Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd
https://doi.org/10.53070/bbd.989226

Journal of Computer Science
ISSN: 2548-1304
Volume: IDAP-2021, Issue: Special , pp: 116-125
@Anatolian Science

Gait based human identification: a comparative analysis

Kubilay Muhammed SÜNNETCİ *[1,3], Muhammed ORDU[2], Ahmet ALKAN[3]

[1]Department of Electrical and Electronics Engineering, Osmaniye Korkut Ata University, Osmaniye, Turkey

[2]Department of Industrial Engineering, Osmaniye Korkut Ata University, Osmaniye, Turkey

[3]Department of Electrical and Electronics Engineering, Kahramanmaraş Sütçü Imam University, Kahramanmaraş, Turkey

(kubilaysunnetci@osmaniye.edu.tr, muhammedordu@osmaniye.edu.tr, aalkan@ksu.edu.tr)

Received: Aug.31, 2021 Accepted: Sep.26,2021 Published: Oct.20, 2021

Abstract — Thanks to gait analysis, many examinations such as person identification, disease detection, and evaluation of neuromusculoskeletal system functions can be performed. In the study, the used dataset includes three different gait parameters obtained from 16 different individuals (7 females and 9 males) using wearable gait analysis sensors, and here there are 321 parameters for one gait of each person. In addition, we classify this data using Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 classifiers. Two different optimization techniques were employed to increase the performance metrics of the classifiers. From the results, it is seen that the Accuracy (%), Error (%), Sensitivity (%), Specificity (%), Precision (%), F1 Score (%), and Matthews Correlation Coefficient (MCC) of Optimizable Ensemble-2 that is the most successful classifier are equal to 97.92, 2.08, 97.92, 99.86, 98.44, 97.86, and 0.9790, respectively.

Keywords: *Classification, linear discriminant, ensemble classification, optimization, gait analysis.*

1. Introduction

Gait is a biological activity that arises with the simultaneous movement of the muscular and bone systems and that humans requires to change their locations. It is also a unique human behavior as some biological features of people (i.e., fingerprints, palm lines) (Gümüşçü, 2019). Due to gait is a physiological movement that can be affected by a number of health problems, gait analysis can provide critical results about patients in clinical settings (i.e., the success of surgeries or the effectiveness of rehabilitation) and provide an important support to decision makers (Caldas et al., 2017). Outside of the laboratory, there exit wearable devices that can be used for gait analysis during both walking and running.

Recently, gait analysis has been carried out for the different purposes in many areas by using data mining techniques. For example, amyotrophic lateral sclerosis diseases detected by (Alaskar and Hussain, 2018), Parkinsonism classified by (Ricciardi et al., 2019), and the Bone Mineral Density of Patients determined by (Recenti et al., 2020). (Del Din et al., 2019) aimed to detect gait disorders in the diagnosis of Parkinson's disease through wearable technologies. (Açıcı et al., 2017) used the random forest method to diagnose Parkinson's disease with gait analysis. In addition, the following problems can be solved using the parameters obtained from gait analysis: gender determination by (Ahad et al., 2020), human recognition by (Arivazhagan, 2017) and age estimation by (Lu and Tan, 2010), (Ahad et al., 2020) and (Pathan et al., 2021).

In the study, Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 classifiers are trained for human identification using 48 gait parameters from the UCI Machine Learning website. Here, three different gait parameters from 16 different individuals are obtained with the aid of a wearable sensor. We also use optimization techniques to increase the performance metrics of ensemble classifiers. Thus, we compare all the classifiers used in the study according to different evaluation metrics.

The study is organized as follows. Section 2 expresses dataset used, and Section 3 describes the proposed method, where we show that training parameters, optimizer options, and so on. Afterward, Section 4 interprets experimental results for Optimizable Ensemble-2 that is the most successful classifier. Finally, Section 5 and Section 6 present Results and Discussion and Conclusion, respectively.

2. Data

The data used in the study were obtained from the results of the experiment conducted by (Gümüşçü, 2019), and we use the dataset given in (Gait-dataset, https://archive.ics.uci.edu). The experimental environment is a walking track shown in Figure1. A total of 16 volunteers, 7 females and 9 males, walked 3 times in 3 tours each. Gait parameter values were recorded for each volunteer in 4 different categories given in Table 1.

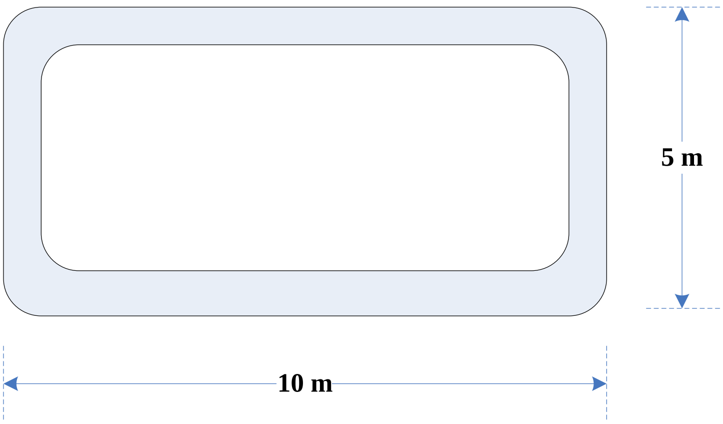


Figure 1: Walking track

Table 1: Walking Attribute Category (Gümüşçü, 2019)

Walking Attribute Category	Attributes
Basis Parameters	Speed, Variability, Symmetry
Temporary Parameters	Heel Press Time, Cycle Time, Cadence, Stance, Swing, Loading, Stepping, Pushing, Dual Support
Spatial Parameters	Stride Length, Stride Velocity, Peak Angle Velocity, Maximum Swing Velocity, Rotation Angle, Step Angle, Lift Angle, Swing Width, 3D Path Length
Height parameters	Maximum Heel Height, Maximum Toe Height, Minimum Toe Height

3. The Proposed Method

In the study, three different gait parameters from 16 different people were classified using five different classifiers. There are 321 features for each person in the dataset used in the study. For this aim, we have used five different classifiers namely Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2. Moreover, we have employed two different optimization techniques for the ensemble method used in the study. Herein, we compared five different classifiers in terms of classification performance metrics (Alkan and Günay, 2012-Solmaz et al., 2013-Zhou, 2009).

Table 2 shows some classification performance metrics and training parameters for the classifiers used in the study. From the table, we see that accuracy, total misclassification cost, and training time values of Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 classifiers are equal to 83.33%-87.50%-91.67-95.83%-97.92%, 8-6-4-2-1, and 9.0381-61.371-60.454-884.78-396.87 sec.

Table 2. Training parameters and classification performance metrics for classifiers used in the study

	Accuracy (%)	Total Misclassification Cost	Prediction Speed ~obs/sec	Training Time sec	Model Type
Linear Discriminant	83.33	8	150	9.0381	Covariance Structure: Full
Ensemble Subspace Discriminant	87.50	6	70	61.371	Number of learners:30, Subspace dimension:161
Ensemble Bagged Trees	91.67	4	110	60.454	Maximum number of splits:47, Number of learner:30
Optimizable Ensemble-1	95.83	2	23	884.78	Optimizable
Optimizable Ensemble-2	97.92	1	32	396.87	Optimizable

The computer used in the study has Windows10 Intel (R) Core (TM) i5-6400 CPU @2.70 GHz 2.71 GHz 8 GB RAM 64-bit operating system, and we use MATLAB R2020b. Herein, the table shows that Optimizable Ensemble-2 is the most successful classifier in terms of accuracy. However, it is seen that the most successful systems in terms of training time are the classifiers that do not use the optimization technique. We have experienced that its classification performance metrics increase when we make use of optimization techniques for this dataset. Therefore, for Optimizable Ensemble-1 and Optimizable Ensemble-2 classifiers, Table 3 is given as follows:

In Table 3, optimizer options and optimized hyper-parameters results are given for Optimizable Ensemble-1 and Optimizable Ensemble-2. From the table, it is seen that the Bayesian optimization technique is used for both classifiers. Yet, the acquisition function used for Optimizable Ensemble-1 and Optimizable Ensemble-2 is defined as the probability of improvement and expected improvement per second plus, respectively.

Table 3. Optimizer options and optimized hyper-parameters results for ensemble classifier

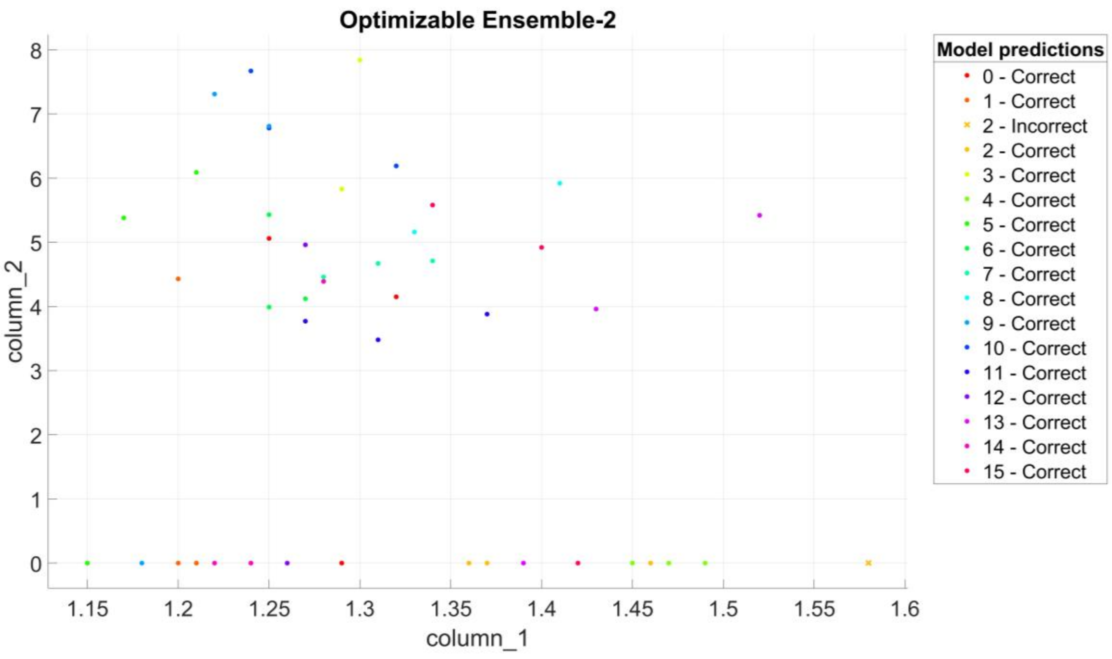
	Optimizable Ensemble-1	Optimizable Ensemble-2
Optimizer Options	Optimizer: Bayesian optimization	
	Acquisition function: Probability of improvement	Acquisition function: Expected improvement per second plus
	Learning rate: 0.001-1	
	Training time limit: False	
	Iterations: 30	
Optimized Hyper-Parameters Results	Ensemble method: Bag	Ensemble method: Bag
	Maximum number of splits: 27	Maximum number of splits: 41
	Number of learners: 395	Number of learners: 248
	Number of predictors to sample: 61	Number of predictors to sample: 7

The learning rate, training time limit, iterations, and ensemble method for both techniques were determined as 0.001-1, false, 30, and bag, respectively. When we examined the optimized hyper-parameters results, we see that the maximum number of splits, number of learners, and number of predictors to sample for Optimizable Ensemble-1 and Optimizable Ensemble-2 are 27-395-61 and 41-248-7, respectively.

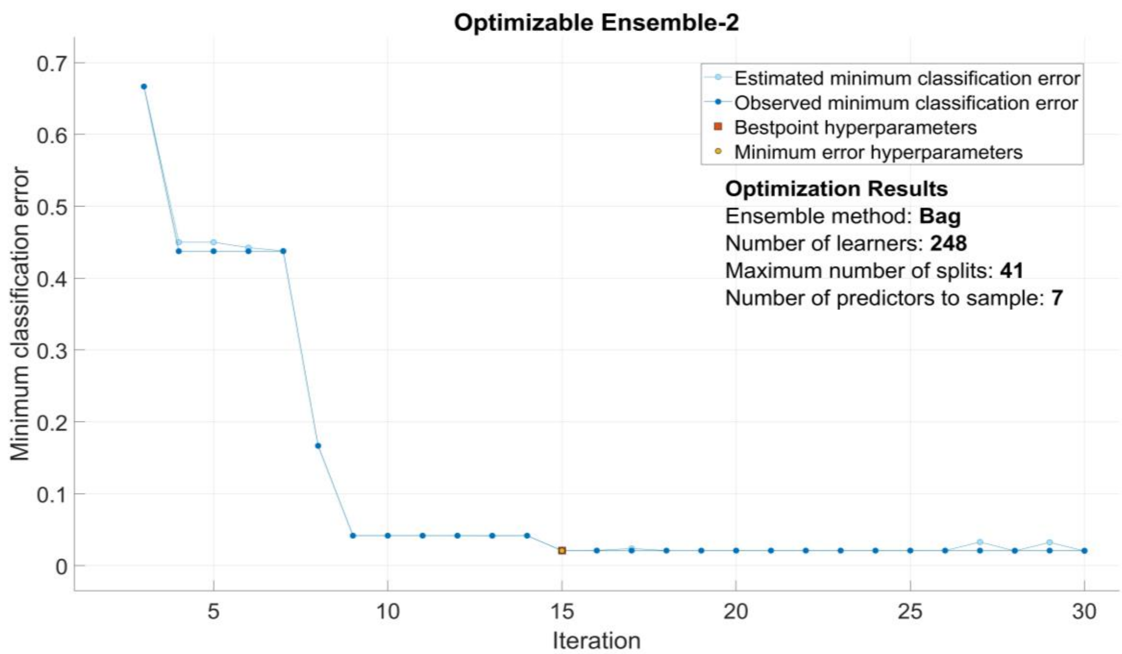
4. Experimental Results

In this section, scatter plot, minimum classification error plot, Receiver Operating Characteristic (ROC) curve for positive class:0, confusion matrix (number of observation, True Positive Rates (TPR)-False Negative Rates (FNR), and Positive Predictive Values (PPV)-False Discovery Rates (FDR)). In addition, Figure 2 and Figure 3 are obtained as follows using 10-fold cross-validation for the dataset used in the study.

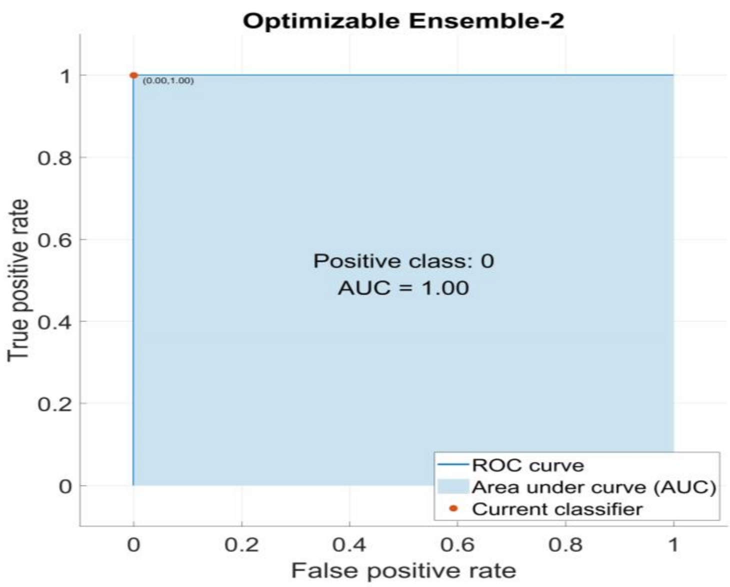
In Figure 2.a-c, 'scatter plot', 'minimum classification error plot', 'Receiver Operating Characteristic (ROC) curve for positive class:0' are given for Optimizable Ensemble-2. When we examine Figure 2.a, it is seen that one of the 48 gaits obtained from sixteen different people is incorrect, and here the model prediction is incorrect for the third person. Likewise, Figure 2.b shows that the best-point and minimum error parameters obtained for this classifier are determined in the 15th iteration.


Optimizable Ensemble-2
Model predictions
0 - Correct
1 - Correct
2 - Incorrect
2 - Correct
3 - Correct
4 - Correct
5 - Correct
6 - Correct
7 - Correct
8 - Correct
9 - Correct
10 - Correct
11 - Correct
12 - Correct
13 - Correct
14 - Correct
15 - Correct
column_2
column_1

(a)


Optimizable Ensemble-2
Estimated minimum classification error
Observed minimum classification error
Bestpoint hyperparameters
Minimum error hyperparameters
Optimization Results
Ensemble method: Bag
Number of learners: 248
Maximum number of splits: 41
Number of predictors to sample: 7
Minimum classification error
Iteration

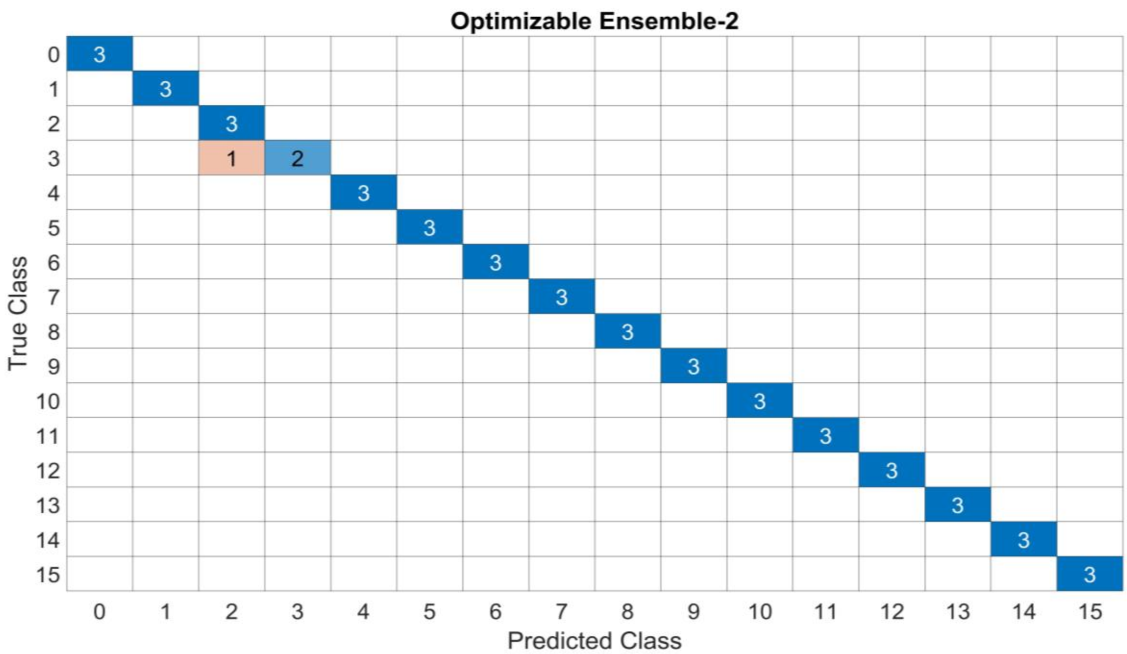
(b)



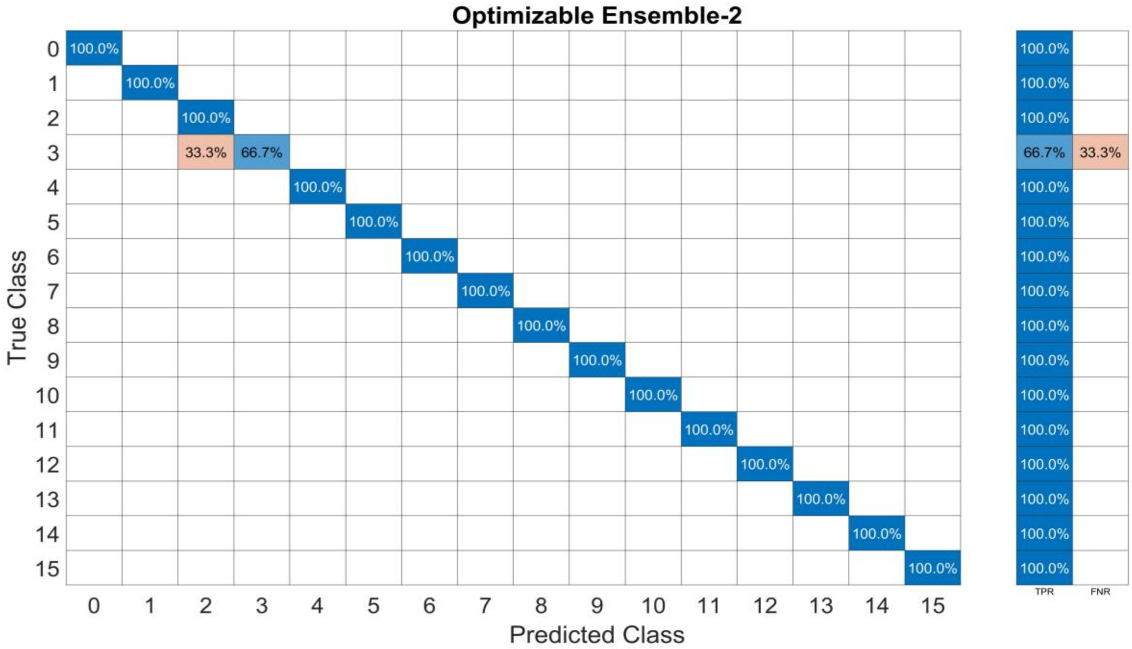
(c)

Figure 2. a) Scatter plot, b) minimum classification error plot, c) Receiver Operating Characteristic (ROC) curve for Optimizable Ensemble-2

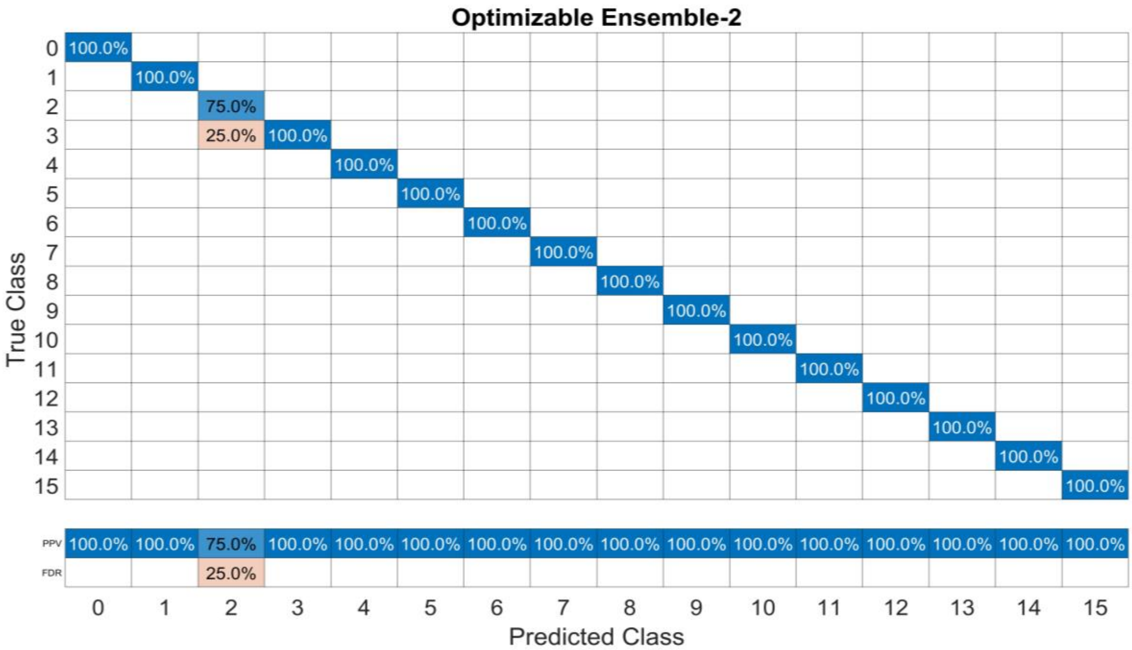
From the best optimization results obtained, it is seen that the ensemble method, the maximum number of splits, the number of learners, and the number of predictors to sample are determined as Bag, 41, 248, and 7, respectively. On the other hand, when examining the ROC curve of Optimizable Ensemble-2, Figure 2.c indicates that the Area Under Curve (AUC) for this classifier is equal to 1. Thus, it can be understood that the learning process for the Optimizable Ensemble-2 has taken place successfully.



(a)



(b)



(c)

Figure 3. Confusion matrices for Optimizable Ensemble-2

Confusion matrices for Optimizable Ensemble-2, the most successful classifier, are given in Figure 3. Here, true class and predicted class are shown in Figure 3.a, and they represent the number of observations. On the other hand, Figure 3.b-c shows confusion matrices according to TPR-FNR and PPV-FDR. From Figure 3.b, the TPR calculated for all individuals except for the 3rd volunteer is equal to 100%, and this value is equal to 66.7% for wrongly identified persons. From Figure 3.c, the PPV obtained for all individuals except for this person is equal to 100%, and this value is equal to 75% for person 2.

5. Results and Discussion

In this study, we classify three different gait parameters from 16 individuals for human identification, and there are 321 features for each volunteer. Afterward, we train classifiers Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 using 10-fold cross-validation. Herein, we use two different optimization techniques to improve the classification performance metrics of ensemble classifiers. In addition, we give the performance metrics of the classifiers used in the study as follows:

Table 4. Classification performance metrics for classifiers used in the study

	Linear Discriminant	Ensemble Subspace Discriminant	Ensemble Bagged Trees	Optimizable Ensemble-1	Optimizable Ensemble-2
Accuracy (%)	83.33	87.50	91.67	95.83	97.92
Error (%)	16.67	12.50	8.33	4.17	2.08
Sensitivity (%)	83.33	87.50	91.67	95.83	97.92
Specificity (%)	98.89	99.17	99.44	99.72	99.86
Precision (%)	86.46	91.67	93.75	96.88	98.44
F_1 Score (%)	82.59	87.80	91.43	95.71	97.86
MCC	0.8276	0.8799	0.9160	0.9580	0.9790

Table 4 shows the classification performance metrics calculated for the classifiers used in the study. From the table, it can be seen that Accuracy (%), Error (%), Sensitivity (%), Specificity (%), Precision (%), F_1 Score (%), and MCC of the classifiers Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 are equal to 83.33-87.50-91.67-95.83-97.92, 16.67-12.50-8.33-4.17-2.08, 83.33-87.50-91.67-95.83-97.92, 98.89-99.17-99.44-99.72-99.86, 86.46-91.67-93.75-96.88-98.44, 82.59-87.80-91.43-95.71-97.86, and 0.8276-0.8799-0.9160-0.9580-0.9790, respectively. Here, we see that the most successful classifier in terms of classification performance metrics is Optimizable Ensemble-2.

In the study given in (Gümüşçü et al., 2018), it is aimed to determine the gender by extracting 321 features for a total of 50 volunteers, where there are 23 females and 27 males. Here, Support Vector Machine, K-Nearest Neighbor, and Decision Tree classifiers are used, and their accuracy is equal to 84%, 68%, and 84%, respectively. Similarly, it is aimed to detect the human for a total of 16 volunteers, where there are 7 females and 9 males, and the K-Nearest Neighbor classifier is used in this study. From the results, it is seen that the accuracy of the proposed system is 97.9% (Gümüşçü, 2019).

6. Conclusion

In this study, it is aimed to identify the human using 48 gait parameters, consisting of 16 volunteers, 7 females and 9 males. To make a comparative study five classifiers namely; Linear Discriminant, Ensemble Subspace Discriminant, Ensemble Bagged Trees, Optimizable Ensemble-1, and Optimizable Ensemble-2 classifiers were employed. Herein, we have used two different optimization techniques by chancing the acquisition function for ensemble classifiers. Thus, five different classifiers used in the study were compared in terms of classification performance metrics. Using the gait analysis from the performance indicators obtained, it was seen that the highest human identification performance was obtained with the Optimizable Ensemble-2.

References

Açıcı K., Erdaş Ç.B., Aşuroğlu T., Toprak M.K., Erdem H., Oğul H. (2017) A Random Forest Method to Detect Parkinson's Disease via Gait Analysis. In: Boracchi G., Iliadis L., Jayne C., Likas A. (eds) Engineering Applications of Neural Networks. EANN 2017. Communications in Computer and Information Science, vol 744. Springer, Cham. https://doi.org/10.1007/978-3-319-65172-9_51

Ahad, M. A. R., Ngo, T. T., Antar, A. D., Ahmed, M., Hossain, T., Muramatsu, D., ... & Yagi, Y. (2020). Wearable sensor-based gait analysis for age and gender estimation. Sensors, 20(8), 2424. https://doi.org/10.3390/s20082424

Alaskar H., Jaafar Hussain A. (2018) Data Mining to Support the Discrimination of Amyotrophic Lateral Sclerosis Diseases Based on Gait Analysis. In: Huang DS., Gromiha M., Han K., Hussain A. (eds) Intelligent Computing Methodologies. ICIC 2018. Lecture Notes in Computer Science, vol 10956. Springer, Cham. https://doi.org/10.1007/978-3-319-95957-3_80

Alkan, A., & Günay, M. (2012). Identification of EMG signals using discriminant analysis and SVM classifier, Expert Systems with Applications, 39(1), 44–47. https://doi.org/10.1016/j.eswa.2011.06.043

Arivazhagan S, Induja P. Gait Recognition-Based Human Identification and Gender Classification. Proceedings of International Conference on Computer Vision and Image Processing; 2017; pp. 533-544.

Caldas, R., Mundt, M., Potthast, W., Lima Neto, F. B., Markert, B. (2017). A systematic review of gait analysis methods based on inertial sensors andadaptive algorithms. Gait & Posture, 57, 204-210. https://doi.org/10.1016/j.gaitpost.2017.06.019

Dataset, https://archive.ics.uci.edu/ml/datasets/Gait+Classification, Gait Classification Data Set, Dr. Abdulkadir Gumuscu,

Del Din, S., Elshehabi, M., Galna, B., Hobert, M. A., Warmerdam, E., Suenkel, U., Brockmann, K., Metzger, F., Hansen, C., Berg, D., Rochester, L., Maetzler, W. (2019). Gait Analysis with Wearables Predicts Conversion to Parkinson Disease. Annals of Neurology, 86(3), 357-367. https://doi.org/10.1002/ana.25548

Gümüşçü, A. (2019). Giyilebilir Yürüyüş Analiz Sensörü ile Kişi Sınıflandırmasının Öznitelik Seçme Algoritmaları ile İyileştirilmesi. Fırat Üniversitesi Müh. Bil. Dergisi, 31(2), 463-471.

Lu JW, Tan YP. Gait-Based Human Age Estimation. Ieee Transactions on Information Forensics and Security 2010; 5(4): 761-770.

Pathan R.K., Uddin M.A., Nahar N., Ara F., Hossain M.S., Andersson K. (2021) Human Age Estimation Using Deep Learning from Gait Data. In: Mahmud M., Kaiser M.S., Kasabov N., Iftekharuddin K., Zhong N. (eds) Applied Intelligence and Informatics. AII 2021. Communications in Computer and Information Science, vol 1435. Springer, Cham. https://doi.org/10.1007/978-3-030-82269-9_22

Recenti, M., Ricciardi, C., Aubonnet, R., Esposito, L., Jónsson, H., & Gargiulo, P. (2020, June). A regression approach to assess bone mineral density of patients undergoing total hip arthroplasty through gait analysis. In 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE.

Ricciardi, C., Amboni, M., De Santis, C., Improta, G., Volpe, G., Iuppariello, L., ... & Unit, T. B. E. (2019). Using gait analysis' parameters to classify Parkinsonism: A data mining approach. Computer methods and programs in biomedicine, 180, 105033. https://doi.org/10.1016/j.cmpb.2019.105033

Solmaz, R., Günay, M., Alkan (2013). Uzman Sistemlerin Tiroit Teşhisinde Kullanılması, Akademik Bilişim, 919–922. http://ab.org.tr/ab13/kitap/olmaz_gunay_AB13.pdf

Zhou, Z.-H. (2009). Ensemble Learning, Encyclopedia of Biometrics, 270–273. https://doi.org/10.1007/978-0-387-73003-5_293