Computers & Security 136 (2024) 103559

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Learning a deep-feature clustering model for gait-based individual identification

Kamal Taha [b], Paul D. Yoo [a,*], Yousof Al-Hammadi [b], Sami Muhaidat [b,c], Chan Yeob Yeun [b]

[a] School of Computing and Mathematical Sciences, University of London, Birkbeck College, Malet Street, Bloomsbury, London WC1E 7HX, United Kingdom
[b] Department of EECS, Khalifa University, Abu Dhabi, United Arab Emirates
[c] Department of SCE, Carleton University, Ottawa, Canada

ARTICLE INFO

Keywords:
Gait biometrics
Forensic analysis
Machine learning
Clustering
Mixed data source

ABSTRACT

Gait biometrics which concern with recognizing individuals by the way they walk are of a paramount importance these days. Human gait is a candidate pathway for such identification tasks since other mechanisms can be concealed. Most common methodologies rely on analyzing 2D/3D images captured by surveillance cameras. Thus, the performance of such methods depends heavily on the quality of the images and the appearance variations of individuals. In this study, we describe how gait biometrics could be used in individuals' identification using a deep feature learning and inertial measurement unit (IMU) technology. We propose a model that recognizes the biological and physical characteristics of individuals, such as gender, age, height, and weight, by examining high-level representations constructed during its learning process. The effectiveness of the proposed model has been demonstrated by a set of experiments with a new gait dataset generated using a shoe-type based on a gait analysis sensor system. The experimental results show that the proposed model can achieve better identification accuracy than existing models, while also demonstrating more stable predictive performance across different classes. This makes the proposed model a promising alternative to current image-based modeling.

1. Introduction

Behavioral biometrics refer to recognizing individuals by analyzing their distinctive behavior characteristics, such as gait, without interfering with the activities of the subject individuals (Yan et al., 2015). In particular, gait biometrics or gait recognition verifies and identifies individuals based on their walking style (Alotaibi and Mahmood, 2015).

Several studies in the literature have focused on the analysis of human gait for identification tasks (Boulgouris and Chi, 2007; Kale et al., 2004; Wang et al., 2010). This includes gender classification (Hu et al., 2011) and age estimation (Lu and Tan, 2010; Makihara et al., 2011), where several benchmark datasets, *e.g.*, (Makihara et al., 2012; Okumura et al., 2010; Yu et al., 2006), were used to evaluate the performance of gait biometrics approaches. Most of the datasets used in gait biometrics studies are generated from 2D/3D images captured by surveillance cameras. These image based approaches however could be problematic in a sense that the quality of a gait-based identification model depends on the resolution of the captured images and variations of the appearance of individuals in a typical surveillance scene (Wang et al., 2016). For example, the appearance of an individual is subject to changes as it is affected by some exterior factors, such as clothing and environmental context, physical and mental conditions, such as age, gender, and illness, *etc*. Yan et al. (2015), observations effects, such as resolution, view angle, and position. A gait-based identification system should be immune to variations of intra-class representations and sensitive to variations of inter-class representations (Wang et al., 2016). Thus, it is important to learn generic discriminative representations of the gait of individuals in an automated fashion.

The proliferation of the Internet of Things (IoT), and its associated devices (*e.g.*, wearable programmable devices, sensors, *etc.*), makes collecting gait data of consenting individuals of such devices in non-interfering fashion possible. This provides an alternative of the image-based data to gait-based applications. In this study, we propose a novel gait-based identification/recognition model based on inertial locomotion information of the gait of individuals. The proposed model relies on newly designed deep feature extraction and clustering-based learning model. We first collect a new gait dataset using a newly developed shoe-type treadmill-based Inertial Measurement Unit (IMU) technology. The gait of each participant is represented by a set of attributes, which are then fed to a deep neural network of stacked sparse autoencoders to

* Corresponding author.
E-mail address: paul.d.yoo@ieee.org (P.D. Yoo).

https://doi.org/10.1016/j.cose.2023.103559
Received 26 March 2023; Received in revised form 21 September 2023; Accepted 21 October 2023
Available online 24 October 2023

generate high level features/representations, immunizing the proposed model to gait variations. The original attributes and generated features are then combined and fed to an algorithm, which selects the most discriminative features/representations of the data. The selected features are used to build a clustering-based learning model, which identifies the physical characteristics of individuals based on their gait information. The main contributions of this study are as follows:

- a new gait dataset is constructed using a shoe-based IMU technology;
- a machine learning model is developed where high-level representations of the data is combined with a randomized multi-layered clustering model;
- the provision of an algorithmic approach to validate the usefulness of Gait-IMU features, preventing incurring loss of information;
- the proposed model is clearly shown to have advantages over conventional learning models in the identification of characteristics of individuals; and,
- we prove that machine learning approaches using IMU technology focusing on gait analysis information have a potential for use in a number of branches of identification tasks.

The rest of this paper is organized as follows: Section 2 reviews related works. Section 3 presents the proposed model. Section 4 describes the experimental setup used and presents a discussion on the obtained results. Section 7 concludes the paper.

2. Related work

Over the last two decades, gait biometric has gained the attention of many researchers due to its applicability to the images generated by video surveillance cameras (Wang et al., 2016). The availability of large amount of gait data makes it a strong candidate solution for surveillance, criminal investigation and forensic applications (Bouchrika et al., 2011; Iwama et al., 2013; Lynnerup and Larsen, 2014). For instance, as we are already witnessing the emergence of Internet of Things (IoT), Closed-Circuit TeleVisions (CCTVs) are ubiquitous and widely deployed in public and private spaces. This facilitates the collection of gait information of individuals without a physical contact (Shiraga et al., 2016). A computational algorithm could then identify the subject individual based on distinctive features extracted from the gait information of subject individual.

Gait biometrics methods can be categorized into two major classes: model-based and model-free approaches (Alotaibi and Mahmood, 2015). Model-based approaches extract parameters that describe subject individuals using human body structure (Bhanu and Han, 2010). On the other hand, model-free approaches focus on the motion of the body where they extract descriptors based on silhouettes of the gait, which then used to build classification models. Han and Bhanu (2006) proposed a new image-based system named Gait Energy Image (GEI), a spatio-temporal gait representation, for individual recognition. The GEI is the average silhouette over one gait cycle, thus, it saves memory and computational time for recognition tasks. Lee and Grimson (2002) described a representation of human gait for individual identification and gender classification, where the representation of a gait appearance is based on simple features, such as moments extracted from orthogonal view video silhouettes. Boulgouris and Chi (2007) proposed a new feature extraction process based on radon transform of binary silhouettes, which are used to compute templates that are subsequently subjected to linear discriminate analysis and subspace projection. Alotaibi and Mahmood (2015) proposed a gait recognition technique using a specialized convolutional neural network. The proposed technique is less sensitive to common variations that affect and degrade the performance of gait recognition. Lu et al. (2014) investigated the problem of identifying human and gender recognition based on gait sequences with arbitrary walking directions. Wang et al. (2016) proposed a deep representation learning with an adaptive margin listwise loss for person re-identification, to deal with imbalance data where negative pairs outnumber positive pairs.

Most recently, Khan et al. (2021, 2023) conducted an exhaustive examination of codebook-based methodologies, elucidating the intricate procedures involved in encoding visual gait sequences while also identifying best practices to yield cutting-edge recognition outcomes. Notably, their research delved into the exploration of two distinct local features for encoding both the stationary visual attributes and dynamic motion characteristics of individuals in motion. Furthermore, they scrutinized twelve diverse feature encoding techniques. This comprehensive analysis encompassed an extensive evaluation on the sizeable CASIA-B gait database, culminating in a thorough presentation of the comparative performance results of these encoding methods.

Most of the aforementioned studies mainly use silhouette/visual gait sequence, which is normally extracted from 2D/3D images, of a subject individual for analysis, and most currently available datasets are geared toward this concept (Lee et al., 2014). However, high resolution images might not be always available due to environmental conditions, setting variations (*e.g.*, view angle), or limited capabilities of resources (*e.g.*, camera of low resolution). This study however uses an inertial locomotion gait dataset using IMU technology and proposes a model that builds a clustering-based learning model using high level representations extracted from a deep neural network of stacked autoencoders. The following subsections describe the generation process of the dataset and the proposed model.

3. Proposed model

The primary objective of this study is twofold. Firstly, it aims to discern and elucidate the distinctive attributes of individuals, encompassing gender, age, height, and weight, through an exclusive analysis of gait information only for forensic purposes. By leveraging advanced deep feature learning and state-of-the-art IMU technologies, high-level representations of gait patterns will be meticulously constructed. Secondly, this research endeavours to investigate the learnability and practicality of the afore-mentioned gait information representations, shedding light on their potential utility in diverse applications. The integration of cutting-edge methodologies and comprehensive data analysis will provide valuable insights into the characterization and applicability of gait-based forensic identification systems.

Recently, deep learning, which is also known as representation learning, has emerged as a new area of machine learning (Deng, 2014) and proven significant performance in different applications. For example, Le (2013) showed that it is possible to learn a face detector using only unlabeled images by training a deep neural network of locally connected sparse autoencoders. A deep neural network generates invariant representations of objects which allow detecting these objects even when they are in different position, size, *etc*. David and Netanyahu (2015).

The proposed approach adopts these principles and applies them for forensic identification of humans based on their gait information. The generated representations of the gait information would be invariant to changes, thus, are capable of identifying individuals even when there are changes in their walking patterns. In order to achieve this, we first need to generate a dataset that represents the gait patterns of individuals as fixed size vectors. A row in this dataset represents a person and a column represents a certain characteristic of the gait of the person, we fed the dataset into a deep neural network of stacked sparse autoencoders that generates high level representations of it. We then select the most discriminative features of the original and extracted features and use them to train a clustering-based learning model. It should be noted that we refer to the original characteristics of a person's gait as attributes whilst we refer to the generated characteristics/representations by the deep neural network as features. In what follows, we describe the generation process of the dataset, the deep features learning and selec-

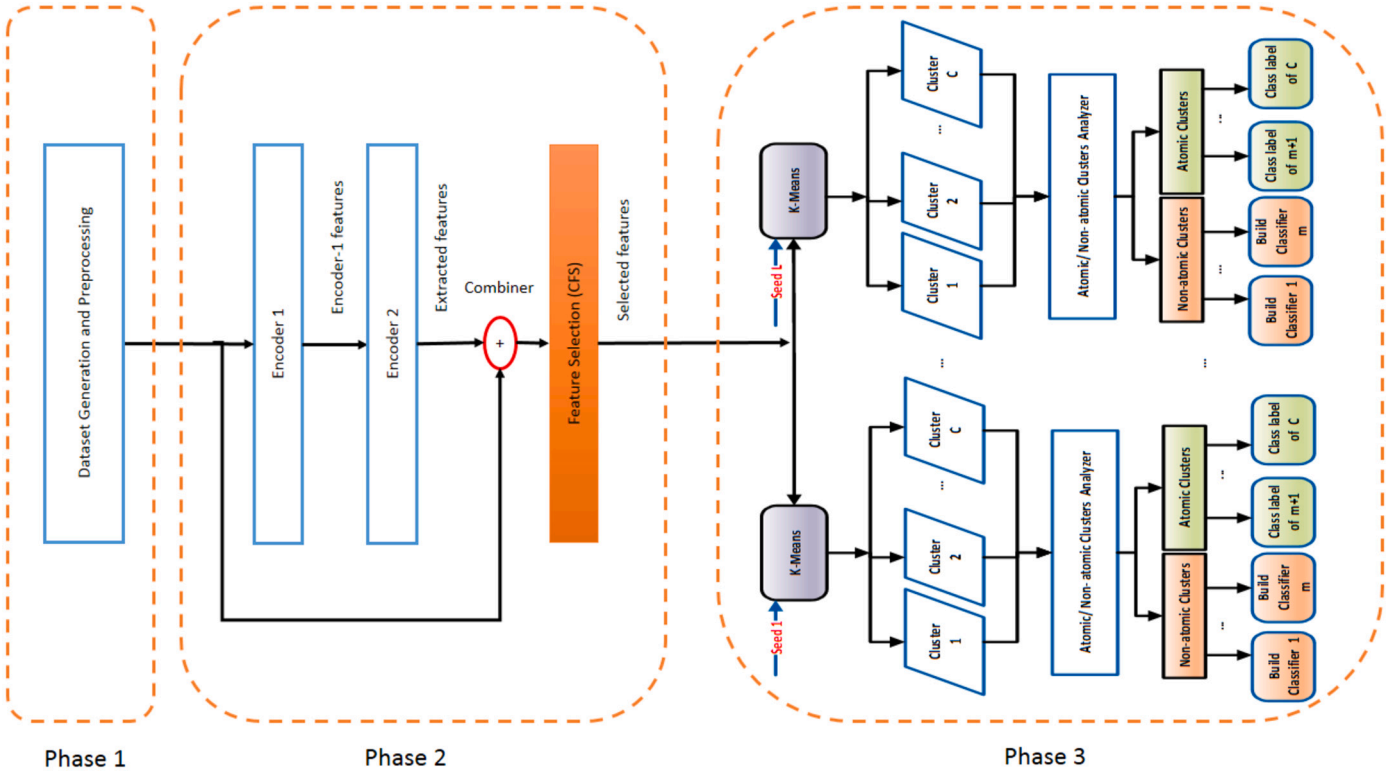

Fig. 1. Stepwise procedure for implementing the proposed model.

tion, and the clustering-based learning model. Fig. 1 shows the stepwise procedure of the proposed model.

3.1. Phase 1: Data collection

The Inertial Measurement Unit (IMU)-based dynamic body balancing analysis system (DynaStab^{TM}, JEIOS) including shoe-type data loggers (Smart Balance, JEIOS, South Korea), data collecting computers, and the motion capture system using nine infrared cameras (MXT10, Vicon) were used to collect the treadmill locomotion data. The data loggers, Smart Balance, including IMU (IMU-3000^{TM}, Inven Sense) can measure the three-axial accelerations (up to 6g) and three-axial angular velocities (up to 500 deg/s) of three orthogonal axes (Joo et al., 2015). These IMUs were located in both outsoles of the shoes. The local coordinate system of the IMU was established with anteroposterior (X-axis), mediolateral (Y-axis), and vertical (Z-axis) directions, respectively. Furthermore, the global coordinate system of the motion capture system was set behind the left side of the treadmill with mediolateral (X-axis), anteroposterior (Y-axis), and vertical directions, respectively.

The body height, weight, knee width, ankle width, and leg length of all 90 participants were measured to make each participant's body model and is used for the motion capture. After completing the somatometry, all the participants wore the spandex shirts, shorts, and the shoe-type data loggers. Sixteen reflective markers with a 14 mm spherical shape to the participants were attached according to the Plug-in gait model for motion capture. All markers were attached to the skin by double-sided tape and kinesiotape to fix the markers. These markers were attached to both left and right side, which were anterior superior iliac spine, posterior superior iliac spine, lateral thigh, lateral femoral epicondyle, lateral tibia, and lateral malleolus, and the calcaneus and second metatarsal head markers were attached on the shoes associated their anatomical landmark. Before starting the test, the participants performed the warm-up such as stretching and joint moving for 5 minutes. After completing the warm-up, they practiced a test walking on the treadmill to adapt to the self-selected speed for 10 minutes. After 30s from the start of the gait, we collected the treadmill gait data for 1 minute.

The data collection using both the IMU system and the motion capture system were sampled at 100 Hz, and the collected data were filtered using a second-order Butterworth low-pass filter with a cut off frequency of 10 Hz. Both measurement systems simultaneously collected the same subject's locomotion data. All the data were synchronized and analyzed based on the gait initiation point using DynaStab-Spotfire (Kim et al., 2016). The comparisons of variables between IMU system and motion capture system were conducted using spatiotemporal gait parameters and resultant linear acceleration variables for the entire treadmill gait trials for 1 minute. The spatiotemporal variables were cadence, step length, stride length, single support phase, and double support phase. The spatiotemporal variables from the motion capture data were calculated as follows: (a) the cadence (step/min) was calculated the total steps divided the 1 minute, (b) the step length was defined as distance between left and right heel strike (HS) during treadmill gait, (c) the stride length was defined as the distance between two subsequent midstance of the same foot. The resultant linear accelerations from the IMU were calculated by the net acceleration of X, Y, and Z axis at each left and right side. Furthermore, the resultant linear accelerations from the motion capture were calculated by the double differential of the heel marker's positions of the X, Y, and Z axis. The spatiotemporal variables from the motion capture and the resultant linear accelerations both IMU and motion capture were also derived using the DynaStab-Spotfire.

A paired t-test was used to compare the spatiotemporal variables between IMU and motion capture. The error between the IMU and motion capture was derived by root mean squares error (RMSE) over the total signal for the resultant linear acceleration (Esser et al., 2012), and a percent error was defined as the ratio between the RMSE value and the averaged peak-to-peak amplitude of the motion capture data (Mayagoitia et al., 2002). Pearson's product-moment correlation analysis was used to compare the validity of the spatiotemporal variables and the resultant linear acceleration between IMU and motion capture. Correlation coef-

ficients were interpreted as weak ($0.0 \leq r \leq 0.25$), fair ($0.25 \leq r \leq 0.50$), moderate to good ($0.50 \leq r \leq 0.75$), or strong ($0.75 \leq r \leq 1.0$) (Portney et al., 2000). The statistical significant level was set at 0.05.

IMU sensors can produce missing or corrupted data due to several factors, including noise, sensor errors, or connectivity issues. Fortunately, as seen in the provided dataset, no missing data observed in our case. However, if missing data is observed, then we suggest to employ the following strategies. Interpolation methods (Imoize et al., 2022), such as linear or cubic spline interpolation, can estimate missing values based on neighboring data points. Data smoothing techniques (Ageng et al., 2021) moving averages and low-pass filters can reduce noise and fill small gaps. Machine learning models, such as recurrent neural networks, can predict missing data, capturing complex patterns. Statistical imputation methods, like mean or median imputation, can replace missing values with summary statistics. Sensor fusion with other data sources can also help estimate missing IMU data (Girbés-Juan et al., 2021). Custom algorithms and data quality assessment are essential, along with documentation to ensure transparency and reproducibility. In dealing with missing IMU data, it is crucial to strike a balance between data recovery and preserving data integrity. A combination of these methods tailored to the specific characteristics of the data and the application's requirements is often necessary. Regular monitoring and calibration of IMU sensors can minimize missing data, while real-time data quality assessment can help detect issues as they arise. Ultimately, a thoughtful and systematic approach to handling missing IMU data enhances the reliability and accuracy of subsequent analyses and applications.

3.2. Phase 2: Deep feature learning and selection

3.2.1. Deep features learning with stacked autoencoders

The previous subsection describes how the movement of a person is converted to a fixed size vector. Given that the gait of an individual is subject to changes, identifying individuals based on their gait information is prone to errors as it depends on the physical and mental conditions of the individuals. For example, patients with Parkinson's Disease (PD) have difficulties in controlling their movement. Therefore, a set of high level representations of the gait of an individual that is resilient to variations, is to be created in order to identify different individuals accurately based on their way of walking. To this end, we construct a deep neural network of sparse autoencoders where the output of an autoencoder is fed into the following autoencoder in a hierarchical architecture. The basic idea is to define higher level representations of the gait of an individual from the lower ones where the same lower level representations help to define the higher level ones (Deng, 2014).

As in Fig. 2, an autoencoder is a deterministic feed-forward artificial neural network of one hidden layer and the output is set as the input. The number of neurons in the output layer is equal to the number of neurons in the input layer and there is a hidden layer of k neurons in between. Usually, the number of neurons in the hidden layer is set to a number that is less than the number of input neurons, in order to create a bottleneck and force the network to learn a compact higher level representation of the input. The model aims to learn an approximation $\boldsymbol{x'}$ of the input which would be more beneficial compared to the raw input $\boldsymbol{x}$ (Karalas et al., 2015).

An input $\boldsymbol{x} \in \mathbb{R}^p$, where p is the number of attributes, is transfered to a hidden representation $\boldsymbol{h}$ of k neurons/units through an encoder function (Karalas et al., 2015):

$$f(\boldsymbol{x}) = \boldsymbol{h} = \alpha_f(W_1\boldsymbol{x} + \boldsymbol{b}_1), \tag{1}$$

where α_f is an activation function. Logistic sigmoid and the hyperbolic tangent are typical examples of nonlinear activation function that are used in traditional autoencoders. As in equation (1), the activation function takes a weight matrix $W_1 \in \mathbb{R}^{k \times p}$, which represents weights learned on the connections between the input to the hidden layer, and a bias

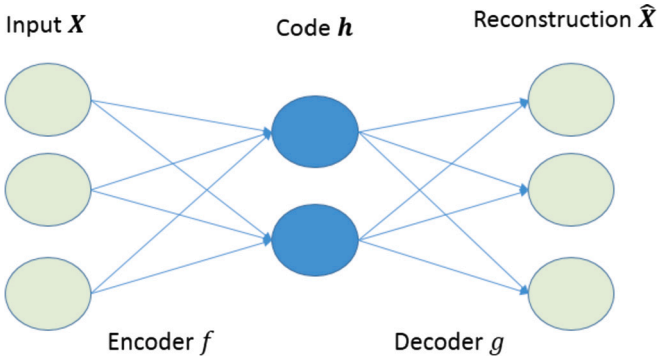

Fig. 2. Architecture of an autoencoder. The encoder takes the input $\boldsymbol{X}$ and computes the latent code $\boldsymbol{h}$. The decoder computes a reconstruction of $\boldsymbol{X}$ from $\boldsymbol{h}$ (Karalas et al., 2015). Note that $\hat{\boldsymbol{X}}$ is set to $\boldsymbol{X}$ and bias units are not presented for simplicity.

vector $\boldsymbol{b}_1 \in \mathbb{R}^{k \times 1}$. Then, the output is computed by mapping the hidden representation $\boldsymbol{h}$ back into a reconstructed $\hat{\boldsymbol{x}} \in \mathbb{R}^{p \times 1}$ using a decoder function as follows Karalas et al. (2015):

$$g(f(\boldsymbol{x})) = \hat{\boldsymbol{x}} = \alpha_g(W_2\boldsymbol{h} + \boldsymbol{b}_2), \tag{2}$$

where $W_2 \in \mathbb{R}^{p \times k}$ and $\boldsymbol{b_2}$ are the learned decoding weight matrix and the bias parameters, respectively. The model is parameterized by four parameters W_1, W_2, $\boldsymbol{b_1}$, and $\boldsymbol{b_2}$. These parameters are estimated by minimization the error between the input and the output based on a loss function, such as the normalized least square error given by Karalas et al. (2015):

$$J_{AE}(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}||\boldsymbol{x}^{(i)} - \hat{\boldsymbol{x}}^{(i)}||^2\right), \tag{3}$$

where $||.||$ is the Euclidean distance. Typical autoencoder is trained using backpropagation with stochastic gradient descent (David and Netanyahu, 2015). In order to prevent the problem of overfitting, a weight decay term is introduced to the cost function (Karalas et al., 2015).

Although the default assumption is that the number of neurons in the hidden layer is small, we can still have insightful structures from the data when the number of neurons in the hidden layer is greater than the input neurons, by imposing sparsity constraints on the hidden neurons, even when the number of hidden neurons is large (Ng et al., 2012).

Sparsity is introduced to the cost function of the autoencoder by penalizing it with a sparsity constant term ρ. Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951), which is a function to measure of the non-symmetric difference between two probability distributions ρ and $\hat{\rho}$, is used as a sparsity term in the cost function of the sparse autoencoder. The KL has a value of 0 when $\hat{\rho} = \rho$ and increases as $\hat{\rho}$ diverges from ρ. KL is given by Ng et al. (2012):

$$KL(\rho||\hat{\rho_u}) = \rho\log\frac{\rho}{\hat{\rho}_u} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_u}, \tag{4}$$

where $\hat{\rho_u}$ is the average latent unit activation given by Karalas et al. (2015):

$$\hat{\rho}_u = \frac{1}{m}\sum_{i=1}^{m}[\alpha_u(\boldsymbol{x}^{(i)})], u = 1, ..., k. \tag{5}$$

The regularized cost function of a sparse autoencoder is given by:

$$J_{SAE}(\theta) = J_{AE}(\theta) + \beta\sum_{j=1}^{k}(KL(\rho||\hat{\rho}_u)), \tag{6}$$

where the parameter β controls the effect of the sparsity regularizer.

Once a sparse autoencoder is trained, we can discard the decoder parameters and fix the parameters of the encoder, in order to keep the

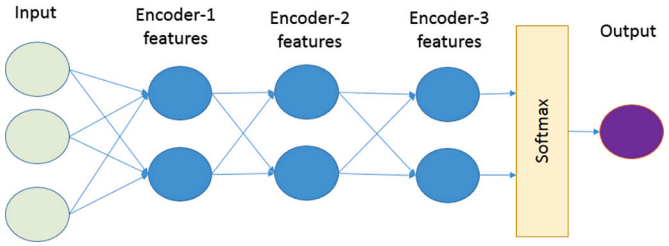

Fig. 3. Architecture of a typical deep neural network comprising stacked sparse autoencoders. It is important to emphasize that the network can accommodate any quantity of autoencoders and hidden neurons exceeding one.

layer unchanged. The output of the hidden layer is then fed into another sparse autoencoder as an input. The second sparse autoencoder is trained in a similar way to the training process of the first autoencoder. As seen in Fig. 3, a typical deep neural network is constructed by combining multiple sparse autoencoders in a hierarchical fashion where the output of an encoder is treated as an input to the following autoencoder. Ultimately, we aim to learn higher level representations of the input which refers to the low level features (*i.e.*, input layer). Conceptually, the subsequent layers correspond to higher level features. In a typical setting of a deep neural network, the output of the last hidden layer in the deep neural network is used to train a supervised softmax classifier. However, this study builds a novel clustering-based learning model based on selected features of the original attributes and generated features of stacked sparse autoencoders as described in Section 3.3.

3.2.2. Feature selection

Although the extracted features give high level representations of the input data, they may not improve the prediction accuracy. Here, we would like to know to what extent the extracted features are important and to see if they have a better significance than the original attributes/features or not. To address these questions, as in Fig. 4, the original attributes are combined with the features extracted by the deep network and then passed to a feature selection algorithm, which selects the most important and discriminative features of the data. In this study, we used a correlation based approach, namely Correlation based Feature Selection (CFS) (Hall, 1999), relying on the hypothesis that a good feature set contains features that are highly correlated with the class, but uncorrelated with each other. The CFS has been considered here as it quickly identifies and screens irrelevant, redundant, and noisy features, and identifies relevant features as long as their relevance does not strongly depend on other features (Hall, 1999). The selected features are then used to build a clustering-based learning model as described in Section 3.3.

3.3. Phase 3: Clustering-based learning

The learned latent features from a stacked autoencoder (SAE) described in Phase 2 are often more discriminative and relevant for clustering tasks compared to the original raw data due to the following reasons: 1) SAEs are trained to encode input data into a lower-dimensional representation, *i.e.*, latent features, which aim to retain the most relevant and discriminative information. In clustering, these features provide a compact yet informative representation of the data, making it easier for clustering algorithms to identify underlying patterns and similarities among data points, 2) SAEs are unsupervised learning models, meaning they do not rely on labeled data during training. As a result, the learned latent features are driven by the intrinsic structure of the data, capturing inherent data relationships that are often not apparent in the original raw data. This unsupervised feature learning helps in identifying meaningful clusters and revealing hidden data structures, 3) SAEs employ non-linear activation functions in their hidden layers, enabling them to model complex relationships in the data. This non-linearity allows the autoencoder to learn intricate decision boundaries, enabling the extracted latent features to better capture cluster boundaries and handle non-linearly separable data, and 4) The architecture of SAEs with multiple hidden layers allows them to learn hierarchical representations of the data. The lower layers capture simple and local features, while the higher layers capture more complex and global patterns. This hierarchical representation aids in identifying clusters at different levels of granularity, leading to more effective clustering results.

In this phase, an ensemble model consists of a set of base classifiers where the final prediction of the ensemble model is derived based on majority voting approach. It should be noted that the decisions of the base classifiers should be different in order to have an accurate decision. This can be reinforced by building heterogeneous or diverse base classifiers that commit non-identical error on a test instance. Diversity among base classifiers can be enforced by different methods, such as:

- using base classifiers of different types (*e.g.*, support vector machine, naive bayes, *etc*.) of an ensemble model.
- using different training datasets for building base classifiers of the same type of an ensemble model.
- using different parameters for base classifiers of the same type (*e.g.*, use different number of hidden layers or neurons for a neural network classifier) of an ensemble model.

Typical to Rahman and Verma (2011), diversity among base classifiers of the proposed model is provided by generating multiple layers of data clustering where different seeds are used in different layers of the ensemble model. The idea is based on the observation that the output of some clustering algorithms (*e.g.*, *K*-means algorithm) depends on the initialization parameters (*e.g.*, number of clusters and seed). In this context, a layer is defined as an instance of *K*-means clustering using a randomly selected seed. To clarify this, consider a dataset of two classes (male and female) as in Fig. 5 (a). This dataset can be partitioned into mutually exclusive clusters. For example, Fig. 5(b) shows resultant clusters of the *K*-means algorithm using a specific seed. On the other hand, Fig. 5(c) shows resultant clusters of the *K*-means algorithm using another seed. As can be seen in Fig. 5(b) and (c), clusters at the same layer are mutually exclusive; however, clusters at different layers are not. In other words, the clusters at different layers are not identical but might have overlapping data. Based on this observation, as in Fig. 5(d), it is possible to build an ensemble model that covers the whole search space by building base classifiers on the clusters at different layers of the *K*-means algorithm. This is because the training datasets of the constituent base classifiers of the ensemble model are not identical which enforces diversity among the base classifiers of the ensemble model.

As in Algorithm 1, the proposed model generates multiple layers of *K*-means clustering. In each layer l, there are C clusters. The resultant clusters of a layer could be one of two types:

- Atomic clusters, which refer to clusters that have only instances of one class (*e.g.*, a cluster that has instances of males only).
- Non-atomic clusters, which refer to clusters that have instances of different classes (*e.g.*, cluster that has males and females).

The proposed model builds classifiers on non-atomic clusters whereas it remembers the class labels of atomic clusters. This would reduce the complexity of the ensemble model as the model builds binary classifiers on non-atomic clusters only.

The testing process of a test instance begins with finding the nearest cluster's centroid at each layer. The cluster that has the minimum distance between its centroid and the testing instance at each layer is selected as the appropriate cluster. The corresponding classifier at each layer is then used to predict the class label of the testing instance. The final label of the testing instance is determined by the majority vote among all the predictions at different layers.

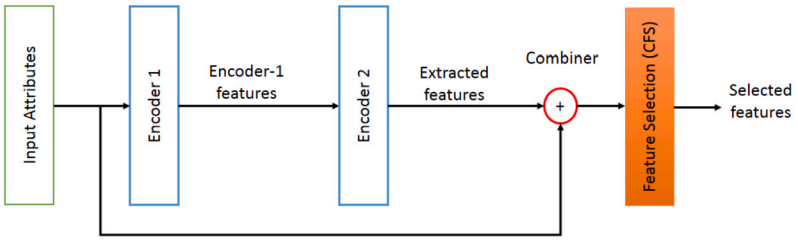

Fig. 4. Stepwise procedure of deep feature learning and selection.

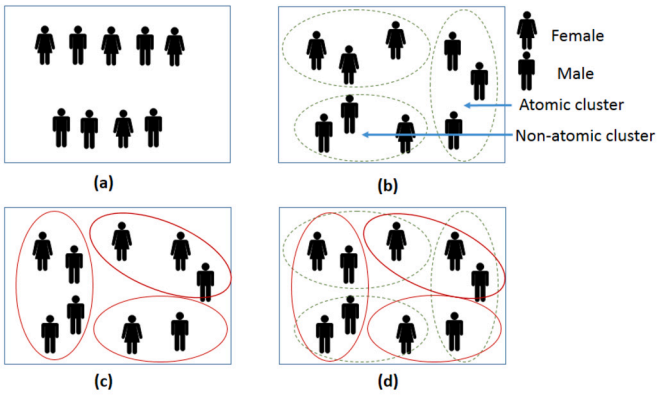

Fig. 5. (a) a dataset of two classes (male and female). (b) resultant clusters of K-means algorithm using seed 1 (*i.e.*, layer 1). (c) resultant clusters of K-means algorithm using seed 2 (*i.e.*, layer 2). (d) a projection of the resultant clusters at different layers.

Algorithm 1 Pseudo-code of building a clustering-based learning model.

```
Initialize the parameters of the model (L: number of layers in the model, C: number of clusters in each layer), and enter a training dataset D of N labeled instances.
for Each layer l do
    Partition D into C clusters using a random seed;
    for Each cluster c do
        if c is an atomic cluster then
            Remember the label of the majority class (Y_(l,c)) of the cluster c in the layer l;
        else
            Build a classifier M_(l,c) on c;
            Save the classifier M_(l,c);
        end if
    end for
end for
```

To model the human gait during locomotion, we utilized the Java and Python programming languages and employed numerical methods for simulation.

4. Experimental setup

4.1. Dataset and preprocessing

The gait dataset has 90 instances where each instance is represented by 162 independent (*i.e.*, explanatory) variables and 4 dependent variables (*i.e.*, response variables), namely, gender, age, height, and weight. The gender variable is categorical having two classes such as male and female while the other variables are numeric. The dataset is exclusively numerical, including the target variables; however, the data analysis technique of this study predicts the class of a nominal space. As such, it is necessary to transform the non-numeric variables into nominal space.

Often, uniform binning is used to perform data transformations for a machine learning model (Dougherty et al., 1995). However, the main problem of this method is with the way it determines the appropriate number of bins. A number of interesting studies have been reported in the literature in relation to finding optimum number of bins (Boulle, 2005; Dougherty et al., 1995; Yang and Webb, 2001, 2002). Dougherty et al. (1995) show the simplest approach, where the number of bins is set to 10 regardless of the number of instances. On the other hand, the K-proportional approach Yang and Webb (2001) set the number of bins to $\lfloor\sqrt{N}\rfloor$ where N is the number of instances. In this study, the number of bins is initially set to 10 and the number of bins is optimized using leave one out approach.[1] Table 1 shows statistics of the used dataset as per target variable after transforming numeric variables into the nominal space using binning.

The dataset and its description are publicly available.[2]

Table 1
Statistics of the dataset per target variable.

Target Label	Classes
Gender	Male (43), Female (47)
Age	-inf–55.6 (1), 55.6–59.2 (2), 59.2–62.8 (6), 62.8–66.4 (13), 66.4–inf (68)
Weight	-inf–46.21 (2), 46.21–50.42 (11), 50.42–54.63 (9), 54.63–58.84 (17), 58.84–63.05 (17), 63.05–67.26 (16), 67.26–inf (18)
Height	-inf–148.3 (8), 148.3–inf (82)

This table presents the class distribution of the target variables within the dataset following the discretization process. For instance, the 'Gender' target variable comprises two classes: the 'Male' class, consisting of 43 instances, and the 'Female' class, comprising 47 instances. Similarly, the 'Age' target variable has been discretized into five classes. The first class encompasses instances with an age below 55.6 years, totaling 1 instance. The second class includes instances with ages ranging between 55.6 and 59.2 years, encompassing 2 instances, and so forth.

4.2. Parameter tuning

As described in Section 3.2, we train a deep neural network consisting of two sparse autoencoders where the output of the first autoencoder is fed into the second one as an input. This process is mainly parameterized by the number of neurons of each autoencoder. To find the optimal number of neurons in each autoencoder, we initially set the number of neurons in each autoencoder to 10% (*i.e.*, 16) of the number of input attributes. We then increase the number of neurons in each autoencoder by 10% of the number of the input attributes iteratively until the number of neurons reaches the number of the input attributes (*i.e.*, 162).

Each iteration computes the recognition/identification accuracy of a softmax classifier. The combination of the number of neurons of the first and the second autoencoders that maximizes the prediction accuracy of the softmax classifier is chosen as the optimal numbers of neurons. This process is performed for each target variable (*i.e.*, gender, age, height, and weight). We also observed that when both the number of neurons of the first autoencoder and the number of neurons of the second autoencoder are equal to 16, the softmax classifier achieves the highest accuracy of predicting the gender. As mentioned in Section 3.2, the fea-

[1] weka.filters.unsupervised.attribute.Discretize.
[2] https://github.com/zaidalmahmoud/Gait-Analysis-Dataset.

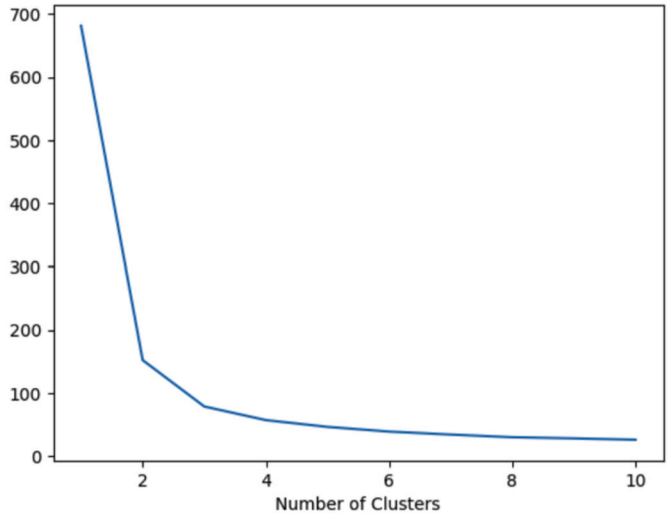

Fig. 6. The Within-Cluster Sum of Squares (WCSS) loss over different number of clusters.

tures extracted by the deep stacked sparse autoencoders are combined with the original attributes of the data and a feature selection algorithm is then used to select the most discriminative features of the data. The features selected by the feature selection algorithm are used to build the clustering-based learning model.

As in Section 3.3, the clustering-based learning model (*i.e.*, *K*-Means) is parameterized by the number of clusters and the number of layers. These parameters are selected by varying the number of clusters and layers and picking the numbers that maximize the prediction accuracy of the proposed learning model empirically. Within the *K*-Means framework, two parameters stand out as particularly pivotal: the number of clusters denoted as "*K*" and the initialization method.

In our quest to determine the ideal value for *K*, we leveraged the Elbow method. This method entails the creation of a graph that portrays the sum of squared distances, commonly referred to as "inertia," for data points with respect to their assigned clusters across various *K* values. Our objective was to discern the "elbow point" within the graph, characterized by a discernible inflection where the inertia diminishes at a notably slower rate. This juncture often offers valuable insights, signifying the optimal number of clusters that best characterize the data's underlying structure.

Following the WCSS (within-cluster sum of squares) analysis presented in Fig. 6, we have determined the most suitable number of clusters for our dataset to be three ($K=3$). Subsequently, we will proceed to construct a *K*-Means model that accommodates these three clusters and apply it to the IMU dataset. The pivotal step in this process is the utilization of the "fit-predict" method, which furnishes us with the cluster assignments or labels for each individual data point within the IRIS dataset. These labels serve as a basis for visualizing and interpreting the clustering outcomes.

Regarding the initialization of cluster centroids, it is vital to acknowledge that this process can exert a substantial influence on both the convergence dynamics and the quality of outcomes produced by the *K*-Means algorithm. Among the established initialization methodologies, notable options include random initialization, the *K*-Means++ initialization strategy, and the manual selection of a subset of data points to serve as initial centroids. The *K*-Means++ method, renowned for its ability to engender superior results compared to purely random initialization, has garnered popularity as a prudent choice.

5. Experimental results and analysis

To evaluate the effectiveness of the proposed method, we compare its performance with the performance of other machine learning methods, including Multi-Layered Perceptron (MLP) (Haykin, 2008), K-Nearest Neighbor (KNN) (Aha et al., 1991), AdaBoostM1 (Freund et al., 1996), and Random Forest (RF) (Breiman, 2001), on the generated gait dataset described in Subsection 3.1. The results of the proposed method were obtained by modifying the Java code of Weka package (Hall et al., 2009).

The most common and well-accepted statistical methods to evaluate the performance of a learning model are cross-validation and bootstrapping. In this study, we used 10 folds cross-validation approach to evaluate the recognition performance of the selected methods as it gives us a better understanding of how these methods perform on new datasets (Kohavi et al., 1995). The gait dataset is divided into 10 random subsets where 9 subsets are used for training and the other 1 subset is used for testing. This process is repeated, iteratively, until all subsets are tested. The classification errors of testing subsets are accumulated followed by computing the mean absolute error.

To demonstrate the significance of the extracted deep features in enhancing the performance of machine learning models, we evaluate the performance of the selected machine learning models on three different feature sets: feature set 1 (FS1), which consists of the original attributes of the data. Feature set 2 (FS2), which consists of the extracted features only; and feature set 3 (FS3), which consists of the selected features by the CFS feature selection algorithm from the combination of FS1 and FS2. The number of extracted features and selected features is not constant for all target variables/labels. For example, the number of extracted features is 16 when the target is gender; however, the number of extracted features is 32 when the target variable is age. This is due to the optimization process that selects the number of hidden neurons that maximizes the accuracy of a softmax classifier for each target variable.

5.1. Gait-based gender recognition

Table 2 shows the recognition accuracy of different models for the target variable 'gender' on each feature set. Accuracy in this classification problem refers to the proportion of correct outcomes, encompassing both True Positives (TP) and True Negatives (TN), out of the total instances assessed. The key performance metrics for this classification problem include Accuracy (ACC), Sensitivity (TPR), Specificity (SPC), False Positive Rate (FPR), and False Negative Rate (FNR), each defined as follows:

$$ACC = \frac{TP+TN}{TP+FN+FP+TN}, \tag{7}$$

$$TPR = \frac{TP}{TP+FN}, \tag{8}$$

$$SPC = \frac{TN}{FP+TN}, \tag{9}$$

$$FPR = \frac{FP}{FP+TN}, \tag{10}$$

$$FNR = \frac{FN}{FN+TP}, \tag{11}$$

where FP is the False Positive and FN is the False Negative.

As can be seen in Table 2, the MLP achieves the highest recognition accuracy of 68.888% when FS1 is used whereas the proposed clustering-based learning model achieves the highest recognition accuracy when FS2 and FS3 are used. Noticeably, the recognition accuracy of all models degrades when FS2 is used. This means that the extracted features, when used solely, do not improve the recognition accuracy of machine learning models for the target variable 'gender'. On the other hand, using a combination of both the original attributes and the extracted features improves the recognition accuracy of machine learning models for the target variable 'gender'. This is evidenced by the results shown in Table 2. The proposed clustering-based model achieves a recognition accuracy of 75.555% using FS3, outperforming all other learning models used in this study.

Table 3 provides a detailed examination of the proposed model's performance when applied to the optimal feature set, denoted as FS3,

Table 2
Recognition accuracy of different models for the target variable 'gender'.

Model	FS1	FS2	FS3
MLP	**68.888**	**53.333**	66.666
K-NN	60.000	46.666	55.555
AdaBoostM1	52.222	52.222	66.666
RF	52.222	38.888	58.888
Clustering-based	62.222	**53.333**	**75.555**

The presented table displays the performance results concerning the recognition accuracy achieved by specific machine learning models when applied to gait data generated for the target variable, which is gender. Within the table, FS1 denotes the inclusion of the original data attributes, FS2 represents the utilization of features extracted by a deep neural network consisting of stacked sparse autoencoders, as depicted in Fig. 4, and FS3 encompasses the features selected through the CFS algorithm, as illustrated in Fig. 4. It is noteworthy that Encoder 1 incorporates 16 hidden neurons, while Encoder 2 employs a similar configuration with 16 hidden neurons.

Table 3
Other performance measures on chosen feature set for the target variable 'gender'.

Model	TPR	SPC	FPR	FNR
MLP (FS1)	0.6809	0.6977	0.3023	0.3191
K-NN (FS1)	0.5778	0.6136	0.3864	0.4222
AdaBoostM1 (FS3)	0.6809	0.6512	0.3488	0.3191
RF (FS3)	0.5319	0.6512	0.3488	0.4681
Clustering-based (FS3)	0.7447	0.7674	0.2326	0.2553

This table presents a comparative analysis of performance metrics extracted from the confusion matrix, specifically, sensitivity (TPR), specificity (SPC), False Positive Rate (FPR), and False Negative Rate (FNR), for various machine learning models utilizing their most effective feature sets.

as initially presented in Table 2. The outcomes elucidate a substantial augmentation in both the Sensitivity (TPR) and Specificity (SPC). These results affirm the model's adeptness in accurately distinguishing instances pertaining to both male and female categories. The concurrent reduction in the False Positive Rate (FPR) serves as a robust indication of the model's improved precision, as it substantially lowers the likelihood of misclassifying an instance as female when, in fact, it belongs to the male category. This heightened precision is particularly advantageous for the accurate identification of female cases.

5.2. Gait-based age recognition

We investigate the recognition accuracy of the selected machine learning models for the target variable 'age' on FS1, FS2, and FS3. As seen in Table 1, most of the participants are of age greater than 66.4 years. Thus, the baseline accuracy is 75.555%.

Table 4 shows the recognition accuracy of the target variable 'age'. The proposed clustering-based model achieves the highest recognition accuracy of 76.666% using FS1. However, its recognition accuracy degrades when FS2 and FS3 are used. The RF classifier behaves similarly to the clustering-based model. However, the MLP, K-NN, and AdaBoostM1 classifiers show the highest recognition accuracy when FS2 is used. This means that the extracted features are good representatives of the data. Although the clustering-based model achieves a recognition accuracy of 76.666%, this is considered as a low recognition accuracy as the baseline accuracy is 75.555%. The low performance of different machine learning models is attributed to the imbalance data where 68 participants are of age greater than 66.4 years. This causes biasness toward the majority class as a machine learning algorithm sees more examples of the majority class during the learning process, which prevents recognizing instances of minority classes.

As shown in Table 5, the proposed algorithm demonstrates noteworthy proficiency in classifying classes from 3 to 5, while the RF model excels in categorizing classes from 1 to 2. In the case of class 1, both algorithms exhibit strong performance, achieving an accuracy of 0.88 for the clustering-based approach and 0.90 for RF. The F1 score, being the harmonic mean of precision and recall (TPR), assumes particular significance in this context. A high F1 score for class 1 indicates a harmonious balance between precision and recall. This implies that the classifier not only makes precise positive predictions (high precision) but also captures a substantial portion of the true positive instances (high recall). Thus, a high F1 score underscores the robust performance of the classifier for this specific class.

Table 4
Recognition accuracy of different models for the target variable 'age'.

Model	FS1	FS2	FS3
MLP	63.333	74.444	60.000
K-NN	55.555	67.777	55.555
AdaBoostM1	71.111	75.555	68.888
RF	74.444	68.888	**72.222**
Clustering-based	**76.666**	**75.555**	**72.222**

This table shows the recognition accuracy of selected machine learning models on the generated gait data for the target variable 'age'. FS1 contains the original attributes of the data. FS2 contains the extracted features by the deep neural network of stacked sparse autoencoders as in Fig. 4. FS3 contains the selected features by the CFS algorithm as in Fig. 4. The number of hidden neurons of Encoder 1 is 16 and the number of hidden neurons of Encoder 2 is 32.

Table 5
Other performance measures on chosen feature set for the target variable 'age' (Clustering-based (FS1) Vs RF (FS1)).

Class	ACC (%)	PPV	TPR	F1
1	83.33 / 85.56	0.95 / 0.98	0.82 / 0.82	0.88 / 0.90
2	84.44 / 86.67	0.47 / 0.53	0.62 / 0.62	0.53 / 0.57
3	90.00 / 87.78	0.33 / 0.27	0.50 / 0.50	0.40 / 0.35
4	96.67 / 93.33	0.33 / 0.0	0.50 / 0.0	0.40 / 0.0
5	98.89 / 95.56	0.50 / 0.0	1.0 / 0.0	0.67 / 0.0

Given the multi-class nature of the target variable 'Age', this table presents additional performance metrics, specifically, Accuracy (ACC), Precision (PPV), Sensitivity (TPR), and F1 Score, for the two top-performing models and their respective feature sets as displayed in Table 4. In each column, the value on the left signifies the performance measure for the clustering-based approach, while the value on the right represents the performance of the Random Forest (RF) model. For instance, in the case of class 5, the accuracy for the clustering-based method is 98.89, whereas it is 95.56 for the RF model.

5.3. Gait-based weight recognition

Table 6 shows the recognition accuracy of different machine learning models for the target variable 'weight'. As seen in Table 6, the clustering-based model shows the best recognition performance when FS1, FS2, and FS3 are used. Of particular interest, most of the models achieve a comparable or better performance denoted by the recognition accuracy using FS2. For example, the MLP achieves a recognition accuracy of 15.555% using FS1; and a recognition accuracy of 20% using FS2. The K-NN achieves a recognition accuracy of 6.666% using FS3; and a recognition accuracy of 20% using FS2. This clearly indicates that features extracted from the deep neural network are good representatives of the input data and can potentially improve the recognition accuracy of machine learning models for the target variable 'weight'. Noticeably, the clustering-based model achieves recognition accuracy of 26.666% using FS2, outperforming other conventional machine learning models.

As depicted in Table 1, the Weight variable encompasses a total of seven distinct classes. Notably, the clustering-based model exhibits exceptional performance, particularly in its ability to accurately classify classes ranging from 5 to 7, as well as class 1, all achieving accuracy levels surpassing 80%. Substantial TPRs are evident in classes 3 and 7. These observations imply that the clustering-based model demonstrates

Table 6
Recognition accuracy of different models for the target variable 'weight'.

Model	FS1	FS2	FS3
MLP	15.555	20.000	17.777
K-NN	22.222	20.000	6.666
AdaBoostM1	17.777	12.222	12.222
RF	20.000	20.000	14.444
Clustering-based	**22.222**	**26.666**	**18.888**

This table shows the recognition accuracy of selected machine learning models on the generated gait data for the target variable 'weight'. FS1 contains the original attributes of the data. FS2 contains the extracted features by the deep neural network of stacked sparse autoencoders as in Fig. 4. FS3 contains the selected features by the CFS algorithm as in Fig. 4. The number of hidden neurons of Encoder 1 is 144 and the number of hidden neurons of Encoder 2 is 48.

Table 7
Other performance measures on chosen feature set for the target variable 'weight' (Clustering-based (FS2) Vs K-NN (FS1)).

Class	ACC (%)	PPV	TPR	F1
1	78.89 / 70.00	0.44 / 0.15	0.22 / 0.11	0.30 / 0.13
2	72.22 / 75.56	0.24 / 0.25	0.25 / 0.19	0.24 / 0.21
3	72.22 / 73.33	0.33 / 0.27	0.47 / 0.24	0.39 / 0.25
4	73.33 / 71.11	0.18 / 0.24	0.12 / 0.24	0.14 / 0.24
5	80.00 / 84.44	0.15 / 0.27	0.22 / 0.33	0.18 / 0.30
6	81.11 / 81.11	0.25 / 0.29	0.27 / 0.36	0.26 / 0.32
7	95.56 / 88.89	0.25 / 0.0	0.50 / 0.0	0.33 / 0.0

Given the multi-class nature of the target variable 'Weight', Table 7 presents additional performance metrics, specifically, Accuracy (ACC), Precision (PPV), Sensitivity (TPR), and F1 Score, for the two top-performing models and their respective feature sets as displayed in Table 6. In each column, the value on the left signifies the performance measure for the clustering-based approach, while the value on the right represents the performance of the K-NN model. For instance, in the case of class 7, the accuracy for the clustering-based method is 95.56, whereas it is 88.89 for the K-NN model.

Table 8
Recognition accuracy of different models for the target variable 'height'.

Model	FS1	FS2	FS3
MLP	85.555	86.666	87.777
K-NN	87.777	83.333	84.444
AdaBoostM1	86.666	90.000	86.666
RF	91.111	85.555	**91.111**
Clustering-based	**92.222**	**92.222**	90.000

This table shows the recognition accuracy of selected machine learning models on the generated gait data for the target variable 'height'. The FS1 contains the original attributes of the data. The FS2 contains the extracted features by the deep neural network of stacked sparse autoencoders as in Fig. 4. The FS3 contains the selected features by the CFS algorithm as in Fig. 4. The number of hidden neurons of Encoder 1 is 48 and the number of hidden neurons of Encoder 2 is 80.

a high degree of effectiveness in correctly identifying instances belonging to these specific classes within the dataset. This high TPR suggests a notably reduced incidence of false negatives, indicating that the classifier seldom fails to identify instances of these classes when they are indeed present.

5.4. Gait-based height recognition

As depicted in Table 1, the target variable 'height' has two classes where the first class has 8 instances and the second class has 82 instances. As such the baseline accuracy is 91.111%. As can be seen in Table 8, all models achieve a recognition accuracy less than the baseline accuracy except the clustering-based model where it achieves a recognition accuracy of 92.222% on FS1 and FS2. In general, the FS1 gives the best recognition accuracy; however, FS2 and FS3 give a comparable or better performance.

Table 9
Other performance measures on chosen feature set for the target variable 'height'.

Model	TPR	SPC	FPR	FNR
MLP (FS3)	0.9146	0.5000	0.5000	0.0854
K-NN (FS1)	0.9024	0.6250	0.3750	0.0976
AdaBoostM1 (FS2)	0.8902	0.6250	0.3750	0.1098
RF (FS1)	0.9268	0.7500	0.2500	0.0732
Clustering-based (FS1)	0.9390	0.7500	0.2500	0.0610

This table presents a comparative analysis of performance metrics extracted from the confusion matrix, specifically, sensitivity (TPR), specificity (SPC), False Positive Rate (FPR), and False Negative Rate (FNR), for various machine learning models utilizing their most effective feature sets.

Table 9 provides an in-depth analysis of the machine learning algorithms when applied to their optimal feature set, as initially presented in Table 8. The results reveal a significant improvement in both TPR and SPC. This implies that the clustering-based algorithm stands out as the only method capable of effectively identifying instances falling within the positive range (*i.e.*, 148.3 and above) as well as those within the negative range (*i.e.*, below 148.3). The low FPR signifies that the model is less likely to incorrectly classify an instance as positive when, in fact, it falls within the negative range. This combination of high TPR, high SPC, and low FPR underscores the model's precision and accuracy in distinguishing between positive and negative instances.

6. Potential applications

To date, gait biometrics technologies have received only a limited exposure in a commercial setting, yet a number of studies have revealed the potential for the use of gait biometrics in industry (*e.g.*, security and forensic) (Mason et al., 2016). However, such approaches still a long way from being commercially applicable (Mason et al., 2016).

Although image-based approaches have received the most research interest in the literature, the first actual commercial incorporation of gait biometrics has come from the wearable and floor sensor industries (Mason et al., 2016). This is made possible by the proliferation of wearable programmable devices which are acceptable by consenting owners (Mason et al., 2016). This motivates several companies to develop gait biometric-based applications and devices. For example, Plantiga (2023) is developing a shoe-based tracking system. It is envisaged that sensor-based approaches will gain more popularity in the near future because they provide more possibilities in terms of integration with other forms of biometrics for human identification purposes (Mason et al., 2016). Thus, we believe that the proposed approach is an alternative solution to image-based approaches and has an opportunity to be deployed in gait biometric-based devices and applications.

From forensic perspective, gait biometrics can be used as forensic tools. For example, it has been reported in engadget (2018) that "investigators have used data from iOS' built-in Health app as evidence in the investigation of a rape and murder case. Police cracked the suspect's phone with the help of an unnamed Munich company and discovered Health data that corresponded with his reported activity the day of the crimes, which included dragging the victim down a river embankment and climbing back up. The suspect's Health app appeared to have registered this last action as two instances of stair climbing, and an officer obtained similar results when replicating the accused's movements.

The Health info (which also included his overall activity levels) was only part of the information investigators collected. They only had incomplete public surveillance video and geodata, but they noticed that his phone contacted a cell tower near the crime scene at a time consistent with video footage, and that there was an unusually long period of inactivity before it had to contact a new cell site. The victim's Nokia phone also sent its last location data shortly after the crime is believed to have taken place."

The IMU sensors found in modern mobile phones are a typically combination of several different types of sensors, including accelerometer, gyroscope, magnetometer and barometer. These sensors work together within the IMU system of a mobile phone to provide data related to motion, orientation, and spatial positioning. The data from these sensors have widely been used by various applications, including gaming, virtual reality, augmented reality, fitness tracking, navigation, and other motion-based interactions. The IMU sensors, we believe, when applied to gait analysis, have the potential to be utilized in various identification tasks due to the mainly following three reasons.

Firstly, IMU-based gait analysis is non-intrusive and unobtrusive, as it requires only small wearable sensors attached to the body. This characteristic makes it suitable for identification tasks where privacy, comfort, and convenience are important considerations. Unlike other biometric methods such as fingerprinting or iris scanning, gait analysis using IMUs can be performed without direct contact or specialized equipment.

Secondly, IMUs can provide continuous monitoring of gait parameters over an extended period. This capability allows for the creation of comprehensive gait profiles that capture subtle variations and changes in an individual's gait pattern over time. Such detailed profiles enhance the accuracy and reliability of identification tasks that rely on gait analysis.

Finally, IMUs can be combined with other sensing technologies, such as accelerometers, gyroscopes, and magnetometers, to provide more comprehensive and accurate gait analysis. By fusing data from multiple sensors, it is possible to capture additional gait features and improve the overall identification performance. For example, combining IMU data with video-based systems can offer both quantitative and qualitative gait information. Gait analysis using IMU technology finds applications in various identification tasks. These include biometric identification for security purposes (*e.g.*, access control, surveillance), healthcare applications (*e.g.*, monitoring and assessing patients' mobility and rehabilitation progress), forensic analysis, and even personalized user recognition in interactive systems or virtual reality environments.

While IMU-based gait analysis using machine learning offers several advantages, it also has some limitations or weaknesses that should be taken into consideration. Firstly, gait patterns can vary significantly among individuals due to factors such as age, health conditions, footwear, and walking surfaces. Machine learning models trained on one population may not generalize well to other populations, which can limit the effectiveness of IMU-based gait analysis across diverse demographics. Adapting the models to accommodate variations in gait patterns becomes essential for improving accuracy and generalization. Secondly, IMU sensors are susceptible to environmental conditions and external interferences. Vibrations, magnetic fields, and other external factors can introduce noise or distort the measurements, impacting the quality of gait data. Robust filtering and preprocessing techniques are required to mitigate the influence of environmental factors and ensure accurate gait analysis. Lastly, while IMU sensors provide valuable data on motion and orientation, the information may be limited compared to other sensor modalities or imaging techniques. IMU-based gait analysis relies primarily on capturing body movements, which may not capture fine-grained details or subtle variations in gait patterns. Incorporating additional sensors or modalities may be necessary to enhance feature representation and improve the accuracy of identification tasks. We believe that addressing these weaknesses through ongoing research and development efforts is essential to unlock the full potential of IMU-based gait analysis using machine learning and ensure its practical applicability across a wide range of identification tasks.

Regarding the offline and online application aspects, we acknowledge that it is essential to clarify the operational aspects of our approach. Our method primarily involves the use of Inertial Measurement Unit (IMU) sensors, which are often integrated into modern mobile devices. The process of extracting latent features from the gait data, as well as the initial training of the machine learning model, can be performed offline. During this stage, the model learns the underlying patterns and relationships from the gait data to create a representation that is effective for subsequent identification tasks.

Once the model is trained and the latent features are obtained, the actual identification tasks can be performed online. The online application involves deploying the trained model on a mobile device or a dedicated system to process real-time gait data and predict the numeric attributes, such as age, height, and weight, of individuals based on their gait patterns. This real-time prediction capability can be invaluable in various contexts, such as forensic analyses, healthcare applications, and user recognition in interactive systems.

7. Concluding remarks and future work

This study highlights the significance of gait analysis as a potential alternative to traditional image-based identification methods. While existing technologies heavily rely on analyzing 2D/3D images captured by surveillance cameras, this study demonstrates the utilization of gait biometrics in combination with deep feature learning and inertial measurement unit (IMU) technologies. By exploring this novel approach theoretically and experimentally, the study reveals compelling results that surpass the accuracy of existing models in individual identification tasks.

Our experimental results undeniably showcase the exceptional performance of the proposed clustering-based model in achieving high identification accuracy. Specifically, in gender identification, the model achieves a recognition accuracy of 75.555% using FS3, outperforming all other examined learning models. Moreover, for age identification, it achieves the highest recognition accuracy of 76.666% using FS1.

The produced representations derived from the clustering-based machine learning model, which captures gait information, exhibit a notable degree of invariance to variations. As a result, these representations demonstrate robustness in identifying individuals, even amidst significant changes occurring in their walking patterns.

Specifically, the use of extracted latent features from a stacked autoencoder in ensemble clustering can be technically suitable and advantageous for several reasons: 1) By extracting latent features via SAE, the high-dimensional input data can be reduced to a lower-dimensional space, which helps in handling computational complexity and improves clustering performance, 2) SAEs are unsupervised learning models that can automatically learn hierarchical and abstract representations of the data. These learned features are often more discriminative and relevant for clustering tasks compared to the original raw data, 3) The learned latent features are likely to capture the underlying structure and patterns in the data while being less affected by noise and minor variations. This robustness can improve the overall performance of ensemble clustering methods, and 4) When using the latent features in an ensemble clustering approach, each base clustering model in the ensemble can be built on different subsets of these features. This diversity enhances the ensemble's ability to capture different aspects of the data distribution and boosts clustering accuracy.

When it comes to weight identification, the proposed clustering-based model consistently demonstrates superior recognition performance when utilizing FS1, FS2, and FS3. Additionally, for height identification, it achieves a recognition accuracy of 92.222% when using FS1 and FS2. We firmly believe that our research, which leverages machine learning approaches and incorporates IMU technology for gait analysis, opens up new possibilities for various identification tasks. Most importantly, it addresses the limitations associated with image quality and individual appearance variations, which are inherent in current image-based methods.

The findings from this study not only contribute to the existing literature on gait analysis but also have significant implications for practical applications. By establishing the potential of IMU-based gait analysis, this research paves the way for advancements in identification tasks across multiple domains. The demonstrated superiority of the proposed

model provides a solid foundation for future research and applications in fields such as security systems, healthcare monitoring, forensic analysis, personalized user recognition, and more.

This study's results underscore the transformative potential of machine learning approaches that leverage IMU technology and prioritize gait analysis information. By overcoming the limitations of image-based methods, this research opens up new possibilities for accurate and reliable identification tasks, marking a substantial advancement in the field of identification technologies.

In the context of designing such machine learning architecture for the precise estimation of human attributes (*e.g.*, age, gender, height, and weight) derived from IMU sensor data, it is imperative to proactively address potential adversarial scenarios. To ensure the model's robustness and fairness, the authors suggest devising a comprehensive plan to mitigate adversarial scenarios. This plan should encompass multiple stages of data preprocessing, involving meticulous data curation to eliminate biases and outliers while anonymizing sensitive information. The augmentation of adversarial data during training enables the model to better withstand potential attacks and perturbations. The plan may also involve rigorous evaluation through fairness metrics, including disparate impact and equal opportunity, to quantify bias in predictions. Bias mitigation techniques, such as re-weighting and adversarial training, are then employed to rectify these biases. Adversarial testing could also be integrated into the evaluation process to assess the model's resilience against targeted attacks. Privacy preservation techniques, such as differential privacy, should be implemented to safeguard sensitive information.

Building upon the findings of this study, there are several promising avenues for future research in the field of gait analysis and identification tasks. Further exploration can be conducted to optimize the proposed clustering-based model by investigating the impact of different feature combinations and fusion techniques on identification accuracy. As discussed, incorporating additional biometric modalities, such as facial recognition or voice recognition, alongside gait analysis could potentially enhance the overall identification performance. Extending the study to larger and more diverse datasets would provide valuable insights into the generalizability and robustness of the proposed model. Moreover, conducting longitudinal studies to investigate the stability and consistency of gait-based identification over time would be beneficial. Extending the ideas above, we also have plan to explore real-time implementation and deployment of the proposed model in practical scenarios, such as security systems or healthcare monitoring, would provide valuable insights into its feasibility and effectiveness in real-world applications.

CRediT authorship contribution statement

K. Taha, **P.D. Yoo**, **Y. Al-Hammadi** & **S. Muhaidat**: Conceptualization, Methodology, Software. **P.D. Yoo**: Data Collection. **K. Taha**, **P.D. Yoo**: Writing original draft preparation. **C.Y. Yeun**: Investigation, Validation. **P.D. Yoo**: Supervision. All: Reviewing and Editing.

Declaration of competing interest

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

Data availability

The link to the dataset is available on page 6.

Acknowledgement

The authors express their sincere appreciation to the DASA machine learning team for their invaluable discussions and feedback. Special gratitude is extended to JH Jeon at JEIOS Ltd for generously providing the dataset. The authors would like to acknowledge the EBTIC, British Telecom's (BT) cyber security team for their insightful and constructive critique of this work.

References

Plantiga. [Online]. Available: http://www.plantiga.com.

engadget. [Online]. Available: https://www.engadget.com/2018/01/11/apple-health-app-data-used-in-rape-investigation/.

Ageng, D., Huang, C.-Y., Cheng, R.-G., 2021. A short-term household load forecasting framework using lstm and data preparation. IEEE Access 9, 911–167 919.

Aha, D.W., Kibler, D., Albert, M.K., 1991. Instance-based learning algorithms. Mach. Learn. 6 (1), 37–66.

Alotaibi, M., Mahmood, A., 2015. Improved gait recognition based on specialized deep convolutional neural networks. In: IEEE Applied Imagery Pattern Recognition Workshop. AIPR. IEEE, pp. 1–7.

Bhanu, B., Han, J., 2010. Model-based human recognition-2d and 3d gait. In: Human Recognition at a Distance in Video. Springer, pp. 65–94.

Bouchrika, I., Goffredo, M., Carter, J., Nixon, M., 2011. On using gait in forensic biometrics. J. Forensic Sci. 56 (4), 882–889.

Boulgouris, N.V., Chi, Z.X., 2007. Gait recognition using Radon transform and linear discriminant analysis. IEEE Trans. Image Process. 16 (3), 731–740.

Boulle, M., 2005. Optimal bin number for equal frequency discretizations in supervized learning. Intell. Data Anal. 9 (2), 175–188.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

David, O.E., Netanyahu, N.S., 2015. Deepsign: deep learning for automatic malware signature generation and classification. In: 2015 International Joint Conference on Neural Networks. IJCNN. IEEE, pp. 1–8.

Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf. Process. 3, e2.

Dougherty, J., Kohavi, R., Sahami, M., et al., 1995. Supervised and unsupervised discretization of continuous features. In: Machine Learning: Proceedings of the Twelfth International Conference, vol. 12, pp. 194–202.

Esser, P., Dawes, H., Collett, J., Feltham, M.G., Howells, K., 2012. Validity and inter-rater reliability of inertial gait measurements in Parkinson's disease: a pilot study. J. Neurosci. Methods 205 (1), 177–181.

Freund, Y., Schapire, R.E., et al., 1996. Experiments with a new boosting algorithm. In: ICML'96, pp. 148–156.

Girbés-Juan, V., Armesto, L., Hernández-Ferrándiz, D., Dols, J.F., Sala, A., 2021. Asynchronous sensor fusion of gps, imu and can-based odometry for heavy-duty vehicles. IEEE Trans. Veh. Technol. 70 (9), 8617–8626.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The weka data mining software: an update. ACM SIGKDD Explor. Newsl. 11 (1), 10–18.

Hall, M.A., 1999. Correlation-Based Feature Selection for Machine Learning.

Han, J., Bhanu, B., 2006. Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. 28 (2), 316–322.

Haykin, S.S., 2008. Neural Networks and Learning Machines. Prentice Hall.

Hu, M., Wang, Y., Zhang, Z., Zhang, D., 2011. Gait-based gender classification using mixed conditional random field. IEEE Trans. Syst. Man Cybern., Part B, Cybern. 41 (5), 1429–1439.

Imoize, A.L., Tofade, S.O., Ughegbe, G.U., Anyasi, F.I., Isabona, J., 2022. Updating analysis of key performance indicators of 4g lte network with the prediction of missing values of critical network parameters based on experimental data from a dense urban environment. Data Brief 42, 108240.

Iwama, H., Muramatsu, D., Makihara, Y., Yagi, Y., 2013. Gait verification system for criminal investigation. IPSJ Trans. Comput. Vis. Appl. 5, 163–175.

Joo, J.-Y., Kim, Y.-K., Park, J.-Y., 2015. Reliability of 3d-inertia measurement unit based shoes in gait analysis. Korean J. Sport Biomech. 25 (1), 123–130.

Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., Chellappa, R., 2004. Identification of humans using gait. IEEE Trans. Image Process. 13 (9), 1163–1173.

Karalas, K., Tsagkatakis, G., Zervakis, M., Tsakalides, P., 2015. Deep learning for multi-label land cover classification. In: SPIE Remote Sensing. In: International Society for Optics, p. 96 430Q.

Khan, M.H., Farid, M.S., Grzegorzek, M., 2021. Vision-based approaches towards person identification using gait. Comput. Sci. Rev. 42, 100432.

Khan, M.H., Farid, M.S., Grzegorzek, M., 2023. A comprehensive study on codebook-based feature fusion for gait recognition. Inf. Fusion 92, 216–230.

Kim, Y.-K., Joo, J.-Y., Jeong, S.-H., Jeon, J.-H., Jung, D.-Y., 2016. Effects of walking speed and age on the directional stride regularity and gait variability in treadmill walking. J. Mech. Sci. Technol. 30 (6), 2899–2906.

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI, vol. 14, pp. 1137–1145.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Le, Q.V., 2013. Building high-level features using large scale unsupervised learning. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 8595–8598.

Lee, L., Grimson, W.E.L., 2002. Gait analysis for recognition and classification. In: Fifth IEEE International Conference on Automatic Face and Gesture Recognition. IEEE, pp. 148–155.

Lee, T.K., Belkhatir, M., Sanei, S., 2014. A comprehensive review of past and present vision-based techniques for gait recognition. Multimed. Tools Appl. 72 (3), 2833–2869.

Lu, J., Tan, Y.-P., 2010. Gait-based human age estimation. IEEE Trans. Inf. Forensics Secur. 5 (4), 761–770.

Lu, J., Wang, G., Moulin, P., 2014. Human identity and gender recognition from gait sequences with arbitrary walking directions. IEEE Trans. Inf. Forensics Secur. 9 (1), 51–61.

Lynnerup, N., Larsen, P.K., 2014. Gait as evidence. IET Biometrics 3 (2), 47–54.

Makihara, Y., Okumura, M., Iwama, H., Yagi, Y., 2011. Gait-based age estimation using a whole-generation gait database. In: International Joint Conference on Biometrics. IJCB. IEEE, pp. 1–6.

Makihara, Y., Mannami, H., Tsuji, A., Hossain, M.A., Sugiura, K., Mori, A., Yagi, Y., 2012. The ou-isir gait database comprising the treadmill dataset. IPSJ Trans. Comput. Vis. Appl. 4, 53–62.

Mason, J.E., Traoré, I., Woungang, I., 2016. Applications of gait biometrics. In: Machine Learning Techniques for Gait Biometric Recognition. Springer, pp. 203–208.

Mayagoitia, R.E., Nene, A.V., Veltink, P.H., 2002. Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. J. Biomech. 35 (4), 537–542.

Ng, A., Ngiam, J., Foo, C.Y., Mai, Y., Suen, C., 2012. Ufldl Tutorial.

Okumura, M., Iwama, H., Makihara, Y., Yagi, Y., 2010. Performance evaluation of vision-based gait recognition using a very large-scale gait database. In: Fourth IEEE International Conference on Biometrics: Theory Applications and Systems. BTAS. IEEE, pp. 1–6.

Portney, L.G., Watkins, M.P., et al., 2000. Foundations of Clinical Research: Applications to Practice, vol. 2. Prentice Hall, Upper Saddle River, NJ.

Rahman, A., Verma, B., 2011. Novel layered clustering-based approach for generating ensemble of classifiers. IEEE Trans. Neural Netw. 22 (5), 781–792.

Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y., 2016. Geinet: view-invariant gait recognition using a convolutional neural network. In: International Conference on Biometrics. ICB. IEEE, pp. 1–8.

Wang, C., Zhang, J., Pu, J., Yuan, X., Wang, L., 2010. Chrono-gait image: a novel temporal template for gait recognition. In: European Conference on Computer Vision. Springer, pp. 257–270.

Wang, J., Wang, Z., Gao, C., Sang, N., Huang, R., 2016. Deeplist: learning deep features with adaptive listwise constraint for person re-identification. In: IEEE Transactions on Circuits and Systems for Video Technology.

Yan, C., Zhang, B., Coenen, F., 2015. Multi-attributes gait identification by convolutional neural networks. In: 8th International Congress on Image and Signal Processing. CISP. IEEE, pp. 642–647.

Yang, Y., Webb, G., 2001. Proportional k-interval discretization for naive-Bayes classifiers. In: European Conference on Machine Learning. Springer, pp. 564–575.

Yang, Y., Webb, G.I., 2002. A comparative study of discretization methods for naive-Bayes classifiers. In: Proceedings of PKAW, vol. 2002. Citeseer.

Yu, S., Tan, D., Tan, T., 2006. Modelling the effect of view angle variation on appearance-based gait recognition. In: Asian Conference on Computer Vision. Springer, pp. 807–816.