Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2023) 31: 448 – 461
© TÜBİTAK
doi:10.55730/1300-0632.3994

An exploratory study on the effect of applying various artificial neural networks to the classification of lower limb injury

Rachel YUN[1,*], **May SALAMA**[2], **Lamiaa ELREFAEI**[2]

[1]Department of Computer Science, The Nueva School, San Mateo, USA

[2]Electrical Engineering Department, Shoubra Faculty of Engineering, Benha University, Cairo, Egypt

Received: 26.09.2022 • **Accepted/Published Online:** 23.02.2023 • **Final Version:** 23.03.2023

Abstract: This paper explores the application of a deep neural network (DNN) framework to human gait analysis for injury classification. The paper aims to identify whether a subject is healthy or has an injury of the ankle, knee, hip, or heel solely based on ground reaction force plate measurements. We consider how three DNNs—the multi-layer perceptron (MLP), fully convolutional network (FCN), and residual network (ResNet)—can be applied to gait analysis when the number of trainable network parameters far exceeds the number of training samples, and benchmark their performance in this context against that of shallow neural networks. The DNN architectures outperformed unsupervised clustering models—self-organizing map and k-means clustering—by a large margin. When tested against support vector machines, which is considered the state-of-the-art approach for supervised gait classification, DNNs performed equal or better despite the propensity for overfitting. We did not find evidence that applying data augmentation to the overfitting problem via timeGAN, a generative adversarial network for time-series data generation, leads to improved classification accuracy. While DNNs are hypothesized to have intrinsic feature extraction capability, the results suggest an advantage to implementing an explicit feature extraction process as a frontend for future applications of deep neural networks: specifically, principal component analysis (PCA) preprocessing improved classification accuracy robustness in all models. In the absence of an explicit feature extraction layer, the use of dropouts in convolutional neural networks caused significant performance degradation; by contrast, the use of a PCA frontend and dropouts synergistically achieved the highest test accuracy among the DNNs studied.

Key words: Gait classification, deep neural networks, feature extraction

1. Introduction

Lower body injuries are common in clinical and athletic settings. This work focuses on the classification of lower-body injuries based on human gait analysis. Gait analysis is a broad yet fruitful field to develop machine learning methods as its many applications extend beyond medicine to authentication, security, and sports [37]. Gait has been used to recognize an individual when it is difficult to obtain other biometric characteristics, such as fingerprints and facial or iris-related information [11, 23, 26]. Gait analysis in athletics is adept at determining biomechanical faults and imbalances that may cause injuries [9]. In the medical field, gait classification is utilized for Clinical Gait Analysis [36], which comprises not only disease detection (e.g., Parkinson's disease, multiple sclerosis [8, 27]) but also assessment tools for evaluating disease progression, patient medication adherence, and the effectiveness of patient training and rehabilitation programs [1].

*Correspondence: rachel.k.yun@gmail.com

Broadly, there are two approaches to the gait data classification problem: statistical classification and machine learning (ML) [4]. Statistical methods compute summary statistics of the input data and apply parametric tests (e.g., Student's t-test, analysis of variance (ANOVA)) to the preprocessed data [4]. They are most frequently employed when the amount of data available is small. ML methods are used for gait analysis when the dataset available is large; both supervised and unsupervised (clustering) methods have been applied for gait classification [4, 2]. Analogous to statistical methods, a preprocessing step may also be introduced in ML methods: feature selection or extraction. Feature selection methods attempt to reduce the set of input features to a "best" subset by an algorithm rather than a human expert, typically by feature scoring. In contrast, feature extraction transforms the original feature data to another set of values in a lower-dimensional space. Principal component analysis (PCA) has been examined along with a combination of SVM and independent component analysis (ICA) as a supervised-unsupervised feature projection [2].

The basic biomechanical variables in gait analysis include anthropometric parameters (age, height), electromyography; spatiotemporal parameters (statistics deemed by an expert to be of clinical relevance), kinematics (joint angle changes over space and time [35]), and kinetics (force data measured by patient movement on a ground force plate). Gait data may be collected by nonwearable sensors, consisting of video motion capture systems for capturing human movement and floor sensors (GRF data); or wearable sensors consisting of inertial motion sensors (accelerometers, gyroscopes) and insole pressure/force sensors [1]. In [13], kinematic and kinetic data are analyzed to predict the identity of an individual. The classification methods applied were linear SVM, linear SGD, MLP, and convolutional neural network architectures (CNN-A and CNN-C3) [13].

Force plate data, or ground reaction force (GRF) data, is a common form of kinetic data collected for gait analysis. Force plate data is a measure of plantar foot pressure typically evaluated over human motion on a force plate and contains three components: point of application, magnitude, and line of action. GRF data can be studied to recognize balance or lack thereof, which is often a sign of pathological disorder or injury. Force plate feedback can be analyzed to make conclusive inferences about abnormalities in balance and gait, which can be used to identify biomechanical imbalances or injuries [1]. In [18], GRF data from the Santos and Duarte dataset [19] was utilized to study the problem of fall-risk prediction. The recurrent neural network, long short time memory (LSTM), one dimensional convolutional neural network (1D-CNN), and a proposed one-one-one deep neural network were studied. GRF measurements were also used in [14] for feature extraction and injury classification. The study found that feature extraction by performing principal component analysis on the raw GRF waveforms was superior to using summarized statistics (e.g., cadence, gait velocity, stance time, and step length). The authors presented results for linear SVM and radial basis function SVM for classification and stated that from previous experimentation, SVM outperforms multilayer perceptron (MLP) and k-nearest neighbors [14].

More recently, a large dataset that includes gait GRF measurements of both healthy subjects and subjects with various lower limb pathologies has been made public [20]. The GaitRec dataset has been applied towards the design of a low-cost alternative to clinician-based gait analysis [16], early detection of gait patterns in populations with high fall risk [17], and explainable ML models for gait classification [3]. Classification tasks for [3] included binary classification between healthy controls and all gait disorders; binary classification between healthy controls and each gait disorder; multiclass classification between healthy controls and all gait disorders; and multiclass classification between all gait disorders. Classification methods consisted of three representative ML approaches—linear SVM, MLP, and CNN—which were compared in terms of prediction accuracy.

While an array of machine learning algorithms including tree-based classifiers, the Bayes classifier, CNNs, and MLPs have been investigated for supervised learning in gait analysis [1–11], the SVM is consistently cited as the most widely used machine learning algorithm for gait analysis [3, 4, 12] and achieves state-of-the-art classification performance in classification [13, 14]. Less studied for gait analysis are deep neural network architectures. Deep neural networks have been theorized to hold the promise of built-in feature extraction from raw, complex, and high-dimensional data [4]. However, they require much more data to train than SVMs and are prone to overfitting. The aim of this paper is threefold:

1. to study how a generalized deep neural network framework [15] can be applied to the problem of injury classification by gait analysis, in which the number of network trainable parameters greatly exceeds the number of training data samples;

2. to benchmark the performance of DNNs against more simple classifiers as well as the state-of-the-art linear and radial basis function SVMs;

3. to test the hypothesis that DNNs incorporate built-in feature extraction capability in the context of gait analysis, by implementing a feature summarization preprocessing layer before the deep neural network architecture and then evaluating performance with and without the preprocessing layer.

This paper is organized as follows:

1. Subsections 2.1–2.3 cover the three deep neural network structures and their specific application to injury classification by gait,

2. Section 2.4 presents the methodology for data preprocessing, model training and testing;

3. And finally, Sections 3–5 analyze the results and discuss areas for future work.

2. Methodology

This study used the large-scale GaitRec dataset [20], consisting of ground reaction force data gathered during 75,732 bilateral walking trials from 2084 patients with various lower body impairments and 211 healthy controls. The data was collected from patients at a rehabilitation center recovering from joint replacement, fractures, ligament ruptures and related ailments. Five analog force plate signals—for vertical GRF, anterior-posterior GRF, medio-lateral GRF, anterior-posterior center of pressure (COP), and medio-lateral COP—were recorded for walking participants and digitized. The input data consisted of 33,939 data samples, each data sample corresponding to five vectors of 101 time series values, amplitude-normalized by body weight and time-normalized to 100% stance. The dataset was manually labeled by an experienced physical therapist based on the medical diagnosis of each patient as one of five classes: healthy control (HC), calcaneus injury (C), ankle injury (A), hip injury (H), or knee injury (K) [20]. The data collected was bilateral, and an injury may be present in either leg; for purposes of this study, we perform the analysis on the right leg dataset. Data for diseased patients with both legs affected were not included. The number of samples in each class were: 7755 (HC), 4978 (C), 6048 (H), 9132 (K), and 6026 (A).

To explore applying DNNs to gait analysis, we used the generalized deep neural network framework for time series data proposed by Wang et al. [15] as our starting point, but we adapted the approach specifically to the gait analysis problem. We chose this framework because it was benchmarked against a variety of other state-of-the-art approaches specifically in time series classification tasks and found to perform well [15]. The framework consists of three baseline DNN model structures for real-world applications: the multilayer perceptron (MLP), the fully convolutional network (FCN), and the residual network (ResNet).

We compared the DNNs with two unsupervised learning approaches: the self-organizing map (SOM) and k-means clustering [2, 14]. As k-means and SOM perform unsupervised clustering and not classification, we performed classification after clustering by assigning to each cluster its most frequent class. SVMs are cited as offering state-of-the-art classification performance in gait analysis [13, 14], so we evaluated both linear and nonlinear boundary SVMs on the GaitRec dataset. The radial basis function (RBF) was used for the nonlinear SVM.

Medical data such as GaitRec are often collected in clinical settings, making it challenging to gather large amounts of data due to high acquisition cost and patient data privacy concerns. GaitRec is the largest dataset of its kind to date, and yet when the deep NN architectures are applied to the data, all three methods have more trainable model parameters than there are training data samples in the GaitRec dataset; in the case of the MLP and the ResNet, the ratio exceeded 10-to-1. We therefore implemented one or more dropout layers and L1 kernel regularization in all three methods and additionally batch normalization in the FCN and ResNet to address overfitting and improve model generalization (Figure 1).

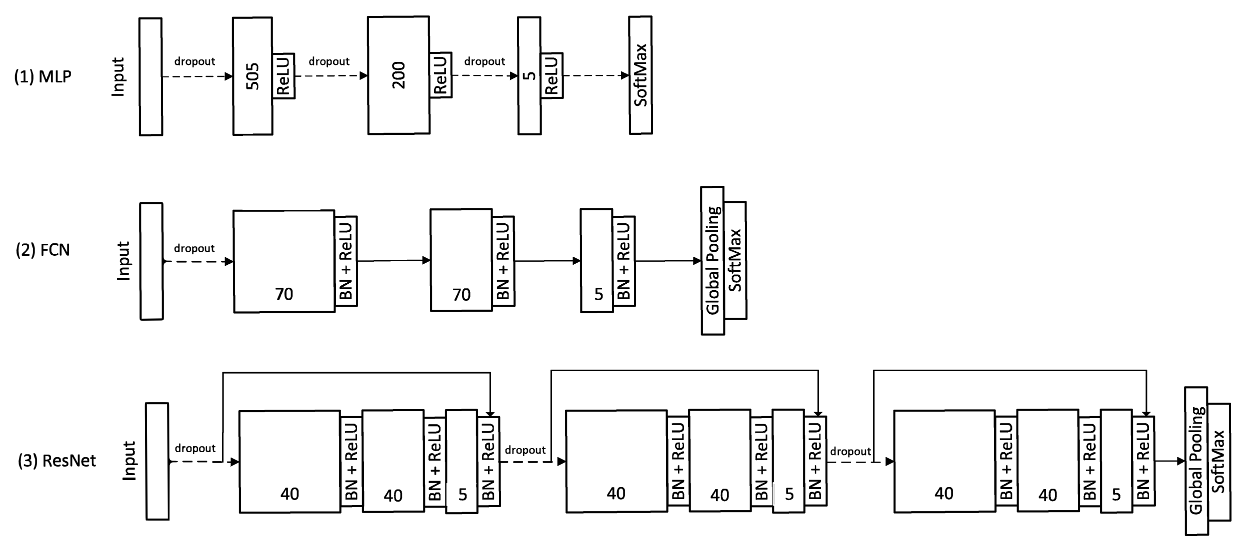


Figure 1. Network structures of three tested models. Dash lines indicate operation of the dropout function; branched arrows show skip connections [15].

Alternatively, data augmentation has been proposed as a means to directly address a mismatch between the number of trainable parameters and available training samples. Uchida et al. surveyed a variety of conventional augmentation methods used in conjunction with NNs for time-series classification and found that, unlike for images, data augmentation can worsen performance nearly as often as it improves it [34]. More recently, TimeGAN, a general adversarial network capable of generating synthetic time-series that preserve the data correlations across time of the real sequences, has reported best in-class performance for time-series data [33]. We have implemented TimeGAN to study the effect of data augmentation on gait GRF data.

2.1. Multilayer perceptron (MLP)

The multilayer perceptron model consisted of three fully connected layers with 505, 200, 5 neurons and ReLU activation, followed by a softmax layer. Dropouts were implemented between successive layers; the dropouts at the three layers were the same. The following dropout values were tested: 0.1, 0.2, and 0.5. A basic layer block is defined as [15]:

$$\begin{aligned} x' &= f_{dropout,p}(x) \\ y &= Wx' + b \\ h &= ReLU(y) \end{aligned} \tag{1}$$

where x is the original input, y is the output, h is the hypothesis, and W and b represent the weights and biases, respectively.

2.2. Fully convolutional network (FCN)

The fully convolutional network comprised three 1-dimensional convolutional layers with ReLU activation and batch normalization (BN) between each layer, followed by global average pooling and softmax layers. Because the gait data were in the form of ground force time-series and not images or video, a multichannel 1-dimensional convolutional kernel was used. A dropout layer with a dropout value between 0.1, 0.2, or 0.5 was optionally implemented at the input to the first convolutional layer. A stride length of 2 was used in the first layer; in the other layers, the stride length was 1. The basic layer block is defined as [15]:

$$\begin{aligned} y &= W \otimes x + b \\ y' &= BN(y) \\ h &= ReLU(y') \end{aligned} \tag{2}$$

where y represents the output when the result of the convolutional block(s) is fed into the $ReLU$ function and the other variables are as defined in (1).

2.3. Residual network (ResNet)

The residual network consisted of 9 layers formed by three repetitions of three stacked residual blocks, with each residual block comprising the FCN described above. A dropout layer with a dropout value of 0.1, 0.2, or 0.5 was optionally implemented at the output of the first convolutional layer. Skip connections were installed in each block to allow the backpropagation signal to reach from the output to the input layers and to minimize training error. The basic layer block is defined as [15]:

$$\begin{aligned} h1 &= Block_{k1}(x) \\ h2 &= Block_{k2}(h1) \\ h3 &= Block_{k3}(h2) \\ y &= h3 + x \\ h &= ReLU(y). \end{aligned} \tag{3}$$

The final ResNet stacks three residual blocks followed by a global pooling layer and a softmax layer. $\mathrm{Block_k}$ denotes the convolutional block with the number of filters k. Lastly, the model applied batch normalization and global average pooling, consistent with the framework in [15].

We also wanted to evaluate a very deep NN against the MLP, FCN, and ResNet for the gait analysis problem. A popular model, such as ResNet50 or ReNet101 [31], takes 2-dimensional data as input and would require the encoding of the gait 1-dimensional time-series into images. The encoding can be done, for example, by using the Gramian Angular Field [32], but this approach would not leverage a strength of ResNet50, that of pretrained weights acquired by transfer learning. More importantly, the Gramian Angular Field captures second order statistics, so performing the encoding may lead to a loss of data information; this is a potential risk of any encoding method. Instead, we decided to take advantage of the fact that the ResNet architecture described above is specifically designed for time-series data and can be extended to more layers by simply concatenating additional residual blocks. To form the very deep NN, then, a total of 15 residual blocks were stacked in tandem to create a 45-layer deep ResNet (”ResNet-45”). This very deep NN architecture can then be complemented with the data augmentation method described above for training.

2.4. Data preprocessing and feature extraction

DNNs are hypothesized to perform built-in feature extraction [4]. We test how well this inherent feature extraction works in practice by comparing the performance of the DNNs with and without a feature extraction frontend layer: to the extent that DNNs perform inherent extraction well, adding a feature extraction layer will provide little benefit to the performance of DNNs.

Feature extraction in gait analysis has traditionally meant that attributes are selected by a human expert based on experience and what the individual believes to be the appropriate level of summarization. These higher-level attributes may be “summary statistics (e.g., mean, variance, correlation, and range) or a parametrization (discrete variables and peak amplitudes) involving measures on a single biomechanical gait data” [2]. Summary statistics and parametrization create smaller datasets that are easier to comprehend; however, as a large part of the data is thrown away, potentially meaningful information may be lost. Rather than selecting a small number of the input features, feature extraction may instead transform the original feature data to another set of values in a lower-dimensional space via self-organizing maps (SOMs), principal component analysis (PCA), etc. [2].

In this paper, we use PCA as the feature extraction approach. The five GRF waveforms were compressed separately to 40 PCA components each, 40 being the sufficient number of parameters to cover roughly 99.995% of data variance.

2.5. Training

We performed stratified five-fold cross-validation on all of the models tested; for each fold partition, 80% of the dataset was used for training and 20% was used for testing. As the GaitRec dataset consisted of multiple walking trials for each subject, the partitioning of samples into a training set and a testing set was performed subject-wise.

A grid search was performed to determine the hyperparameters of each model. This included the best cost hyperparameter and the best cost and gamma hyperparameters for the linear and nonlinear SVMs, respectively, the number of layers and the number of neurons in each layer in the MLP, and the number of filters and the filter size in each layer in the ResNet and the FCN, and the dropout value and L1 kernal regularization factor for all of the DNNs.

To train the TimeGAN model for generating synthetic GRF data, we used the default hyperparameters of [34] (module = 'gru', hidden dimensions = 24, number of layers = 3, number of iterations = 10,000, and batch size = 128). The amount of synthetic data generated was set so as to increase the number of training samples threefold, so that the ratio of synthetic data to real data during training with data augmentation was 2-to-1.

The number of parameters for the MLP, FCN, and ResNet models with PCA were 81,000, 35,000, and 317,000, respectively. Without PCA, the number of parameters for MLP, FCN, and ResNet models was significantly greater at 358,000, 118,000, and 380,000, respectively. The very deep ResNet-45 NN was investigated with PCA only and had 1,585,000 parameters.

The model optimizer was set to Adam, a stochastic gradient descent optimizer, and the loss was the categorical cross entropy function.

All the ML models studied were implemented using the TensorFlow/Keras framework with the additional Python libraries numpy, pandas, and scikit.learn. Model training was performed with Google Colab Pro.

3. Results

All the unsupervised and supervised models were evaluated by the metric of test classification accuracy. Figure 2 shows the accuracies of the models tested, with and without a PCA frontend layer. As a baseline for comparison, we included the zeroR classifier, which simply predicts the majority category and achieved a classification accuracy of 29.0%. The tested implementations, in order of best to worst performance, were as follows: 1) MLP with PCA, 2) MLP without PCA, 3) FCN without PCA, 4) ResNet with PCA, 5) RBF SVM with PCA, 6) RBF SVM without PCA, 7) ResNet without PCA, 8) FCN with PCA, 9) linear SVM with PCA, and 10) linear SVM without PCA. While CNNs are typically used to classify image data, 1D-CNNs are effective for the classification of gait ground force time-series, as seen in the performance of the FCN and ResNet NNs relative to that of the benchmark SVMs.

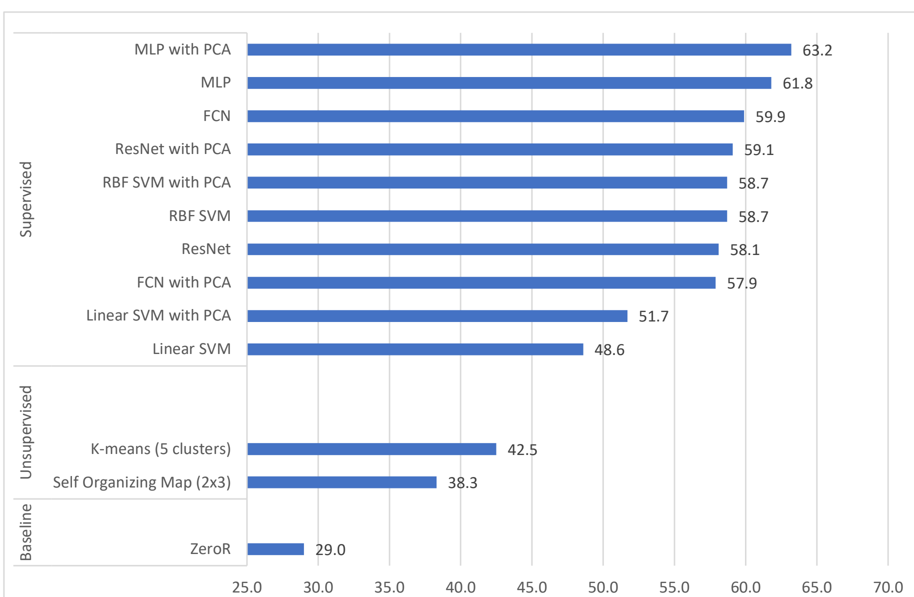


Figure 2. Test classification accuracy of all models ordered from best top of chart to worst (bottom of chart).

In addition to accuracy, the precision, recall, and F1 scores for the three NNs without and without a PCA frontend layer and for different dropout values are shown in Table 1. The scores for the three metrics are similar in each test case, an indication that the GaitRec dataset is fairly balanced.

Table 1. Precision, recall, and F1 scores for the MLP, ResNet, and FCN models.

			MLP			FCN			ResNet	
PCA	Dropouts	Precision	Recall	F1 score	Precision	Recall	F1 Score	Precision	Recall	F1 score
Yes	None	0.638	0.624	0.624	0.590	0.566	0.572	0.594	0.562	0.568
	0.5	0.654	0.616	0.626	0.594	0.536	0.546	0.590	0.590	0.584
No	None	0.618	0.614	0.616	0.624	0.586	0.578	0.586	0.578	0.574
	0.5	0.574	0.492	0.496	0.440	0.354	0.312	0.208	0.258	0.158

Figure 3 examines the effect of the interaction between the dropout values and the inclusion or exclusion of PCA feature extraction on classification accuracy in the three DNNs.

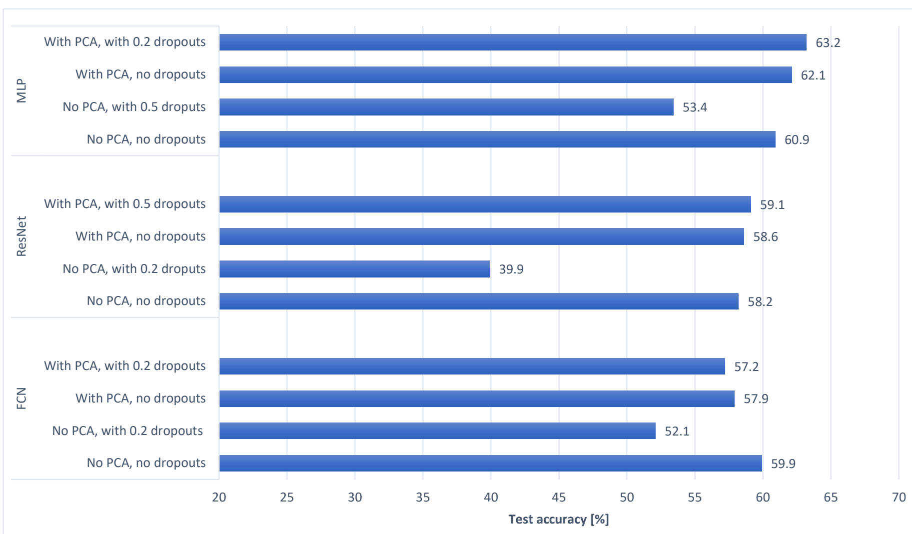


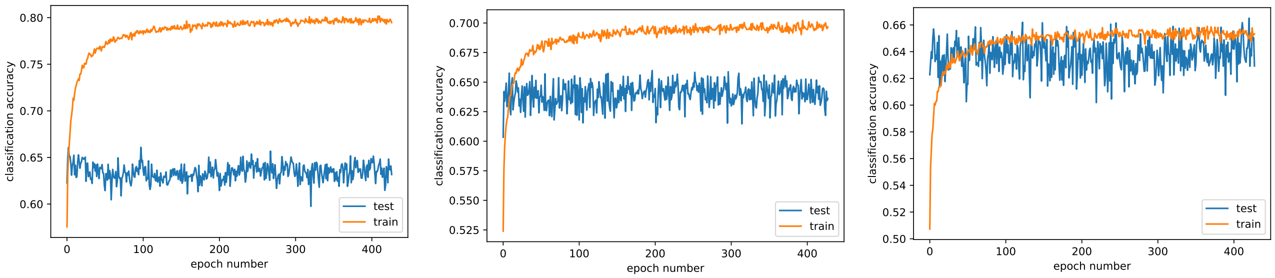
Figure 3. Classification accuracy for different approaches of PCA preprocessing and dropouts in the DNN models.

To fine tune hyperparameters for addressing model overfitting, the training and test accuracies were compared over the training epochs. This is illustrated in Figure 4 for one sample training k-fold of the MLP NN with PCA preprocessing tested for dropout values of 0.2, 0.3, and 0.35.

Table 2 compares the accuracy rates reported from previous gait classification studies for a variety of classification problems and ML techniques which were applied.

Table 3 shows the results from supplementing actual training data with synthetic GRF data generated by the TimeGAN framework to train the MLP, ResNet, and FCN networks.

Lastly, Table 4 compares the classification performance of the 9-layer ResNet against the very deep 45-layer ResNet, with and without training data augmentation.



dropout value = 0.2   dropout value = 0.3   dropout value = 0.35

Figure 4. Training and test accuracy versus training epochs for MLP with PCA preprocessing, L1 regularization factor $= 10^{-4}$ (from one sample fold in 5-fold cross validation).

4. Discussion

Of other relevant papers in the field, only [14] and [3] addressed a problem similar to ours, with other authors undertaking a simpler classification problem (binary classification of healthy vs diseased subjects [16, 17]); a different classification problem (identifying individual subjects [13]); or both a different classification problem and on a different data set [18] (Table 2). The work in [14] addressed the multiclass classification task by applying linear and RBF SVM approaches, which were comparable to our results but performed worse than many of our higher-performing NNs. The work in [3] addressed a similar but slightly simpler (4-class) classification problem, opting to omit the class for injuries of the calcaneus. As with [14], the models in [3] performed slightly worse than ours, producing average accuracy rates in the 40s and 50s for their MLP, SVM, and CNN models.

Consistent with the observation of [14] that SVM outperformed k-nearest neighbor, our results show that, as a class, unsupervised learning methods were uncompetitive with supervised learning methods for time series gait classification, performing only marginally better than the ZeroR baseline (Figure 2). However, among the supervised learning methods, we found that unlike [14], where linear SVM slightly outperformed RBF SVM, we observed that RBF SVM significantly outperforms linear SVM (Figure 2). Since the nature of gait input—digitized ground force waveforms—makes it unlikely that the data points would be linearly separable, it is not an unexpected result that the RBF SVM, with its ability to form nonlinear decision boundaries, would have a performance advantage. Although SVMs are the most commonly used ML method for gait data analysis and have yielded the best performance in [14] and [13], we found that the deep neural networks outperformed SVM to varying degrees and the MLP in particular offered the highest performance in gait classification. In particular, we found that the ResNet architecture, which is essentially three FCNs in tandem and the deepest NN studied, had similar performance as the FCN (Figure 2). This is in contrast to [15], which suggested that "the FCN achieves premium performance to other state-of-the-art approaches."

A key issue in the application of deep neural networks was susceptibility to overfitting. In this study, the MLP with no PCA feature extraction layer had over 358,000 trainable parameters and the 9-layer ResNet model had over 380,000 trainable parameters; yet both were trained on fewer than 32,000 data samples. Initially, without taking any measures, a performance gap between the training accuracy and test accuracy of 30%–35% was observed. The results for the other NNs are similar: training accuracy approaches 90%–99% in most cases, and yet the test accuracies ranged from 52% to 65%.

We examine the effects of data augmentation on overfitting. Looking at Table 3, augmenting the training data with TimeGAN synthetic waveforms did not appear to have a positive impact: classification performance

Table 2. Tabulated results for studies of relevance, detailing author, data type, dataset, approach, classification problem studied and performance.

Study	Data type	Dataset	Methods studied	Classification problem	Accuracy rates (%)
Slijepcevic et al. [14]	GRF	Australian Workers' Compensation Board (AUVA)	Linear SVM with PCA RBF SVM with PCA	Healthy/ calcaneous/ ankle/ hip/ knee injury	54.3% 51.2%
Horst et al. [13]	GRF	Gutenberg dataset	Linear SVM Linear Stochastic Gradient Descent Fully-connected MLP Deep CNN	Individual subjects	100% 95.4% 98.8% 99.2%
Boompelli et al. [16]	GRF	GaitRec Dataset	1D-CNN	Fractured/ healed patients	88.7%
Iber et al. [17]	GRF	GaitRec Dataset	Linear SVM	Healthy/ pathological gait	91%
Slijepcevic et al. [3]	GRF	GaitRec Dataset	CNN SVM MLP	Healthy/ hip/ knee/ ankle injury	~53% ~53% ~49%
Savadkoohi et al. [18]	GRF	Santos and Duarte dataset	1D-CNN + LSTM + Deep Neural Network	Low/ medium/ high fall risk	99.9%
Jung et al. [35]	Marker data + joint angle feature extraction	Moissnet / Horst / Jung Datasets	k-NN CART Random forest	Gait phase	82% 87% 92%
Chen et al. [36]	Video data	Video data from 7 subjects	CNN SVM KNN LSTM	normal vs. pelvic-obliquity vs knee-hyperextension gait	87.6% 94.9% 94.0% 83.6%

with data augmentation was slightly worse than without augmentation for every test case, whether training is performed directly using the time-series samples or after transformation of the GRF waveforms into PCA components. This would indicate that the TimeGAN framework had difficulty modeling the gait waveforms and the synthesized data were capturing similar, but slightly inaccurate or incomplete, statistics of the actual gait GRF time-series. One can further infer that generating more synthetic data is unlikely to improve training efficiency.

Referring to Tables 3 and 4, the performance of a very deep NN, the 45-layer ResNet, was no better than that of the three DNNs studied. This appeared to be the case irrespective of the dropout value and whether data augmentation was applied or not: the performance of the 45-layer ResNet was similar to that of the 9-layer ResNet. This was likely due to a performance plateau due to the small training sample size relative to the number of trainable parameters that was already reached by the 9-layer ResNet; adding additional layers would not address the issue. Data augmentation can potentially tackle this problem. However, the performance of the 45-layer ResNet was slightly worse than the 9-layer ResNet with data augmentation, suggesting any benefit from augmentation for reducing the mismatch between the number of training samples to number of network parameters is offset by the imperfect statistics introduced by the synthetic data, which skewed the classification decision boundaries.

Table 3. Test classification accuracy of ResNet, MLP, and FCN models with and without TimeGAN data augmentation.

ML model	PCA	Dropout	Without data augmentation	With TimeGAN
FCN	Yes	No	57.9	56.4
	No	No	59.9	59.3
9-layer ResNet	Yes	No	58.6	57.7
	No	No	58.1	57.4
MLP	Yes	No	62.1	60.2
	No	No	60.9	59.1
45-layer ResNet	Yes	No	58.8	57.2
	Yes	0.2	57.5	57.2
	Yes	0.5	57.9	55.7

Table 4. Performance accuracy of 45-layer ResNet versus 9-layer ResNet.

TimeGAN data augmentation	Dropout	9-layer deep ResNet	45-layer deep ResNet
No	No	58.6	58.8
	0.2	55.8	57.5
	0.5	59.1	57.9
Yes	No	57.7	57.2

Dropouts are a potential solution to overfitting. The deep neural network framework of [15] proposed the use of dropouts in the MLP, but not the FCN or ResNet architectures, which instead incorporate batch normalization for regularization. We found that the use of batch normalization alone resulted in significant overfitting in both the FCN and ResNet architectures. Figure 4 shows the effect of varying the dropout value on training and test accuracy during the course of training for one sample k-fold of the MLP with PCA. Through the implementation of L1 regularization and fine-tuning the dropout value, we were eventually able to almost completely eliminate overfitting in all of the DNN architectures. However, the changes impacted the test accuracy only marginally: as seen in the figure, the train and test accuracies leveled but both remained in the 60s, with the primary change being a decrease in training accuracy.

We found that adding an explicit feature extraction layer by PCA led to higher performance for most, but not all, supervised classifiers that were tested. Referring to Figure 2, the test accuracy with PCA was 3.1%, 2.9%, and 1.0% higher than without PCA for linear SVM, MLP, and ResNet, respectively, unchanged for RBF SVM, while being 1.8% lower for RCN. In the case of linear SVM, transformation of the data into a lower dimensional space by PCA allowed the SVM's linear decision boundaries to more easily resolve the differences between classes. For the CNN-based neural networks, the closeness in performance with and without PCA feature extraction suggests they are performing feature extraction at some level. However, this built-in feature extraction appears to be sensitive to the presence of dropouts: in the absence of PCA, dropouts can be highly detrimental to classification performance. This is evident from Figure 3, which shows the impact of the dropout values on classification accuracy with and without PCA feature extraction in the three DNNs: including a PCA frontend makes the classification performance much less sensitive to variations in dropout values. In addition to higher classification robustness, PCA feature extraction also led to higher training efficiency in all the supervised ML methods to which it was applied. This is important for deep NNs, as they can require much more computational effort to train compared with shallower networks.

5. Conclusion

For gait classification, deep neural networks, despite having a huge number of trainable parameters and risk of overfitting, generally outperformed more shallow networks. Very deep NNs, however, showed no performance advantage over merely deep NNs, likely due to bounds on performance reached from the relatively few samples for training the large number of network parameters. The addition of a PCA feature extraction layer improves the gait classification performance for most, but not all, architectures studied but had a greater effect on shallow network architectures. This is perhaps owing to the hypothesis that an inherent part of a deep neural network serves a similar role as a feature extraction layer; thus, the latter's inclusion renders non-DNNs like a minimalist form of DNNs. Ultimately, our findings in PCA lead us to suggest the use of PCA as a machine learning frontend for training robustness. The effect of the dropouts in the convolutional layers of an FCN or ResNet is detrimental to performance when unaccompanied by PCA feature extraction, yet performance-enhancing when applied in conjunction with PCA. The incorporation of PCA feature extraction and dropouts in deep neural network architectures produced the highest classification performance for gait data, surpassing current state-of-the-art SVM methods. Due to the pressing need for medical datasets and their inherent scarcity, an area for future work is discovering effective time-series data synthesis methods for data augmentation.

References

[1] Matsushita Y, Tran DT, Yamazoe H, Lee JH. Recent use of deep learning techniques in clinical applications based on gait: a survey. Journal of Computational Design and Engineering 2021; 8 (6): 1499-1532. doi: 10.1093/jcde/qwab054

[2] Phinyomark A, Petri G, Marcelo EI, Osis ST, Ferber R. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. Journal of Medical and Biological Engineering 2018; 38: 244–260. doi: 10.1007/s40846-017-0297-2

[3] Slijepcevic D, Horst F, Horsak B, Lapuschkin S, Raberger A et al. Explaining Machine Learning Models for Clinical Gait Analysis. ACM Transactions on Computing for Healthcare 2021; 3 (2): 1–27. doi: 10.1145/3474121

[4] Abid M, Mezghani N, Mitiche A. Knee Joint Biomechanical Gait Data Classification for Knee Pathology Assessment: A Literature Review. Applied Bionics Biomechanics 2019; 2019: 7472039. doi: 10.1155/2019/7472039

[5] Gadaleta M, Rossi M. IDNet: Smartphone-based gait recognition with convolutional neural networks. Pattern Recognition 2018; 74:25–37.

[6] McGinnis RS, Mahadevan N, Moon Y, Seagers K, Sheth N et al. (2017). A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis. PLoS One 2017; 12 (6):e0178366.

[7] Wafai L, Zayegh A, Woulfe J, Begg R. Automated classification of plantar pressure asymmetry during pathological gait using artificial neural network. IEEE 2nd Middle East Conference on Biomedical Engineering 2014; 17-20. doi: 10.1109/MECBME.2014.6783244

[8] Trentzsch K, Schumann P, Sliwinski G, Bartscht P, Haase R et al. Using Machine Learning Algorithms for Identifying Gait Parameters Suitable to Evaluate Subtle Changes in Gait in People with Multiple Sclerosis. Brain Sciences 2021; 11:1049.

[9] Tedesco S, Crowe C, Ryan A, Sica M, Scheurer S et al. Motion Sensors-Based Machine Learning Approach for the Identification of Anterior Cruciate Ligament Gait Patterns in On-the-Field Activities in Rugby Players. Sensors (Basel) 2020;20 (11):3029.

[10] Prakash C, Kumar R, Mittal N. Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. Artificial Intelligence Review 2018; 49:1-40.

[11] Wan C, Wang L, Phoha V. A survey on gait recognition. ACM Computing Surveys 2018, 51:5(89).

[12] Lai DTH, Levinger P, Begg RK, Gilleard WL, Palaniswami M. Automatic Recognition of Gait Patterns Exhibiting Patellofemoral Pain Syndrome Using a Support Vector Machine Approach. IEEE Transactions on Information Technology in Biomedicine 2019; 13 (5):810-7.

[13] Horst F, Lapuschkin S, Samek W, Muller KR, Schollhorn WI. Explaining the unique nature of individual gait patterns with deep learning. Scientific Reports 2019; 9:2391. doi: 10.1038/s41598-019-38748-8

[14] Slijepcevic D, Zeppelzauer M, Gorgas AM, Schwab C, Schuller M et al. Automatic Classification of Functional Gait Disorders. IEEE Journal of Biomedical and Health Informatics 2017; 22 (5).

[15] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline. International Joint Conference on Neural Networks 2017. doi:10.1109/IJCNN.2017.7966039

[16] Boompelli SA, Bhattacharya S. Design of a Telemetric Gait Analysis Insole and 1-D Convolutional Neural Network to Track Postoperative Fracture Rehabilitation. IEEE 3rd Global Conference on Life Sciences and Technologies 2021; 484-488.

[17] Iber M, Dumphart B, Oliveira VAJ, Ferstl S, Reis J et al. Mind the Steps: Towards Auditory Feedback in Tele-Rehabilitation Based on Automated Gait Classification. Proceedings of the ACM AM '21 2021. doi: 10.1145/3478384.3478398

[18] Savadkoohi M, Oladunni T, Thompson LA. (2021). Deep neural networks for human's fall-risk prediction using force-plate time series signal. Expert Systems with Applications 2021; 182:115220.

[19] dos Santos DA, Duarte M. A public data set of human balance evaluations. PeerJ 2016; 4:e2648; DOI 10.7717/peerj.2648.

[20] Horsak B, Slijepcevic D, Raberger AM, Schwab C, Worisch M et al. GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. Scientific Data 2020; 7:143.

[21] Horst F, Slijepcevic D, Simak M, Schöllhorn WI. Gutenberg Gait Database, a ground reaction force database of level overground walking in healthy individuals. Scientific Data 2021 ;8 (1):232.

[22] Zhang Y, Zhang Y. Sports Training System Based on Convolutional Neural Networks and Data Mining, Computational Intelligence and Neuroscience 2021: 1331759.

[23] Kale A, Cuntoor N, Yegnanarayana B, Rajagopalan A, Chellappa R. Gait analysis for human identification. Lecture Notes in Computer Science 2003; 2688:1058-1058.

[24] Eskofier BM, Kraus M, Worobets JT, Stefanyshyn DJ, Nigg BM. Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking. Computer Methods in Biomechanics and Biomedical Engineering 2012; 15 (5):467-74.

[25] Osis ST, Hettinga BA, Leitch J, Ferber R. Predicting timing of foot strike during running, independent of striking technique, using principal component analysis of joint angles. Journal of Biomechanics. 2014 ;47 (11):2786-9.

[26] Schreiber C, Moissenet F. A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. Scientific Data 2019; 6:111.

[27] Kinsella S, Moran K. Gait pattern categorization of stroke participants with equinus deformity of the foot. Gait Posture. 2008;27 (1):144-51.

[28] Yu L, Mei Q, Xiang L, Liu W, Mohamad NI et al. Principal Component Analysis of the Running Ground Reaction Forces With Different Speeds. Frontiers in Bioengineering and Biotechnololgy 2021;9:629809.

[29] Bramah C, Preece SJ, Gill N, Herrington L. Is There a Pathological Gait Associated With Common Soft Tissue Running Injuries? The Americal Journal of Sports Medicine 2018 ;46 (12):3023-3031.

[30] Gafurov D, Snekkenes E, Bours P. Spoof attacks on gait authentication system. IEEE Trans Information Forensics Security 2007; 2 (3):491-502.

[31] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conf Computer Vision and Pattern Recognition 2016 June. doi:10.1109/CVPR.2016.90

[32] Wang Z, Oates T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. 29th AAAI Conf Artificial Intelligence 2015 Jan.

[33] Yoon J, Jarrett D, Schaar M. Time-series generative adversarial networks. 33rd Conf Neural Information Processing Systems 2019; 5508-5518.

[34] Iwana B, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. PLoS ONE 2021;16 (7): e0254841.

[35] Jung E, Lin C, Contreras M, Teodorescu M. Applied machine learning on phase of gait classification and joint-moment regression. Biomechanics 2022; 2 (1): 44-65.

[36] Chen B, Chen C, He J, Sayeed Z, Qi J et al. Computer Vision and Machine Learning-Based Gait Pattern Recognition for Flat Fall Prediction. Sensors 2022; 22 (20): 7960.

[37] Harris E, Khoo I, Demircan E. A Survey of Human Gait-Based Artificial Intelligence Applications. Front. Robot. AI, 2022 Jan 3.