SAPIENZA
UNIVERSITÀ DI ROMA

FACULTY OF INFORMATION ENGINEERING,
INFORMATICS AND STATISTICS

Gait Recognition with Wii Balance Board

BIOMETRIC SYSTEMS PROJECT

Ballanti Chiara - ballanti.1844613@studenti.uniroma1.it
Barreca Federico - barreca.1736423@studenti.uniroma1.it
Bevilacqua Paolo Pio - bevilacqua.2002288@studenti.uniroma1.it
De Sio Ilaria - desio.2064970@studenti.uniroma1.it

Academic Year 2022/2023

Contents

1 Introduction

The term "Biometric Recognition System" usually refers to the use of distinctive physiological or behavioral characteristics for the automatic recognition of individuals. Based on this concept, several implementations have been made for different biometrics. In this project, we decided to work on the gait trait and exploit the gait patterns to recognize an individual.

The fundamental requirements for a biometric trait are:

- *Universality*: the trait must be owned by any person;
- *Uniqueness*: any pair of people should be different according to the biometric trait;
- *Permanence*: the biometric trait should not change in time;
- *Collectability*: the biometric trait should be measurable by some sensor;
- *Acceptability*: involved people should not have any objection to allowing collection or measurement of the trait.

A biometric trait is considered *hard* if it meets all the requirements described above. On the other side, there are *soft* biometric traits, such as gait, which lack permanence. Since gait is a genotypic and a behavioural trait, it is characterized by a very high uniqueness.

All the data considered in this project were taken directly from us using the *Nintendo Wii Balance Board* (WBB) (Nintendo, Kyoto, Japan), which is an inexpensive, portable and accessible alternative to costly laboratory-grade force plates for measuring the vertical ground reaction force (vGRF) and center of pressure (CoP).
The WBB has generated significant interest in its application as a postural control measurement device in both the clinical and (basic, clinical, and rehabilitation) research domains [4, 9, 13]. Additional research confirmed the viability of the WBB as a valid and reliable center of pressure (CoP) measurement device [10, 1]. And in [8] the authors verified the accuracy of the WBB for measuring kinetic gait parameters. However, the validity of the WBB as a recognizer system has not been evaluated yet.

2 Data Processing

Figure 1 shows the pipeline of the process of creating a template from the sample of a subject.

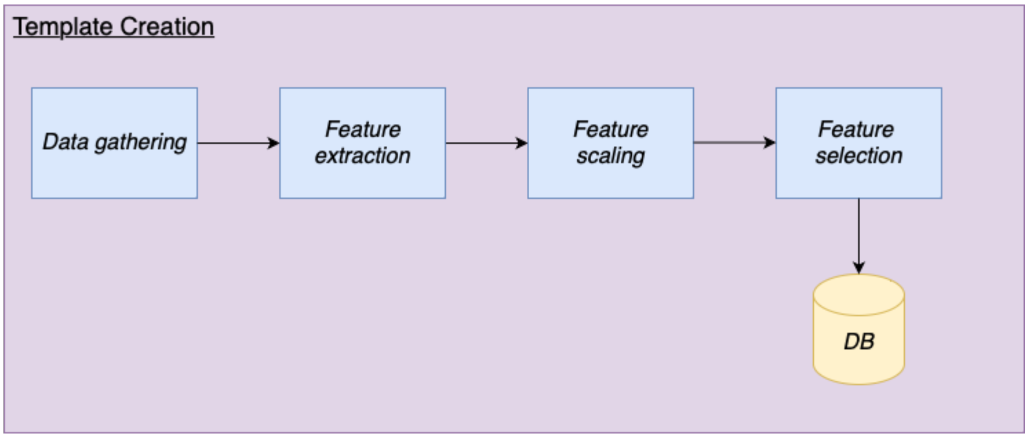


Figure 1: Pipeline.

2.1 Software and Acquisition Interface

The *Nintendo Wii Balance Board* (WBB) allows to determine and control the position of the center of mass of the whole body (COM). The center of mass represents the point around which the mass is evenly distributed. It is therefore possible to perform a static analysis to determine the coordinates of the center of mass in two dimensions. The WBB, as shown in figure 2, can measure the vertical ground reaction force (vGRF) and the center of pressure (COP) using four strain gauge load cells located near each of the four corners of the device.

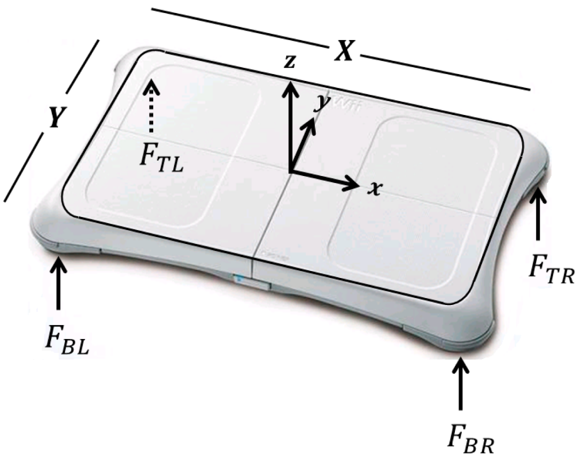
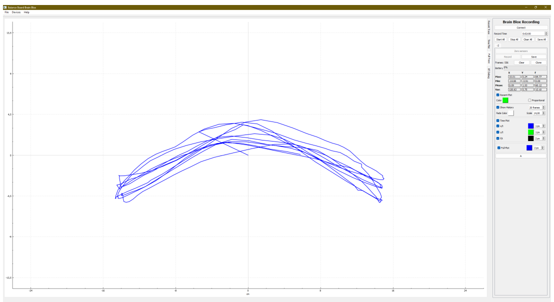


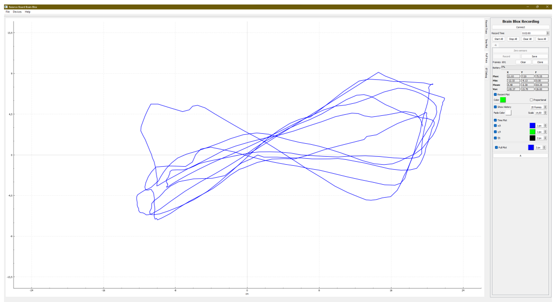
Figure 2: Sensors on Wii Balance Board.

The dataset has been built from scratch by us, taking samples of people of different ages (in a range between 11 and 66 years old) and in different physical conditions. The data has been collected by performing an on-site walk on the balance, taking 30 samples per person, each consisting of 10 steps. The final dataset is made up of 24 different subjects, with a total of 720 samples.

Data were gathered through a software called *BrainBLOX* ([5]), which provides an interface to capture, record, and visualize data acquired by this device. The *BrainBLOX* software interface consists of a graphical plot region and a control panel.
In figure 3, two examples of final sampling are presented: the first of a 26-year-old gymnast and the second of a 66-year-old.



(a) Sample of a 26-year-old gymnast



(b) Samples of a 66-year-old

Figure 3: Samples examples.

Data are time series saved in a tabular format. The key below indicates which data is saved in which column.

- Column A: Time in milliseconds elapsed since the program was opened;
- Column B: Force from sensor 1 (top left) in kilograms;
- Column C: Force from sensor 2 (top right) in kilograms;
- Column D: Force from sensor 3 (bottom left) in kilograms;
- Column E: Force from sensor 4 (bottom left) in kilograms;
- Column F: COP distance from the center in the x-direction in centimeters;
- Column G: COP distance from the center in the y-direction in centimeters;
- Column H: Total force (sum of columns 2-5) in kilograms.

Once sampling was finished, the file associated with the person was defined as follows:

- *ID*: user identifier;
- *M/F*: gender of the user;

- *AGE*: age of the user;
- *SX/DX*: foot with which sampling begins (*SX* for left and *DX* for right);
- *R/MS/S*: physical state in which the user at the time of sampling. "At rest", "medium-tired" and "tired" respectively.

An example of file naming is *S1_F_31_DX_R*.

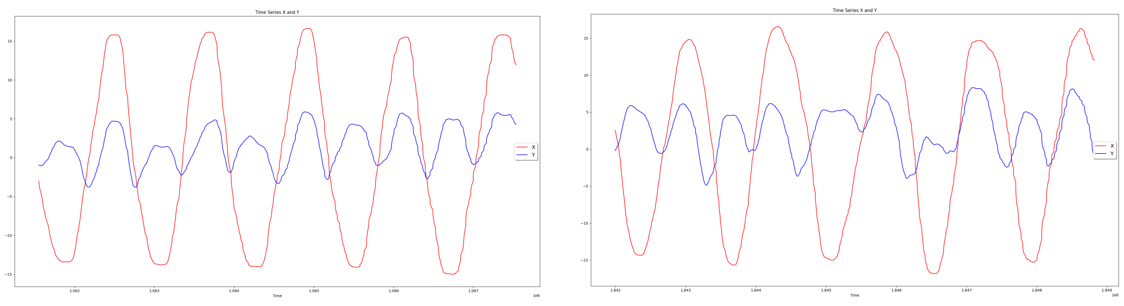
2.2 Features Extraction and Filtering

Since we did not need all the information provided by the resulting sample, we decided to select only the most relevant ones. Thus, we created a time series for each sample containing four elements:

- ID: Ground truth, the label that represents the subject's true identity;
- Column A: Time in milliseconds elapsed since the program was opened;
- Column F: COP distance from the center in the x-direction in centimeters;
- Column G: COP distance from the center in the y-direction in centimeters.

Time series analysis helps to understand the underlying causes of trends and patterns over time. The data are easily visualized and the final template size is about 30 Kb. They are computationally inexpensive, it takes about one minute to extract features from 720 samples using parallelization.

In figure 4, the two final time series of the samples of figure 3 are presented: the first of a 26-year-old gymnast and the second of a 66-year-old.



(a) Time series of a 26-year-old gymnast. (b) Time series of a 66-year-old.

Figure 4: Final time series examples.

We split the dataset into two sub-datasets of equal size:

- Training set: used to perform the features selection;
- Testing set: used to perform the evaluation.

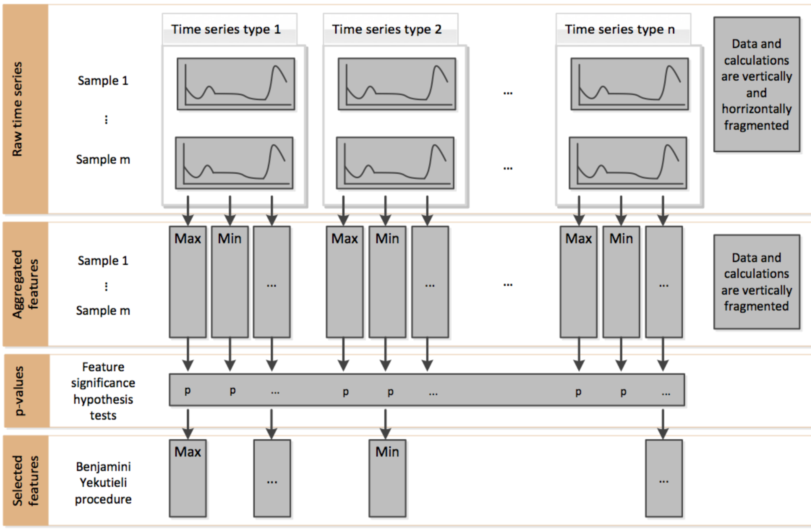


Figure 5: Phases of the filtering process provided by *tsFresh*.

We used the *tsfresh* [2] library to handle the problem of the feature extraction. The result is a raw dataframe of all the features identified by their names. The algorithm characterizes time series with comprehensive and well-established feature mappings and considers additional features describing meta-information.

Feature normalization is mostly needed to eliminate the effect of several quantitative features measured on different scales. We chose Min-Max Scaling, a process that normalizes value into a given range (min, max). The *Min-Max Scaler* is provided by the *scikit-learn* [12] library and the transformation is given by:

$$s'_k = \frac{s_k - min}{max - min} \tag{1}$$

We used *tsfresh* to handle the problem of feature selection, which is the identification of all strongly and weakly relevant attributes. It is a very powerful tool when each label is associated with several time series and meta-information simultaneously.
To limit the number of irrelevant features *tsfresh* deploys the *fresh algorithm* (*fresh* stands for *FeatuRe Extraction based on Scalable Hypothesis tests*) [3]. It is an efficient, scalable feature extraction algorithm, which filters the available features in an early stage of the machine learning pipeline with respect to their significance for the regression or classification task while controlling the expected percentage of selected but irrelevant features.

Starting from the selected features, using tsfresh we constructed a relevance table that allows us to sort by p-value and reduce to the most relevant k-features. For the evaluation, different values of k were tested: 50, 100, 200, 400 and 877. The best results were obtained by setting k equal to 200.
The number of features after each macro step is shown below:

- Feature extraction: 1566;
- Feature selection: 877;
- Reduction to k most relevant features: 200.

3 Evaluation

Biometric systems are susceptible to a variety of faults. In this section, we provide an overview of the most often used performance measures in the literature.
Before starting to evaluate the system, let us define some concepts [6].
There are two types of recognition operations:

- **Verification**: The user claims an identity and the system performs a 1:1 matching to verify the identity.
- **Identification**: There is no identity claimed and the system performs a 1:N matching to determine the correspondence with one of the subjects in the gallery.

Identification can be subdivided into two categories:

- **Open set**: where some probes may not belong to any subject in the gallery and the system has a reject option.
- **Closed set**: where all probes belong to an enrolled subject, the system doesn't have a reject option but may return the wrong identity.

We performed the evaluation by computing encompassing statistics using a complete **ALL-against-ALL distance matrix** (all templates are compared with any other template). We calculated the distance matrix between all samples in the Testing set, using the Euclidean distance (distance values are not normalized).
The distances are computed once and for all and used for performance evaluation of different kinds of applications: verification, identification closed set and identification open set, considering multiple templates per subject in the gallery.

In the following sections, the results of the evaluation phase will be presented.

3.1 Verification

False Acceptance Rate, False Rejection Rate, Genuine Acceptance Rate, Equal Error Rate, Detection Error Trade-Off and Receiver Operating Characteristic are the metrics used in literature to evaluate a Verification system.

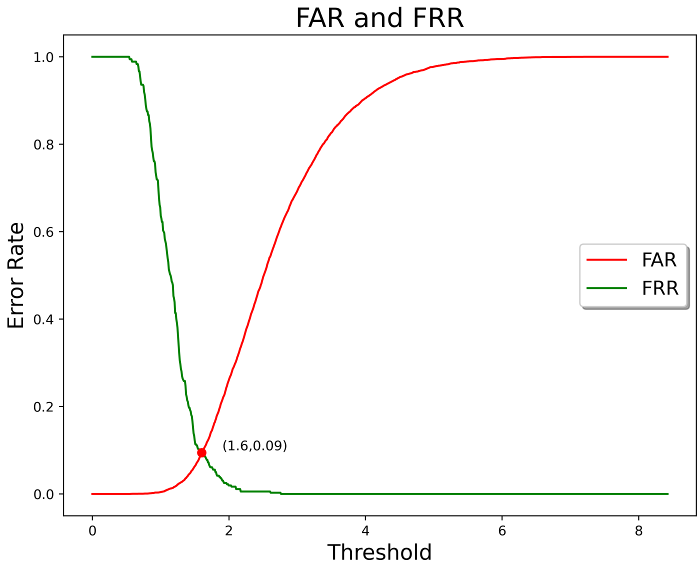
The **False Acceptance Rate** (FAR) is defined as the percentage of recognition operations with an impostor claim in which false acceptance occurs. It is the probability of an unauthorized person being identified as an authorized person.

The **False Rejection Rate** (FRR) is defined as the percentage of recognition operations with a genuine claim in which false rejection occurs. It is the probability of an authorized person being identified as an unauthorized person.

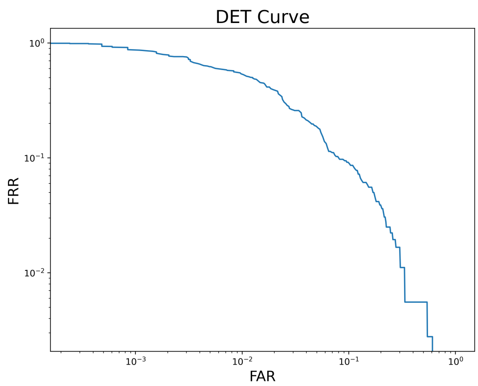
The **Equal Error Rate** (EER) is the error rate when FAR=FRR.

The **Detection Error Trade-Off** (DET) depicts the probability of False Reject (FRR) of the system vs False Accept Rate (FAR) variation.

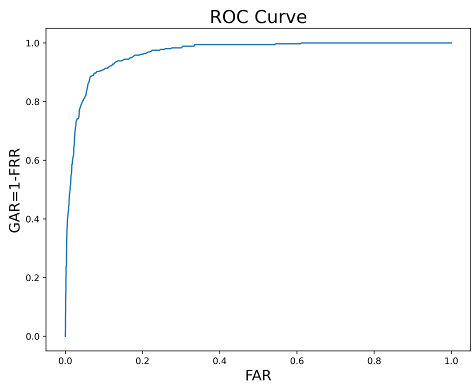
The **Receiver Operating Characteristic** (ROC) depicts the probability of Genuine Accept (GAR) of the system, expressed as 1-FRR, vs False Acceptance Rate (FAR) variation.



(a) FAR, FRR and EER.



(b) DET.



(c) ROC.

Figure 6: Results of verification evaluation.

3.2 Open-Set Identification

Detection and Identification Rate, False Acceptance Rate, False Rejection Rate and Receiver Operating Characteristic are the metrics used in literature to evaluate a Verification system.

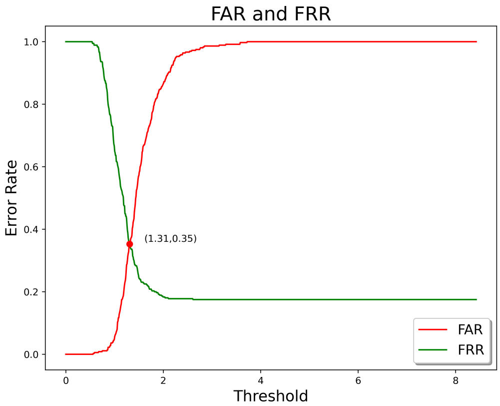
Let's define $rango(p_j)$ as the position in the list where the first template for the correct identity is returned.

The **Detection and Identification Rate** at rank k (DIR(t, k)) is the probability of correct identification at rank k (the correct subject is returned at position k), given by the ratio between the number of individuals correctly recognized within rank k and the number of probes belonging to individuals in the gallery.

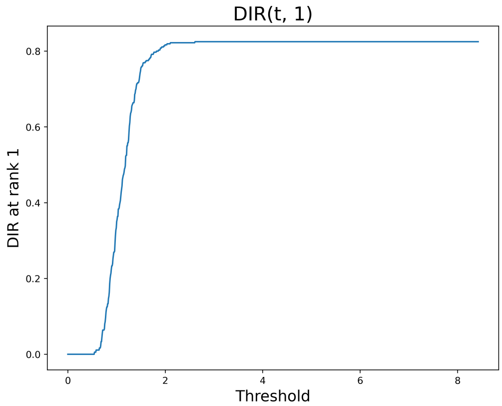
The **False Acceptance Rate** (FAR) is given by the number of times in which the system returns an incorrect alarm, running trials with probes belonging to subjects not in the gallery. That is, it is the ratio between the number of impostors recognized by mistake and the total number of impostors not in the gallery.

The **False Rejection Rate** (FRR) is the probability of false reject, at rank 1, expressed as 1-DIR(t, 1).

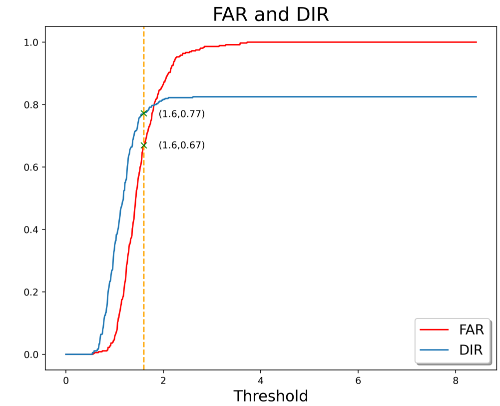
The **Open-set (Watchlist) Receiver Operating Characteristic** (ROC) depicts the probability of Detection and Identification Rate at rank 1 (DIR) of the system vs False Acceptance Rate (FAR) variation.



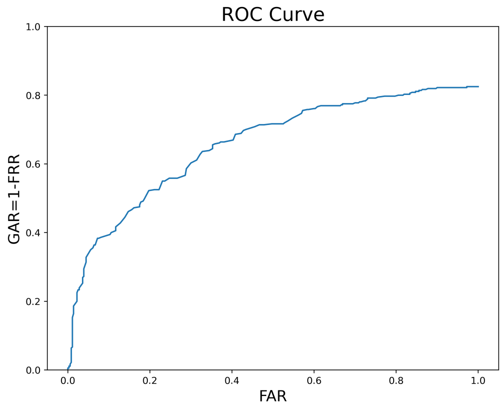
(a) FAR, FRR and EER.



(b) DIR(t, 1).



(c) FAR and DIR.



(d) ROC.

Figure 7: Results of verification evaluation.

3.3 Closed-Set Identification

Cumulative Match Score and Cumulative Match Characteristic are the metrics used in literature to evaluate a Closed-Set Identification system.

The **Cumulative Match Score** (CMS) at a certain rank k, can be defined as the probability of identification at that rank k. It judges the ranking capabilities of an identification system.
The Cumulative Match Score at rank 1 is also defined as Recognition Rate.
It is calculated by computing the ratio between the number of individuals correctly predicted among the number of ranks and the total number of individuals in the system.

The **Cumulative Match Characteristics** (CMC) curve visualizes the Cumulative Match Score for a certain number k of ranks. It reports the probability that the classifier predicts the correct class in the first k places.

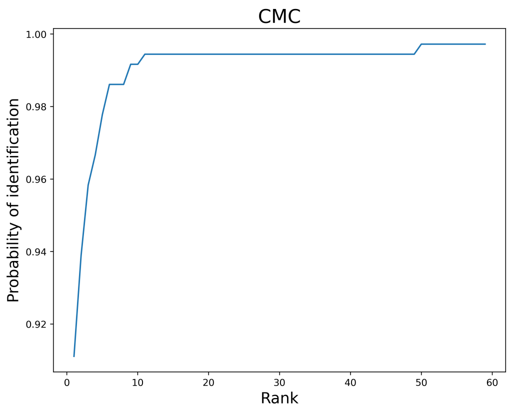


Figure 8: Results of closed-set identification evaluation.

3.4 Doddington zoo

Since every user has completely different interactions with each authentication system, several techniques have been developed in the literature to model distinctive users categories. Doddington zoo [7, 14] is a biometric menagerie that defines and labels user groups with animal species to reflect their behavior with the biometric systems. Eight categories of animals were defined, which are:

- *Sheep*: concerns users who can easily be recognized;
- *Goats*: represents users who are particularly difficult to recognize;
- *Lambs*: contains users who are easy to imitate;
- *Wolves*: consists of users who can easily imitate others;
- *Chameleons*: correspond to users who are easy to recognize and easy to attack;
- *Phantoms*: depict the users characterized by rejections of genuine and impostor queries;
- *Doves*: represent the best users because they are easy to recognize and difficult to attack;
- *Worms*: regroup the worst users as they are difficult to recognize and easy to attack.

The classification is based on the Average Genuine Score and the Average Impostor Score.

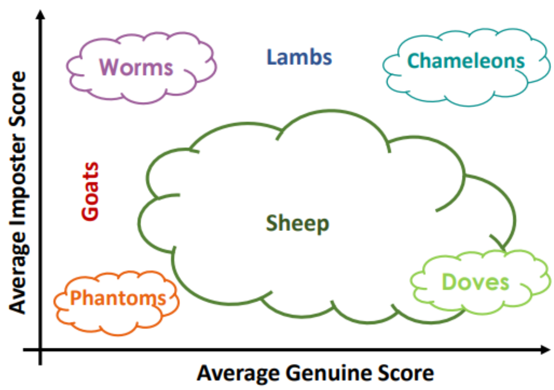


Figure 9: Animals of the Doddington zoo Biometric Menagerie

This menagerie was developed for different biometric modalities and we used it to understand the behavior of the individuals in our dataset during the verification phase. Our classification is based on the error rates of the system, which are the False Acceptance Rate (FAR) and the False Rejection Rate (FRR).

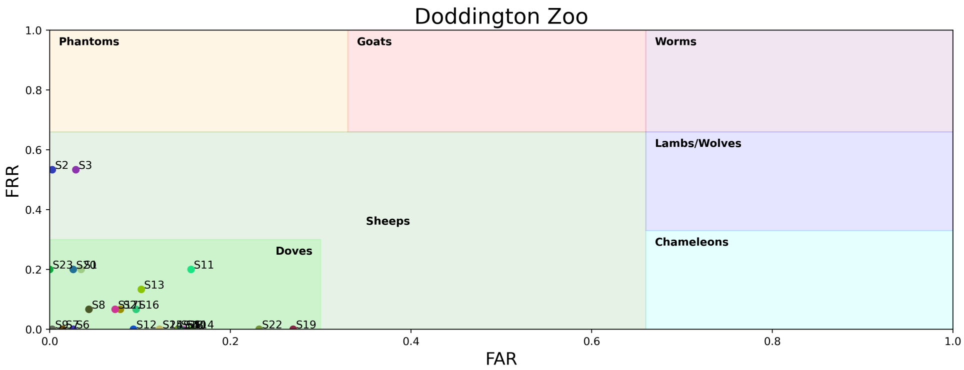


Figure 10: Doddington zoo results with division into classes.

As can be seen in figure 10, our dataset consists of only sheep and doves. There are no subjects with particularly abnormal behavior. We suspect that those results are due to the data contained in the dataset, which is not wide and varied enough. Figure 11 shows a close-up of our results:

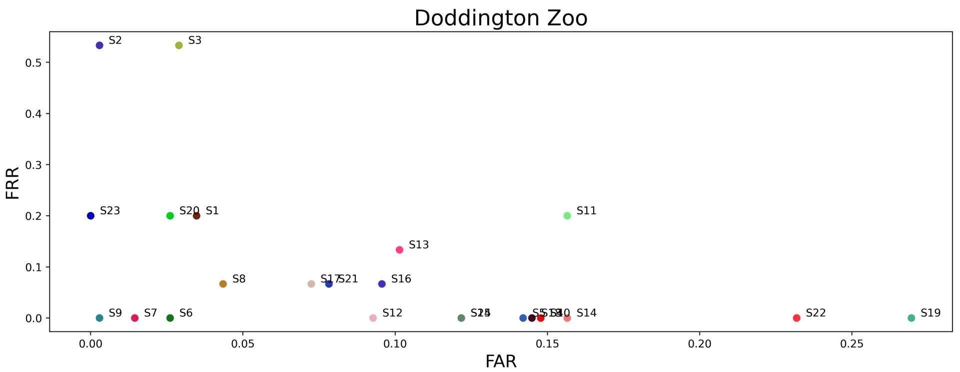


Figure 11: Close-up of Doddington Zoo results.

4 Demo

The demo consists of a python script on CLI.
A probe and its ground truth will be submitted. The system can perform the following operations:

- a verification operation by specifying the input probe and the claimed identity. At the end of this operation, it will print the outcome: "*Identity verified, you are user ID*" or "*Access denied*".
- an identification operation by specifying the input probe. At the end of this operation, it will print the outcome: "*You are user ID*" or "*No match in dataset*".

Note that for both verification and identification we set the *acceptance threshold* equal to 1.6.

5 Future Works

Since time was limited, we collected data from only 24 people. The size of the dataset can be expanded by continuing to acquire samples.
By integrating the WBB drivers into our software, the sampling can be independent from the use of *BrainBLOX*, making the execution flow unique. In this way, we can more easily expand the dataset and it will be possible to perform tests on a wider and bigger dataset.

In order to collect more detailed and informative data, we could add sensors or integrate our system with wearable devices such as a smartwatch. *BrainBLOX* offers an additional functionality, not yet mentioned, that could be useful for some possible future work. In fact, the software offers the possibility of integrating Wiimote controllers.

Since gait is a trait that is vulnerable to camouflaging, but difficult to spoof, a multi-biometric approach could be applied to reinforce the recognition system.

A possible evolution of our project could be the development of a "Smart Carpet", a system which is able to recognising people entering a public building, a private office or a home.

Beyond the biometric context, attention can be focused on possible medical and sporting developments. The WBB could be used for the classification of pathologies or disorders. Furthermore, there are countless sports in which balance is crucial and WBB is a device that would assist many athletes, both beginners and experts.

6 Conclusions

Considering that no one before us had tested the WBB for the recognition of individuals, we believe we have obtained fairly good results.
As shown in section 3, the verification results are great. The Equal Error Rate (EER) is very low, with a threshold of 1.6 the EER is equal to 0.10. On the other hand, identification's results report a high EER, which is also confirmed by the trend of the ROC Curve.
As expected, it is a system with a low threat of spoofing and a high risk of camouflage, characterized by a low FAR and an average FRR, respectively.
Therefore, we can confirm the initial premises on gait recognition requirements with a focus on the execution/operating speed. Indeed, it remains mediocre due to the attended and controlled settings, despite the fast and inexpensive computation of time series.

References

[1] Harrison L Bartlett, Lena H Ting, and Jeffrey T Bingham. Accuracy of force and center of pressure measures of the wii balance board. *Gait & posture*, 39(1):224–228, 2014.

[2] M Christ, Nils Braun, and J Neuffer. Tsfresh documentation, 2021. URL: `https://tsfresh.readthedocs.io/en/latest/`.

[3] Maximilian Christ, Andreas W Kempa-Liehr, and Michael Feindt. Distributed and parallel time series feature extraction for industrial big data applications. *arXiv preprint arXiv:1610.07717*, 2016.

[4] Ross A Clark, Adam L Bryant, Yonghao Pua, Paul McCrory, Kim Bennell, and Michael Hunt. Validity and reliability of the nintendo wii balance board for assessment of standing balance. *Gait & posture*, 31(3):307–310, 2010.

[5] Joseph Cooper, Alaa Ahmed, and Katrina Siegfried. Brainblox software readme documentation, 2017. URL: `https://www.colorado.edu/neuromechanics/sites/default/files/attached-files/software_readme_2017.pdf`.

[6] Maria De Marsico. *Biometric Systems Evaluation*, pages 1–6. Springer Berlin Heidelberg, Berlin, Heidelberg, 2019. `doi:10.1007/978-3-642-27739-9_1605-1`.

[7] George Doddington, Walter Liggett, Alvin Martin, Mark Przybocki, and Douglas Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. Technical report, National Inst of Standards and Technology Gaithersburg Md, 1998.

[8] Ryo Eguchi and Masaki Takahashi. Validity of the nintendo wii balance board for kinetic gait analysis. *Applied Sciences*, 8(2):285, 2018.

[9] Jeffrey D Holmes, Mary E Jenkins, Andrew M Johnson, Michael A Hunt, and Ross A Clark. Validity of the nintendo wii® balance board for the assessment of standing balance in parkinson's disease. *Clinical Rehabilitation*, 27(4):361–366, 2013.

[10] Julia M Leach, Martina Mancini, Robert J Peterka, Tamara L Hayes, and Fay B Horak. Validating and calibrating the nintendo wii balance board to derive reliable center of pressure measures. *Sensors*, 14(10):18244–18267, 2014.

[11] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. Analysis of doddington zoo classification for user dependent template update: Application to keystroke dynamics recognition. *Future Generation Computer Systems*, 97:210–218, 2019.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[13] Giacomo Severini, Sofia Straudi, Claudia Pavarelli, Marco Da Roit, Carlotta Martinuzzi, Laura Di Marco Pizzongolo, and Nino Basaglia. Use of nintendo wii balance board for posturographic analysis of multiple sclerosis patients with minimal balance impairment. *Journal of neuroengineering and rehabilitation*, 14:1–14, 2017.

[14] Neil Yager and Ted Dunstone. Worms, chameleons, phantoms and doves: New additions to the biometric menagerie. In *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, pages 1–6. IEEE, 2007.