BioDeep: A Deep Learning System for IMU-based Human Biometrics Recognition

Abeer Mostafa[1][a], Samir A. Elsagheer[1,3][b] and Walid Gomaa[1,2][c]

[1]*Cyber-Physical Systems Lab, Egypt Japan University of Science and Technology, Alexandria, Egypt*
[2]*Faculty of Engineering, Alexandria University, Alexandria, Egypt*
[3]*Faculty of Engineering, Aswan University, Egypt*

Keywords: IMU, Transfer Learning, Convolutional Neural Networks, Age and Gender Recognition, Deep Learning.

Abstract: Human biometrics recognition has been of wide interest recently due to its benefits in various applications such as health care and recommender systems. The rise of deep learning development, together with the massive data acquisition systems, made it feasible to reuse models trained on one task for solving another similar task. In this work, we present a novel approach for age and gender recognition based on gait data acquired from Inertial Measurement Unit (IMU). BioDeep design is composed of two phases, first of which is applying a statistical method for feature modelling, the autocorrelation function, then building a Convolutional Neural Network (CNN) for age regression and gender classification. We also use random forest as a baseline model to compare the results achieved by both methods. We validate our models using four publicly available datasets. The second phase is doing transfer learning over these diverse datasets. We train a CNN on one dataset and reuse its feature maps over the other datasets for solving both age and gender recognition problems. Our experimental evaluation over the four datasets separately shows very promising results. Furthermore, transfer learning achieved $20 - 30x$ speedup in the training time in addition to keeping the acceptable prediction accuracy.

1 INTRODUCTION

Recognition of human biometrics such as age and gender has been widely been studied in the recent decades due to its important use in many applications such as speech analysis (Markitantov, 2020), (Albuquerque et al., 2021), recommendation systems (Sun et al., 2017), and of course health care applications (Rosli et al., 2017). Researchers use various types of data for performing age and gender recognition starting from images, voice signals to inertial measurement unit (IMU) signals. In this paper, we build our age and gender recognition system based on IMU data. IMUs have two essential sensors: accelerometer and gyroscope and sometimes more additional sensors are included. The accelerometer sensor produces a tri-axical signal corresponding to the proper acceleration of the moving body accelerometer-X, accelerometer-Y, accelerometer-Z,

whereas the gyroscope sensor is used to determine the angular velocity of the moving body, and also produces a tri-axical signal gyroscope-X, gyroscope-Y, and gyroscope-Z.

The motivation for this work is that there is a big gap in research regarding the analysis of human biometrics based on IMU data since most of the literature uses IMU data for human activity recognition only. In addition, nowadays, IMUs are embedded in all of the wearable devices we use in our everyday life so it will be suitable for our analysis to work on IMU data.

Our contributions are illustrated as follows. We propose a novel methodology for the analysis of human biometrics based on IMU data and applying it specifically on age regression and gender classification using various datasets. In addition, our work is the first of its kind to perform cross-testing (transfer learning) for age regression and gender classification using IMU data. Furthermore, we measure the timings for our experiments and report the computational speedup gained in our approach.

We begin firstly by applying autocorrelation function on the accelerometer and gyroscope signals for

---

[a] https://orcid.org/0000-0002-8971-4311
[b] https://orcid.org/0000-0003-4388-1998
[c] https://orcid.org/0000-0002-8518-8908

feature modeling. The reason for choosing this statistical feature is that it is simple, efficient, and helps in reducing the input data size which is better in terms of processing time and memory. Afterwards, we apply modern machine learning techniques for age regression and gender classification. We begin by using random forest as it is considered a powerful machine learning model. Additionally, we investigate the deep learning approach and compare both results. In the deep learning approach, we apply a Convolutional Neural Network (CNN) on each dataset. Consequently, We apply transfer learning from one dataset to another for the sake of minimizing the number of learning parameters and, hence, reducing training time dramatically.

We apply our proposed methodology on four publicly available datasets: EJUST-GINR-1 (Mostafa et al., 2020), OU-ISIR (Ngo et al., 2014), GEDS (Miraldo et al., 2020), and HuGaDB (Chereshnev and Kertész-Farkas, 2017). We begin by applying both models (random forest and CNN) on each dataset separately for gender classification and age regression, then we perform cross testing by training the CNN on one dataset then fine-tune and test on the others.

The rest of the paper is organised as follows. In section 2, we review the state of the art research work that have been published in the area of age and gender estimation and also in the area of transfer learning. In section 3, we provide a brief description for all of the four considered datasets. In section 4, we illustrate our proposed methodology in detail. In section 5, we explain the setup for our experiments. In section 6, we show and discuss the results achieved by our methodology. Finally, we summarize our work in section 7 and conclude the paper.

2 RELATED WORK

In this section, we briefly provide a literature review on the latest work that have been published related to two research areas. Firstly, research that addresses solving the problem of age and gender estimation. Secondly, research in the area of deep learning applied to solve classification or regression problems using IMU signals with focus on doing transfer learning among various datasets.

2.1 Age and Gender Recognition

Age and gender recognition has been a common research area for many years. Almost all types of data have been used in research to identify these characteristics for various applications. The authors

in (Mostafa et al., 2020), proposed a robust method for gender recognition based on IMU data acquired from 8 sensors placed in 8 different positions on the human body during gait activity. The proposed method included using wavelet transform as a feature extractor along with various classifiers. They evaluated their approach on EJUST-GINR-1 dataset and had successfully reached accuracy up to 96.74% using the sensor placed on the left cube.

With reference to (Ngo et al., 2019), a competition on gender and age recognition was conducted using OU-ISIR Gait dataset (Ngo et al., 2014), which includes IMU signals extracted during gait activity. According to the competition results, most of the participated teams got relatively low accuracy in gender classification. However, the results of age regression were obviously better. The best results were presented in (Garofalo et al., 2019) which showed accuracy of gender classification reaching 75.77% and age regression with mean absolute error equals to 5.3879 by using orientation independent AE-GDI representation along with a CNN.

In the work proposed by (Riaz et al., 2015), a system for gender, age, and height estimation was presented based on IMU gait signals of four sensors located on the moving body. The authors used random forest classifier along with two validation methods: 10-fold cross-validation and subject-wise cross-validation. They categorized the age to three groups to perform age classification: less than 40 years, between 40 and 50, and older than 50 years. The highest accuracy for gender classification was achieved by the chest sensor and equals to 92.57%. Regarding age classification, the accuracy reached 88.82% using chest and lower back sensors.

The authors in (Jain and Kanhangad, 2016) investigated gender recognition using accelerometer and gyroscope data from the built-in smartphone sensors. The authors used multi-level local pattern (MLP) and local binary pattern (LBP) as feature extractors. To classify the extracted features, the authors used support vector machine (SVM) and aggregate bootstrapping (bagging). To evaluate these models, 252 gait signals collected from 42 subjects were used. The final result for gender classification reached 77.45% by applying MLP along with bagging.

2.2 Deep Learning and Transfer Learning on IMU Data

Research in deep learning has gained a huge popularity in the recent few years and proved to be a powerful methodology for solving machine learning problems. Some researchers applied deep learning and transfer

learning on IMU signals to perform classification or regression tasks which resulted in good performance.

The authors in (Abdu-Aguye and Gomaa, 2019) proposed a robust method for doing transfer learning over IMU data in the domain of human activity recognition. Their approach was to train a CNN on one dataset and use the convolutional filters as feature extractor, then train a feed forward neural network as a classifier on the extracted features for other datasets. This method proved to be very promising as the results were within 5% compared to the CNNs which were trained from scratch, end-to-end, with time speedup of $24 - 52x$.

In (Fu et al., 2021), the authors designed a compact wireless wearable sensing node that combines an air pressure sensor and IMU to be used for human activity recognition. Their method is to apply a transfer learning algorithm that consists of a joint probability domain adaptive method with improved pseudo-labels (IPL-JPDA). The authors used this method to recognise 7 daily human activities. The average recognition accuracy of different subjects reached 93.2%.

With reference to (Du et al., 2019), the authors used cascade learning and compared it to end-to-end learning in many transfer learning experiments applied to human activity recognition for IMU data. The method proved to be robust and reliable. They achieved 15% improvement of F1 score compared to other methods.

The authors in (Ashry et al., 2020) Proposed a deep learning method called CHARM-Deep to perform offline/online continuous human activity recognition based on IMU data streams collected from smartwatches. They built a cascaded bi-directional long short term memory (Bi-LSTM) to classify the feature vector extracted from IMU data. Statistical features such as autocorrelation, entropy, median were extracted from the signals, then fed to the Bi-LSTM. The proposed method achieved high accuracy of 94.2% to 97.2% in classification of all activities. In addition, it reduced the processing time by 86% compared to training Bi-LSTM on raw data with also space reduction by 97.7%.

From the previous literature, it is apparent that deep learning methods including transfer learning can demonstrate very good performance on accelerometer and gyroscope signals. However, these methods were only used in human activity recognition. In this work, we investigate the use of deep learning and, in particular, transfer learning on accelerometer and gyroscope data to be applied for age and gender estimation and seeing if deep learning can learn the features that distinguish the gender of a person and the features that

accurately estimate their age during the particular activity of "walking". In addition, we investigate doing transfer learning (cross-testing) among four datasets to see if the learned feature space from one dataset can effectively be used on another dataset and gain time speedup compared to training the network on each dataset from scratch.

3 DATASETS

Before we illustrate our proposed methodology, we provide a brief description of the datasets we used to run our experiments to its test the validity and effectiveness. We used four publicly available datasets. Each dataset was collected in a different place with different environmental setup. Furthermore, the datasets were collected using different devices, so they produce other modalities beside accelerometer and gyroscope, however, we consider only accelerometer and gyroscope signals in our analysis. These two sensors are available in all of the considered datasets, so it will make possible to do transfer learning. Another variation among the datasets is sensor placement, meaning that the IMU sensors were placed differently in each dataset, however, we consider the placements they have in common. For example, if a dataset has all sensors placed on legs only and the other dataset has sensors placed on arms and legs, we consider the data of leg sensors only to do transfer learning. In the following four subsections, we briefly explain each dataset.

3.1 EJUST-GINR-1 Dataset

EJUST-GINR-1 dataset (Mostafa et al., 2020), (Adel et al., 2020) was collected using six IMU sensors called MetaMotionR in addition to two Apple smart watches series-1 at each hand. The sensor placement, as shown in Figure 1, are: waist, back, left upper arm, right upper arm, left cube, and right cube. The data acquisition system was able to synchronize the eight sensing elements together when a subject was walking and capture the accelerometer and gyroscope signals with 50Hz frequency.

Twenty subjects participated in collecting this dataset, 10 males and 10 females, with age ranging from 19 years to 33 years. The procedure of collecting the data was that each person walks naturally on a straight ground for 20 minutes decomposed into small sessions. The total number of samples in the dataset is $5,292$ fixed-length samples.
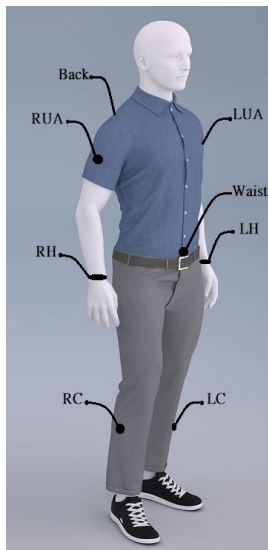
Figure 1: EJUST-GINR-1 dataset sensor placement.

3.2 OU-ISIR Gait Dataset

The OU-ISIR gait dataset (Ngo et al., 2014) is the largest IMU gait dataset in the world. The dataset was collected using three IMU sensors named as IMUZ and a Motorola smartphone placed on a belt around the waist of the person. The dataset includes a tri-axial accelerometer and a triaxial gyroscope gait signals of 744 subjects (389 males and 355 females) with age ranging from 2 years to 78 years. The data was recorded at frequency of 100Hz. The dataset was collected in an exhibition and formulated for many different research purposes. Each person provided two signals of level walk gait. In addition, another protocol of the dataset was released that has up-slope and down-slope data for 495 subjects. However, in this paper, we consider only level walk data in order to be used in transfer learning (cross-testing) with other datasets. The total number of samples for level walk only is 1,488 samples. We obtained the permission to use this dataset in our research by a signed agreement from EJUST university to Osaka university.

3.3 Gait Events DataSet (GEDS)

The Gait Events DataSet (GEDS) (Miraldo et al., 2020) is a publicly open dataset which was collected using six wireless sensors. Four of them produce accelerometer and gyroscope signals with other modalities, and two of them are force-sensitive resistor (FSR) sensors. In this paper, we work with the data captured from the sensors placed at the tibialis anterior muscles in the right and left legs which are named (TaR) and (TaL), and the sensors placed over the tibia

bones at the right and the left legs which are named (TbR) and (TbL). The number of subjects who participated in collecting the dataset is 22 (10 males and 12 females) including one female with a foot drop gait abnormality. The ages of the subjects range from 18 years to 50 years. The dataset contains a total of 9,661 gait strides. It also contains gait data corresponding to 3 different speeds: slow walk, fast walk, and comfortable walk. In this work we consider the comfortable gait style.

3.4 Human Gait Database (HuGaDB)

The Human Gait Database (HuGaDB) (Chereshnev and Kertész-Farkas, 2017) consists of recordings for 12 human activities such as walking, running, sitting, standing and so on. The dataset was collected using six IMU sensors that produce accelerometer and gyroscope signals in addition to two electromyography (EMG) sensors. The IMU sensors were located at the right and left thighs, shins, and feet. Eighteen healthy subjects participated in collecting the dataset (14 males and 4 females), their ages range from 18 to 35 years. The sampling rate of the dataset is 56.35Hz. The total records of the data equaled 10 hours, then the dataset was segmented and annotated. In this work, we consider only the walking activity data and the IMU sensors.

4 PROPOSED SCHEME

We now describe our proposed methodology to build a reliable age and gender recognition system. Figure 2 illustrates the stages of our proposed methodology. We begin by loading the IMU data consisting of triaxial accelerometer (accelerometer-X, accelerometer-Y, accelerometer-Z) and triaxial gyroscope (gyroscope-X, gyroscope-Y, and gyroscope-Z). The data preprocessing stage and feature extraction methods are the same in both gender classification and age regression. After the feature extraction phase, we feed the feature vector to either a classifier for learning to classify gender, or to a regressor to learn to estimate age.

4.1 Feature Modeling

To be able to extract useful information out of our raw data, we propose to first apply a statistical method in the data preprocessing stage. One of the most useful statistical properties in the analysis of timeseries is the *autocorrelation function*. The autocorrelation function is a measure of similarity among the signal
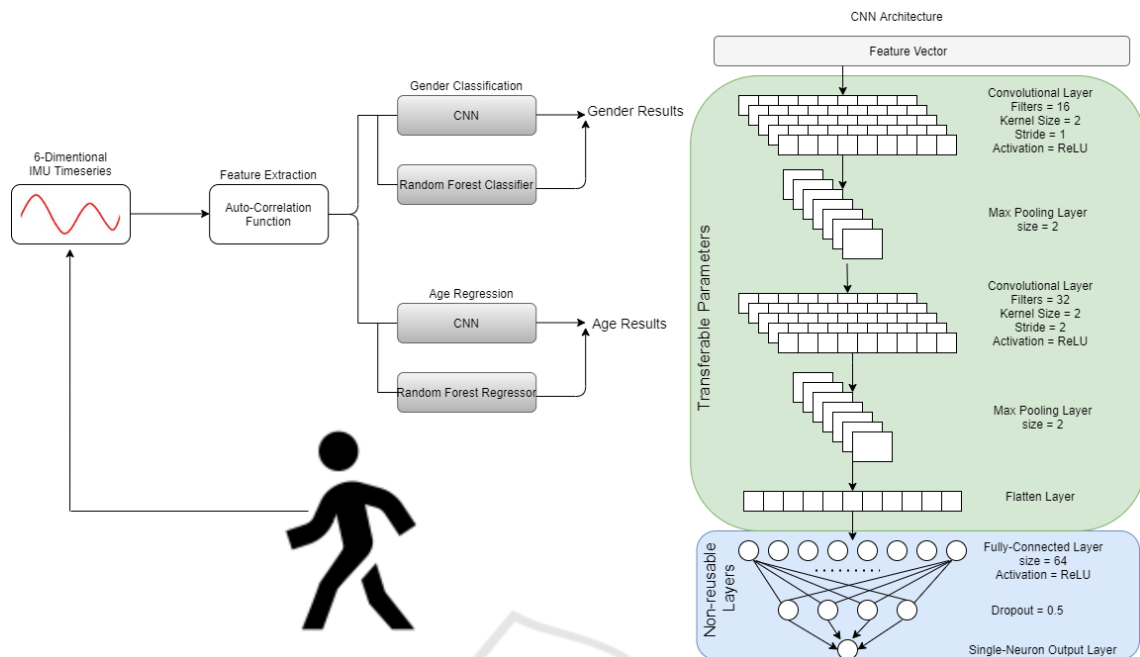
Figure 2: Proposed BioDeep System Architecture.

and time-shifted versions of itself. It is particularly useful in the analysis of signals to find the repeated patterns such as gait data in our case as the motion is repeated periodically. We calculate the autocorrelation function *acf* for each sensory signal up to a certain lag which is specified by experiments. The sample autocorrelation function is calculated as indicated in Equations (1) and (2).

$$acf(h) = \frac{\gamma(h)}{\gamma(0)} \qquad (1)$$

$$\gamma(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}) \qquad (2)$$

Where $\bar{x}$ corresponds to the sample mean, n represents the signal length, and h is the lag (Gomaa et al., 2017). The output of the autocorrelation function is a vector of length (6 x *number of lags*) as we have 6-axes signals. This vector is then considered the *feature vector* to be fed to a classifier or a regressor. Another benefit of using the autocorrelation function is dimensionality reduction. At the beginning, our data consists of 6-dimensional timeseries. If we take EJUST-GINR-1 dataset as an example, each sample has 250 sequential data points which result in dimensions of (6x250) for each gait sample. However, after applying the autocorrelation function, it results in a dramatic reduction in the dimension. From the previous analysis, *acf* is best suitable for modeling the features of our data.

4.2 Age Estimation

In this work, we consider age estimation as a regression problem to approximately estimate the exact age of a person. A supervised machine learning approach is employed as all of the considered datasets provide age information attached to the gait signals. We fix the feature vector as the output of the autocorrelation function; after that, we feed this feature vector to a regressor. We first try simple machine learning technique such as a random forest regressor, then we go for the deep learning approach in which we use Convolutional Neural Network (CNN).

Random Forest is a well-known robust method in machine learning. As shown in (Mehrang et al., 2018), (Feng et al., 2015), (Mostafa et al., 2020), and (Casale et al., 2011), random forest has proven to be very promising applied to the analysis of IMU signals. Random forest has such power because it relies on ensemble learning, which means that not only one big model is created to analyze the features of the given dataset, but many different smaller models are created using decision trees. Furthermore, each model of those not only uses a subset of the data but also performs random feature selection so as to reduce the variance of prediction and hence increase accuracy (Mostafa et al., 2020). Here, we apply a random forest technique for doing regression on the age with model specifications illustrated in section 5.

Another method we use here for regression is CNN. The reason for that is because we want to make use of the power of deep learning for solving this supervised learning problem and compare its results to the results of a traditional method such as random forest. CNNs are known to be good at solving machine learning problems based on multi-dimensional data. In our work, we construct the network shown in Figure 2. The input to the first convolutional layer is the feature matrix resulted from the autocorrelation function which is considered six channels corresponding to the six timeseries autocorrelation functions. The network is composed of 2 convolutional layers in order to create the feature maps for age and each one is followed by a max pooling layer to create a summarized version of those feature maps. Following that, we flatten the output of the second max pooling layer then feed it to a fully-connected layer. Consequently, we add a dropout layer for regularization to prevent overfitting and to boost the computational performance by minimizing the large number of learning parameters caused by the fully-connected layer. Finally, the output layer consists of one neuron that will be produce the prediction for the age. More specification about the model hyperparameters are provided in section 5.

4.3 Gender Recognition

Gender recognition is considered a binary classification problem. We use the same scheme proposed for age regression as illustrated in Figure 2 except for the activation function in the output layer of the CNN and the loss function for estimating the error.

4.4 Transfer Learning

Transfer learning is a machine learning paradigm which has proved to be very effective recently. It means that the model which had previously learned on a specific task can be reused as a starting point to learn another semi-similar task which means that the learning time for the latter task will be reduced significantly. In this context, we apply transfer learning among the four previously mentioned datasets. First, we train the CNN architecture shown in Figure 2. The top layers which are highlighted in green, that include the convolutional and max pooling layers, learn the features of the input by creating the corresponding feature maps. Consequently, the bottom layers which are highlighted in blue in Figure 2 are then used to classify the gender or estimate the age according to the application at that experiment (keeping in mind that the activation function at the last layer

and the loss function will be different in regression as illustrated in the model specification in section 5). The same pretrained top layers are taken after training on one dataset with the same parameters, then we attach to them new bottom layers and do fine-tuning with this new model on another dataset. We call this "cross-testing" The fine-tuning process is not like training the network from scratch. In fine-tuning, we train the model for 2 or 3 epochs only to capture the features of the new dataset. Afterwards, we evaluate the model on the test samples from the second dataset which the model haven't seen before.

5 EXPERIMENTAL SETUP

To evaluate the effectiveness of our proposed methodology, we apply our model over the four previously mentioned datasets. Due to the variation among the datasets, we establish a systematic protocol of our experiments. In this section, we illustrate the specifications of those experiments. First, we calculate the *acf* for all sensor data of each dataset. We take different lag values up to 10. For each sample, we have 6-dimensional signal: accelerometer-X, accelerometer-Y, accelerometer-Z, gyroscope-X, gyroscope-Y, and gyroscope-Z. So the resulted feature space is in $\mathbb{R}^{66}$. We tried many lag values then chose the best value that resulted in the highest accuracy which was 10.

Secondly, we train and test random forest and CNN classification models on each dataset separately for gender classification. Consequently, we train and test random forest and CNN classification models on each dataset separately for age regression. We run these experiments on each sensor data to compare the performance and to recognize which body parts best provide age and gender information. After that, we perform cross-testing, that is, we train the CNN classification model on one dataset then fine tune and test on another dataset. We take the data from the same sensor location in both datasets. We begin our training on EJUST-GINR-1 dataset which includes the largest number of sensor locations. We use the learned features from the shin sensor in EJUST-GINR-1 (RC) and transfer them to the data of taR sensor in GEDS dataset. Similarly, we do the same on shin sensor in HuGaDB which is (RS). The OU-ISIR dataset doesn't contain shin sensors, so we train on the waist sensor data in EJUST-GINR-1 dataset and use the waist sensor in OU-ISIR dataset. After that, we repeat these experiments but with choosing another dataset for training so that we make a sort of cross testing among the four datasets. Eventually, we compare the results to those of random forest. Additionally, we perform

some timing evaluation experiments in order to explore the computational speed we gained by applying transfer learning compared to training from scratch.

The experimental model specifications for the random forest classifier are as follows: we set the number of decision trees in the model to be 100 trees because it is a suitable choice to give a fair result as random forest takes the majority vote. We also use Gini index as a measure of the split quality. We apply bootstrap aggregation to select random subsets of the data and random subsets of the feature set thus prevent overfitting. We run each experiment 10 times and write down the average accuracy.

The CNN model specifications, shown in Figure 2, are as follows: The first convolutional layer has 16 filters with stride 1, the second convolutional layer has 32 filters with stride 2 and the fully-connected layer has 64 neurons. These parameters were selected experimentally by applying a specific setup and evaluating the model performance then choosing the best evaluated setup. All of the three layers use Rectified Linear Unit (ReLU) as activation function whereas the final single-neuron output layer uses sigmoid activation in the case of gender classification and linear activation in the case of age regression. We use Adam optimizer and set the batch size to 10 samples. We apply binary cross-entropy as a loss function in the case of gender classification and mean absolute error in the case of age regression. In addition, we scale the values of the age to be from 0 to 1 in order to have a better training process and faster convergence. The code for all of these experiments was uploaded on GitHub and available upon request.

6 RESULTS AND DISCUSSION

In this section, we include all the results achieved by applying our methodology to the four mentioned datasets. We firstly show the results of each sensor in each dataset separately for age regression and gender classification. Consequently, we show the results of cross testing i.e., training on a sensor data of one dataset then testing on the other datasets.

6.1 Results of Age Estimation

The results of applying our approach for age regression are shown in Figure 3. We refer to EJUST-GINR-1 dataset sensors right cube as RC, left cube as LC, left upper arm as LUA, right upper arm as RUA, left hand as LH, and right hand as RH. For HuGaDB dataset, there was some corrupted data, so we consider the right-side sensors only and refer to

right foot as RF, right shin as RS, and right thigh as RT. For the GEDS dataset, we evaluate the sensors placed at the tibialis anterior muscles in the right and left legs and refer to them as TaR and TaL, and the sensors placed over the tibia bones at the right and the left legs referred to as TbR and TbL. For the OU-ISIR database, the data was extracted automatically for the subjects using the center IMUZ placed on waist. Age regression evaluation was done by measuring the mean absolute error. For consistency, the results shown were achieved by taking the age range from range 15 to 35 years as this is the most common range of subjects' ages in all the datasets and any other data outside this age range is considered outliers. This means we consider in our evaluation all of 20 subjects of EJUST-GINR-1 dataset, all of 18 subjects of HuGaDB, 19 subjects of GEDS, and 385 subjects of OU-ISIR database. Figure 3 shows the results of using autocorrelation function followed by random forest regressor in blue and the results of CNN in orange using the same autocorrelation function features for each sensor in each dataset. This allows us to compare between using a traditional machine learning approach (random forest) and deep learning. It also allows us to determine which sensor location(s) can be used best to estimate age.

The overall performance indicates that the mean absolute error in age estimation for EJUST-GINR-1 dataset lies between 1.9 for the LH sensor and 2.37 for the LUA sensor in the case of using random forest, and lies between 0.88 for the RC sensor and 1.8 for the RH sensor in the case of using CNN. For HuGaDB, the mean absolute error lies between 1.38 for the RF sensor and 1.57 for the RT sensor in the case of using random forest, and lies between 0.835 for the RS sensor and 0.93 for the RT sensor in the case of using CNN. For GEDS, the mean absolute error lies between 2.2 for the TbR sensor and 2.29 for the TaL sensor in the case of using random forest, and lies between 1.3 for the TbR sensor and 2.37 for the TaL sensor in the case of using CNN. For OU-ISIR database, the result achieved using random forest was 3.99 and 4.28 using CNN.

From these results, we can observe that CNN performed better than random forest in most of the cases maybe this is due to the fact that CNN convolutional filters can capture more complex features such that in the case of mapping the gait signal to the corresponding subject's age. The results also show that lower sensor locations on the body result in better age estimation. That's reasonable as most of gait patterns can be featured from leg sensors.
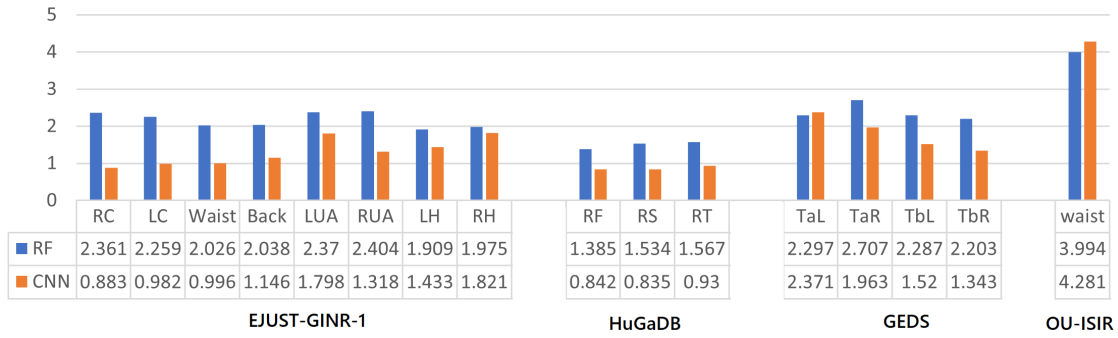
Figure 3: Age regression results for each dataset evaluated in mean absolute error.
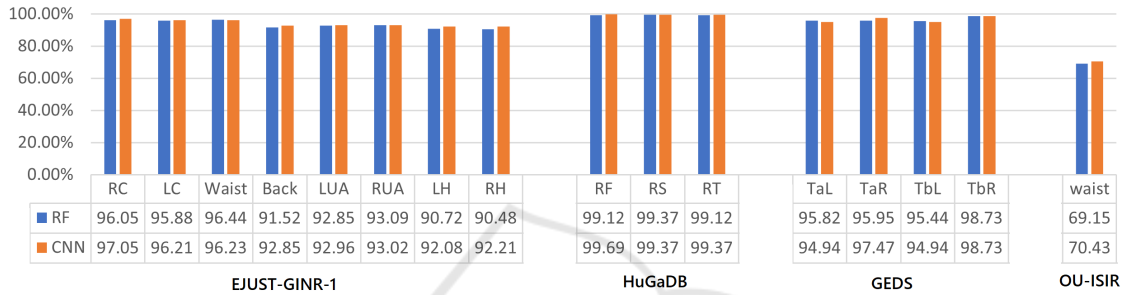

Figure 4: Gender classification results for each dataset evaluated in percentage accuracy.

Table 1: Results of cross-testing in age regression evaluated in mean absolute error.

Training Dataset	Testing Dataset			
	EJUST-GINR-1(RC)	HuGaDB(RS)	GEDS(TaR)	OU-ISIR(Waist)
EJUST-GINR-1(RC)	0.883	2.0	2.5982	NA
HuGaDB(RS)	2.5685	0.835	3.6507	NA
GEDS(TaR)	2.232	2.19	1.963	NA
EJUST-GINR-1(Waist)	NA	NA	NA	5.1209

Table 2: Results of cross-testing in gender classification evaluated in percentage accuracy.

Training Dataset	Testing Dataset			
	EJUST-GINR-1(RC)	HuGaDB(RS)	GEDS(TaR)	OU-ISIR(Waist)
EJUST-GINR-1(RC)	97.05%	98.13%	96.56%	NA
HuGaDB(RS)	95.58%	99.37%	96.2%	NA
GEDS(TaR)	95.9%	98.03%	97.47%	NA
EJUST-GINR-1(Waist)	NA	NA	NA	67.93%

6.2 Results of Gender Classification

The results of applying our approach in gender classification are shown in Figure 4. We use the same abbreviations mentioned in the previous subsection. Our metric for evaluation here is the classification percentage accuracy.

In Figure 4, we can see that the overall performance for EJUST-GINR-1 dataset lies between 90.48% for the RH sensor and 96.44% for the waist sensor in the case of using random forest, and lies between 92.08% for the LH sensor and 97.05% for the RC sensor in the case of using CNN. For HuGaDB, the classification accuracy lies between 99.1% for the

RF and RT sensors and 99.37% for the RS sensor in the case of using random forest, and lies between 99.37% for the RS and RT sensors and 99.69% for the RF sensor in the case of using CNN. For GEDS, the classification accuracy lies between 95.44% for the TbL sensor and 98.37% for the TbR sensor in the case of using random forest, and lies between 94.94% for the TbL and TaL sensors and 98.73% for the TaR sensor in the case of using CNN. For OU-ISIR database, the result achieved using random forest was 69.15% and 70.43% using CNN. from these results, we can observe that CNN had a slight improvement in performance compared to random forest.

6.3 Cross-testing

The results of applying transfer learning, for cross-testing, in age regression are shown in Table 1. The diagonal elements represent training and testing over the same dataset, so their results are considered the baseline for our experimental comparisons. The off-diagonal elements represent the results of transfer learning across different datasets. For example, the first row shows the results of training on EJUST-GINR-1 RC sensor and testing on other datasets with the same sensor location. In the last row, we used the waist sensor in EJUST-GINR-1 datast to train, then test on OU-ISIR dataset. The overall average loss in performance in the case of transfer learning is 1.2 in mean absolute error compared to training and testing on the same dataset, however, when measuring the training time for each, transfer learning provided $20 - 30x$ speedup in the training time compared to training from scratch. The training was performed on the Nvidia GeForce GTX 1650 GPU with 4GB of memory. The same conventions are applied in Table 2 to show the results of applying transfer learning in gender classification evaluated in percentage accuracy. The overall average accuracy loss is 1.4% in the case of transfer learning compared to the case of training and testing on the same dataset, however, we also achieved $20 - 30x$ speedup in the training time for gender classification.

It can be observed that EJUST-GINR-1 dataset consistently has the highest transfer learning accuracy in gender classification and age regression. This can be due to the fact that it consists of long sequences of gait signals and a balanced age and gender distribution over the participated subjects causing the convolutional filters of the CNN to efficiently model the features. Additionally, it can be observed that although OU-ISIR dataset has the largest number of subjects, the results achieved by using it for testing have the largest error in age regression and poorest accuracy in gender classification. Our reasoning for this is that the dataset contains two gait sequences for each subject and each sequence consists of only a few seconds which may not be enough for capturing the gender and age pattern features for each subject.

7 CONCLUSION

In this work, we proposed a novel scheme for age and gender recognition using IMU gait signals. Our design begins by applying the autocorrelation function on the triaxical accelerometer and triaxical gyroscope timeseries for feature modeling. Consequently, we

investigated, using two machine learning techniques random forest and CNN for age regression and gender classification. Furthermore, we applied transfer learning among datasets to reduce the training time and validate our model generalizability. We train the CNN on one dataset then fine-tune and test on other datasets. We used four publicly available datasets: EJUST-GINR-1, OU-ISIR, GEDS, and HuGaDB.

The results obtained from our experimental evaluation indicate that our proposed methodology yields a good performance in both age regression and gender classification. In addition, the transfer learning experiments yielded outstanding results compared to the baseline CNN-based models trained from scratch. The cross-testing average loss in performance was 1.2 in mean absolute error compared to networks that were trained from scratch in the case of age regression and 1.4% loss in classification accuracy in the case of gender classification. The speedup gained by transfer learning reached $20 - 30x$ in the training time. We believe these results should open the way in using pre-trained models for age and gender recognition using IMU timeseries the same way pretrained models are used nowadays in computer vision.

In the future, we intend to extend this work to investigate other various human activities such as sitting, climbing stairs, running, etc. We may also analyse electroencephalogram (EEG) signals to do age and gender recognition based on brain signals.

REFERENCES

Abdu-Aguye, M. G. and Gomaa, W. (2019). Versatl: Versatile transfer learning for imu-based activity recognition using convolutional neural networks. In *ICINCO (1)*, pages 507–516.

Adel, O., Nafea, Y., Hesham, A., and Gomaa, W. (2020). Gait-based person identification using multiple inertial sensors.

Albuquerque, L., Oliveira, C., Teixeira, A., and Figueiredo, D. (2021). Eppur si muove: Formant dynamics is relevant for the study of speech aging effects. *14th BIOSIGNALS, Online*.

Ashry, S., Ogawa, T., and Gomaa, W. (2020). Charm-deep:

Continuous human activity recognition model based on deep neural network using imu sensors of smartwatch. *IEEE Sensors Journal*, 20(15):8757–8770.

Casale, P., Pujol, O., and Radeva, P. (2011). Human activity recognition from accelerometer data using a wearable device. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 289–296. Springer.

Chereshnev, R. and Kertész-Farkas, A. (2017). Hugadb: Human gait database for activity recognition from wearable inertial sensor networks. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 131–141. Springer.

Du, X., Farrahi, K., and Niranjan, M. (2019). Transfer learning across human activities using a cascade neural network architecture. In *Proceedings of the 23rd international symposium on wearable computers*, pages 35–44.

Feng, Z., Mo, L., and Li, M. (2015). A random forest-based ensemble method for activity recognition. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5074–5077. IEEE.

Fu, Z., He, X., Wang, E., Huo, J., Huang, J., and Wu, D. (2021). Personalized human activity recognition based on integrated wearable sensor and transfer learning. *Sensors*, 21(3):885.

Garofalo, G., Argones Rúa, E., Preuveneers, D., Joosen, W., et al. (2019). A systematic comparison of age and gender prediction on imu sensor-based gait traces. *Sensors*, 19(13):2945.

Gomaa, W., Elbasiony, R., and Ashry, S. (2017). Adl classification based on autocorrelation function of inertial signals. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 833–837.

Jain, A. and Kanhangad, V. (2016). Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings. In *2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*, pages 597–602.

Markitantov, M. (2020). Transfer learning in speaker's age and gender recognition. In Karpov, A. and Potapova, R., editors, *Speech and Computer*, pages 326–335, Cham. Springer International Publishing.

Mehrang, S., Pietilä, J., and Korhonen, I. (2018). An activity recognition framework deploying the random forest classifier and a single optical heart rate monitoring and triaxial accelerometer wrist-band. *Sensors*, 18(2):613.

Miraldo, D., Watanabe, R., and Duarte, M. (2020). An open dataset of inertial, magnetic, foot-ground contact, and electromyographic signals from wearable sensors during walking.

Mostafa, A., Barghash, T. O., Assaf, A. A.-S., and Gomaa, W. (2020). Multi-sensor gait analysis for gender recognition.

Ngo, T. T., Ahad, M. A. R., Antar, A. D., Ahmed, M., Muramatsu, D., Makihara, Y., Yagi, Y., Inoue, S., Hossain,

T., and Hattori, Y. (2019). Ou-isir wearable sensor-based gait challenge: Age and gender. In *Proceedings of the 12th IAPR International Conference on Biometrics, ICB*.

Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., and Yagi, Y. (2014). The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, 47(1):228–237.

Riaz, Q., Vögele, A., Krüger, B., and Weber, A. (2015). One small step for a man: Estimation of gender, age and height from recordings of one step by a single inertial sensor. *Sensors*, 15(12):31999–32019.

Rosli, N. A. I. M., Rahman, M. A. A., Balakrishnan, M., Komeda, T., Mazlan, S. A., and Zamzuri, H. (2017). Improved gender recognition during stepping activity for rehab application using the combinatorial fusion approach of emg and hrv. *Applied Sciences*, 7(4):348.

Sun, M., Li, C., and Zha, H. (2017). Inferring private demographics of new users in recommender systems. In *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems*, pages 237–244.