshekari, M., Fagert, J., Bonde, A., Mengshoel, O. J., Noh, H. Y., & Zhang, P. (2017). FootprintID: Indoor pedestrian identification through ambient structural vibration sensing. Proc. ACM IMWUT Conf., 1, 1–31.
[35] Xu, W., Shen, Y., Zhang, Y., Bergmann, N., & Hu, W. (2017). Gait-watch: A context-aware authentication system for smart watch based on gait recognition. In Proc. IEEE/ACM IoTDI conf. (pp. 59–70).
[36] Sprager, S., & Juric, M. B. (2015a). An efficient HOS-Based gait authentication of accelerometer data. IEEE Transactions on Information Forensics and Security, 10, 1486–1498.
[37] Sprager, S., & Juric, M. B. (2015b). Inertial sensor-based gait recognition: A review. Sensors, 15, 22089–22127.
[38] Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., & Yagi, Y. (2015). Similar gait action recognition using an inertial sensor. Pattern Recognition, 48, 1289–1301.
[39] Juefei-Xu, F., Bhagavatula, C., Jaech, A., Prasad, U., & Savvides, M. (2012). Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics. Proc. IEEE BTAS Conf, 8–15
[40] Zhong, Y., Deng, Y., & Meltzner, G. (2015). Pace independent mobile gait biometrics. In Proc. IEEE BTAS conf (pp. 1–8).
[41] Nickel, C., & Busch, C. (2013). Classifying accelerometer data via hidden markov models to authenticate people by the way they walk. IEEE Aerospace and Electronic Systems Magazine, 28, 29–35.
[42] Zhang, Y., Pan, G., Jia, K., Lu, M., Wang, Y., & Wu, Z. (2015). Accelerometer-Based gait recognition by sparse representation of signature points with clusters. IEEE Trans. Cybern., 45, 1864–1875.
[43] Sun, H., & Yuao, T. (2012). Curve aligning approach for gait authentication based on a wearable accelerometer. Physiological Measurement, 33, 1111.
[44] Shen, C., Chen, Y., & Guan, X. (2018). Performance evaluation of implicit smartphones authentication via sensor-behavior analysis. Informing Science, 430, 538–553.
[45] Papavasileiou, I., Zhang, W., & Han, S. (2017a). Real-time data-driven gait phase detection using ground contact force measurements: Algorithms, platform design and performance (pp. 34–49). Smart Health.
[46] Papavasileiou, I., Zhang, W., Wang, X., Bi, J., Zhang, L., & Han, S. (2017b). Classification of neurological gait disorders using multi-task feature learning. In Proc. IEEE/ ACM CHASE conf (pp. 195–204).
[47] L. Tran, T. Hoang, T. Nguyen, H. Kim and D. Choi, "Multi-Model Long Short-Term Memory Network for Gait Recognition Using Window-Based Data Segment," in IEEE Access, vol. 9, pp. 23826-23839, 2021.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ``ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097_1105.
[49] K.-T. Nguyen, T.-L. Vo-Tran, D.-T. Dinh, and M.-T. Tran, ``Gait recognition with multi-region size convolutional neural network for authentication with wearable sensors," in Proc. Int. Conf. Future Data Secur. Eng. Cham, Switzerland: Springer, 2017, pp. 197_212.
[50] L. Tran and D. Choi, ``Data augmentation for inertial sensor-based gait deep neural network," IEEE Access, vol. 8, pp. 12364_12378, 2020.
[51] S. Hochreiter and J. Schmidhuber, ``Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735_1780, 1997.
[52] P. Fernandez-Lopez, J. Liu-Jimenez, K. Kiyokawa, Y. Wu, and R. Sanchez-Reillo, ``Recurrent neural network for inertial gait user recognition in smartphones," Sensors, vol. 19, no. 18, p. 4054, 2019.
[53] D. Gafurov, “Performance and security analysis of gait-based user authentication,” Ph.D. dissertation, University of Oslo, 2008.
[54] S. J. Morris, “A shoe-integrated sensor system for wireless gait analysis and real-time therapeutic feedback,” Ph.D. dissertation, Massachusetts Institute of Technology, 2004.
[55] B. Huang, M. Chen, P. Huang, and Y. Xu, “Gait modeling for human identification,” in IEEE Int. Conf. on Robotics and Automation. IEEE, 2007, pp. 4833–4838.
[56] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S.- M. Makela, “Identifying people from gait pattern with accelerometers,” in Defense and Security. Int. Society for Optics and Photonics, 2005, pp. 7–14.
[57] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, “Identification of individual walking patterns using gait acceleration,” in Int. Conf. on Bioinformatics and Biomedical Engineering, 2007, pp. 543–546.
[58] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, “A wearable acceleration sensor system for gait recognition,” in IEEE Conference on Industrial Electronics and Applications, 2007, pp. 2654–2659.
[59] B. Huang, M. Chen, P. Huang, and Y. Xu, “Gait modelling for human identification,” in IEEE Int. Conf. on Robotics and Automation. IEEE, 2007, pp. 4833–4838.
[60] M. J. Marín-Jiménez, F. M. Castro, N. Guil, F. de la Torre, and R. Medina-Carnicer, ``Deep multi-task learning for gait-based biometrics," in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 106_110.
[61] M. Alotaibi and A. Mahmood, ``Improved gait recognition based on specialized deep convolutional neural networks," in Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR), Oct. 2015, pp. 1_7.
[62] H. Su and F.-G. Huang, ``Human gait recognition based on motion analysis," in Proc. Int. Conf. Mach. Learn. Cybern., vol. 7, Aug. 2005, pp. 4464_4468.
[63] F. M. Castro, M. J. Marín-Jimènez, N. G. Mata, and R. Muñoz-Salinas, ``Fisher motion descriptor for multiview gait recognition," Int. J. Pattern Recognit. Artif. Intell., vol. 31, no. 01, p. 1756002, 2017.
[64] M. J. Marín-Jiménez, F. M. Castro, Á. Carmona-Poyato, and N. Guil, ``On howto improve tracklet-based gait recognition systems," Pattern Recognit. Lett., vol. 68, pp. 103_110, Dec. 2015.

[65] Z.Wei,W. Qinghui, D. Muqing, and L. Yiqi, ``A new inertial sensor-based gait recognition method via deterministic learning," in Proc. 34th Chin. Control Conf. (CCC), Jul. 2015, pp. 3908_3913.
[66] M. Gadaleta and M. Rossi. (Oct. 2016). ``IDNet: Smartphone-based gait recognition with convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1606.03238
[67] M. Gadaleta and M. Rossi, ``IDNet: Smartphone-based gait recognition with convolutional neural networks," Pattern Recognit., vol. 74, pp. 25_37, Feb. 2018.
[68] R. Delgado-Escaño, F. M. Castro, J. R. Cózar, M. J. Marín-Jiménez, and N. Guil, ``An end-to-end multi-task and fusion CNN for inertial-based gait recognition," IEEE Access, vol. 7, pp. 1897_1908, 2019.
[69] Q. Zou, Y.Wang, Q.Wang, Y. Zhao, and Q. Li, ``Deep learning-based gait recognition using smartphones in the wild," IEEE Trans. Inf. Forensics Security, vol. 15, pp. 3197_3212, 2020.
[70] J. R. Kwapisz, G. M.Weiss, and S. A. Moore, ``Cell phone-based biometric identification," in Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS), Sep. 2010, pp. 1_7.
[71] Capela, N.A., Lemaire, E.D., Baddour, N., Rudolf, M., Goljar, N., Burger, H., 2016. Evaluation of a smartphone human activity recognition application with able- bodied and stroke participants. Journal of Neuroeng. Rehabil. 13 (1), 1–10.
[72] Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J.M., 2016. Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors 16 (4), 1–24.
[73] Antos, S.A., Albert, M.V., Kording, K.P., 2014. Hand, belt, pocket or bag: practi- cal activity tracking with mobile phones. J. Neurosci. Methods 231, 22–30.
[74] Ganti, R.K., Srinivasan, S., Gacic, A., 2010. Multisensor fusion in smartphones for lifestyle monitoring. In: 2010 International Conference on Body Sensor Net- works, pp. 36–43.
[75] Gadaleta, M., Rossi, M., 2018. IDNet: smartphone-based gait recognition with convo- lutional neural networks. Pattern Recognit. 74, 25–37.
[76] Ngo, T.T., Makihara, Y., Nagahara, H., Mukaigawa, Y., Yagi, Y., 2014. The largest iner- tial sensor-based gait database and performance evaluation of gait-based per- sonal authentication. Pattern Recognit. 47 (1), 228–237.
[77] Derawi, M., Bours, P., 2013. Gait and activity recognition using commercial phones. Comput. Secur. 39 (July 2015), 137–144.
[78] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, ``Gait recognition using dynamic time warping," in Proc. IEEE 6th Workshop Multimedia Signal Process., Sep./Oct. 2004, pp. 263_266.
[79] D. Gafurov, E. Snekkenes, and P. Bours, ``Improved gait recognition performance using cycle matching," in Proc. IEEE 24th Int. Conf. Adv. Inf. Netw. Appl. Workshops, Apr. 2010, pp. 836_841.
[80] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, ``A wearable acceleration sensor system for gait recognition," in Proc. 2nd IEEE Conf. Ind. Electron. Appl., May 2007, pp. 2654_2659.
[81] M. O. Derawi, P. Bours, and K. Holien, ``Improved cycle detection for accelerometer-based gait authentication," in Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process., Oct. 2010, pp. 312_317.
[82] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. Reyes-Ortiz, Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. Berlin, Germany: Springer, 2012, pp. 216_223.
[83] Nakano, K., Chakraborty, B., 2017. Effect of dynamic feature for human activity recognition using smartphone sensors. In: 2017 IEEE 8th Intl. Conf. on Awareness Science and Technology (iCAST, pp. 539–543.
[84] O. Dehzangi, M. Taherisadr, and R. ChangalVala, ``IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," Sensors, vol. 17, no. 12, p. 2735, 2017.
[85] L. Kaliciak, H. Myrhaug, A. Goker, and D. Song, ``On the duality of specific early and late fusion strategies," in Proc. 17th Int. Conf. Inf. Fusion (FUSION), Jul. 2014, pp. 1_8.
[86] S. Choi, I.-H. Youn, R. LeMay, S. Burns, and J.-H. Youn, ``Biometric gait recognition based on wireless acceleration sensor using k-nearest neighbor classi_cation," in Proc. Int. Conf. Comput., Netw. Commun. (ICNC), Feb. 2014, pp. 1091_1095.
[87] Z. Wei,W. Qinghui, D. Muqing, and L. Yiqi, ``A new inertial sensor-based gait recognition method via deterministic learning," in Proc. 34th Chin. Control Conf. (CCC), Jul. 2015, pp. 3908_3913.
[88] http://www.am.sanken.osaka-u.ac.jp/BiometricDB/InertialGait.html
[89] Sprager, S. , Zazula, D , 2009. A cumulant-based method for gait identification using accelerometer data with principal component analysis and support vector ma- chine. WSEAS Trans. Signal Process. 5 (11), 369–378 .
[90] Derawi, M.O., Nickel, C., Bours, P., Busch, C., 2010. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In: 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 306–311.
[91] Frank, J. , Mannor, S. , Precup, D , 2010a. Activity and gait recognition with time-delay embeddings time-delay embeddings. In: AAAI Conference on Artificial Intelli- gence, 2010, pp. 407–408
[92] Kwapisz, J.R., Weiss, G.M., Moore, S.A., 2011. Activity recognition using cell phone accelerometers. ACM SIGKDD Explor. Newslett. 12 (2), 74–82.
[93] Nickel, C. , Brandt, H. , Busch, C. , 2011a. Classification of acceleration data for bio- metric gait recognition on mobile devices. In: BIOSIG 2011 –Proceedings of the Biometrics Special Interest Group. Gesellschaft für Informatik e.V, Bonn S. 57-66 .
[94] Nickel, C., Brandt, H., Busch, C., 2011b. Benchmarking the performance of SVMs and HMMs for accelerometer-based biometric gait recognition. In: 2011 IEEE Inter- national Symposium on Signal Processing and Information Technology (ISSPIT), 2011, pp. 281–286.
[95] Nickel, C., Derawi, M.O., Bours, P., Busch, C., 2011e. Scenario test of accelerometer- based biometric gait recognition. In: 2011 Third International Workshop on Se- curity and Communication Networks (IWSCN), pp. 15–21.
[96] Hestbek, M.R. , Nickel, C. , Busch, C. , 2012. Biometric gait recognition for mobile devices using wavelet transform and support vector machines. In: 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 205–210 .
[97] Nickel, C., Wirtl, T., Busch, C., 2012. Authentication of smartphone users based on the way they walk using k-NN algorithm. In: 2012 Eighth International Con- ference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 16–20.
[98] Muaaz, M., Mayrhofer, R., 2013. An analysis of different approaches to gait recog- nition using cell phone based accelerometers. In: Proceedings of International Conference on Advances in Mobile Computing & Multimedia, pp. 293–300.
[99] Hoang, T., Nguyen, T., Luong, C., Do, S., Choi, D, 2013b. Adaptive cross-device gait recognition using a mobile accelerometer. J. Inf. Process. Syst. 9 (2), 333–348.
[100] Watanabe, Y., 2014. Influence of holding smart phone for acceleration-based gait au- thentication. In: 2014 Fifth International Conference on Emerging Security Tech- nologies, pp. 30–33.
[101] Watanabe, Y., 2015. Toward application of immunity-based model to gait recogni- tion using smart phone sensors: a study of various walking states. Proc. Com- put. Sci. 60, 1856–1864.
[102] H. Alobaidi, N. Clarke, F. Li, A. Alruban, "Real-world smartphone-based gait recognition", Computers & Security 113 (2022) 102557, pp. 1-11.

e-ISSN : 0976-5166
p-ISSN : 2231-3850

[103] N. Al-Naffakh, N. Clarke, P. Haskell-Dowland, F. Li, "A Comprehensive Evaluation of Feature Selection for Gait Recognition Using Smartwatches", International Journal for Information Security Research (IJISR), Volume 6, Issue 3, Sep. 2016, pp. 691-700.
[104] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof Attacks on Gait Authentication System", IEEE Transactions on. Information. Forensics Security., vol. 2, no. 3, pp. 491–502, Sep. 2007.
[105] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015, pp. 1–6.
[106] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach", in 2016 IEEEEMBS International Conference on Biomedical and Health Informatics (BHI), 2016, pp. 426–429.
[107] E. Vildjiounaite, S.-M. M¨akel¨a, M. Lindholm, R. Riihim¨aki, V. Kyll¨onen, J. M¨antyj¨arvi, and H. Ailisto, "Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices," in 4th International Conference Pervasive Computing, 2006, pp. 187–201.
[108] Y. Zhang et al., "GaitSense: Towards Ubiquitous Gait-Based Human, Identification with Wi-Fi", ACM Trans. Sensor Netw., Vol. 1, No. 1, Article. Publication date: May 2021, pp. 1-25.
[109] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit", IEEE Transactions on Information Theory, Vol. 53 (12), pp. 4655-4666, 2007.
[110] D. Donoho, "De-noising by soft-thresholding", IEEE Transactions on Information Theory, Vol. 41(3), 613-627, 1995.

Authors Profile

Ambika K curenty working as an Assistant Professor in the department of Electronics and Telecommunication, BMS College of Engineering. Bengaluru. She received her BE degree from Kalpataru Institute of Technology in 2000. She obtained her MTech degree in Digital communication and Networking from SJCIT, Chikkaballapur. Currently Pursuing PhD in the stream of Biometrics under the guidance of Dr.Radhika K R

Dr Radhika K R, an academician, by choice, presumes in dedicated service towards teaching, research and academic activities. She has a rich experience of 25 years in teaching a wide spectrum of subjects in the areas of Information Technology at BMSCE. She has about 50+ publications in refereed journals. The noteworthy publications include Springer book chapter and Elsevier journals, Pattern Recognition, Applied Soft Computing, Pattern Recognition Letters and Journal of Visual Communication and Image Representation. Dr. Radhika has taken up initiative and lead role for various academic activities at BMSCE. She is a Senior member of IEEE. She was the key resource person for VTU e-learning content development and training as part of undergraduate curriculum. She is a Wipro Mission 10X Master Trainer and has organized faculty orientation programs at BMSCE. As an inspiring educationalist, she has mentored several students in different facets of their learning. Under her guidance, one of the research scholars has been awarded PhD. She is currently guiding five research scholars in specialized areas of Biometrics.