BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 14, 2026
Research Article

Explainable Parkinson's Disease Detection from Keystroke Dynamics Using Optimized AdaBoost

Ferdi Ozbilgin[1]

***Abstract*— Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting motor functions, and its early detection remains a significant challenge. This study proposes an explainable machine learning framework for the detection of Parkinson's disease using keystroke dynamics. A total of 33 statistical features derived from hold time, flight time, and latency were extracted from the publicly available Tappy Keystroke dataset. An AdaBoost classifier optimized via Bayesian optimization was employed and compared with several machine learning models, including Decision Tree, Naive Bayes, K-Nearest Neighbor, Artificial Neural Network, RUSBoost, and Bagging. The results demonstrate that the proposed approach achieves the highest performance with an accuracy of 87.16% and an AUC of 0.9129. Furthermore, SHAP-based analysis was used to interpret model decisions, revealing that variation-based features related to Hold Time and Flight Time play a critical role in classification. The findings indicate that keystroke dynamics can serve as a non-invasive and scalable digital biomarker for PD detection.**

Index Terms— **AdaBoost ensemble learning, Bayesian optimization, explainable AI, keystroke dynamics, Parkinson's disease, SHAP**

I. Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by the gradual loss of dopaminergic neurons in the substantia nigra. The disease was first documented in 1817 by the English physician James Parkinson, who described it as "shaking palsy" [1]. Following Alzheimer's disease, PD is the second most common neurodegenerative disorder and mainly occurs in individuals over the age of 60, affecting roughly 1% of this population [2]. It is estimated that more than 10 million individuals worldwide live with the disease, and this number is projected to reach 25.2 million by 2050 [3]. In Türkiye, prevalence studies indicate that PD affects approximately 111 individuals per 100,000 people [4]. Progression of the disease and the aging of the population make it a growing social problem. Diagnosing the disease early and accurately is therefore a matter of great concern for the quality of life of patients as well as for health care systems.

Initially the diagnosis of PD is carried out by a specialist neurologist based on information collected through an examination of the patient's motor activity composed of bradykinesia (slowness of movement) associated with either tremor at rest or rigidity in their limbs [5]. The onset of motor symptoms may occur years before they are clinically recognized, significantly delaying the official diagnosis. From research data this condition occurs when about 60-80% of the neurons in the substantia nigra (PD's primary affected area of the brain) have been destroyed [6]. Recent studies aim to utilize novel biomarkers that show that some of the behavioral characteristics of the motor-symptoms associated with PD can be measured and thus possible to be used for the purposes of a possible pre-diagnostic test. Examples of such characteristics can include voice analysis such as e.g. voice-tremors, breath-insufficiency and phonetic-unpreciseness, or examples of gait-analysis based on motion-sensors such as e.g. length of steps, the degree of asymmetry in the movements and the variability in the swing phase or examples of keystroke analysis of keyboard use for detecting characteristics in the timing of key-presses and key-releases (in this case the characterizes represents characteristics of the fine motor skills required when handling a keyboard) [7, 8]. It has become widely acknowledged throughout the world in the medical and scientific communities that Machine Learning (ML) and Deep Learning (DL) have immense opportunities in the field of medicine. Research which utilizes AI for pre-diagnostic tests has in recent years seen great progress and numerous examples of high-performance have been reported on different diseases [9, 10]. New research is now emerging rapidly, particularly in the field of signal-analysis based on machine learning. As of today there is an increasing amount of research focusing on finding alternative ways of diagnosing PD which will make the current clinical diagnosis carried out by specialist neurologists possible to replace by new non-invasive AI-based pre-diagnostic tests, such as those described in the paper by means of signal-biomarkers and ML.

Keystroke dynamics refers to the temporal patterns of typing, specifically key press duration and spacing between keys. These metrics are considered as digital biomarkers to non-invasively and objectively monitor motor changes. Alfalahi et al. carried out a systematic review and meta-analysis of 41 works that validated that keystroke dynamics metrics have moderate to high accuracy in identifying motor decline associated with PD [11]. Initial works were mainly conducted in laboratory settings by requiring participants to type predefined phrases and achieved significant results by applying conventional machine learning algorithms such as SVM, Decision Trees, Random Forest and other ensemble techniques [12]. Later works attempted to generalize the results to unconstrained home

[1] **Ferdi Ozbilgin**, is with Department of Electrical Electronics Engineering Giresun University, Giresun, Türkiye, (e-mail: ferdi.ozbilgin@giresun.edu.tr). https://orcid.org/0000-0003-4946-7018
Manuscript received Mar 9, 2026; accepted Apr 13, 2026. DOI: 10.17694/bajece.1905893
Ozbilgin, F., (2026). Explainable Parkinson's Disease Detection from Keystroke Dynamics Using Optimized AdaBoost. Balkan Journal of Electrical and Computer Engineering, 14(1), 225-231.

environment and free-text typing scenarios. Tripathi et al. carried out a work on keystroke dynamics features by using a set of 281 participants in a home environment study using convolutional neural networks and obtained promising results for the detection of PD [13]. Roy et al. dealt with the issue of class imbalance and proposed a new ensemble learning approach that exploits both fixed-text and free-text keystroke dynamics features, thereby improving the performance of models on imbalanced medical datasets [14]. Tripathi et al. carried out further work on keystroke dynamics and proposed the application of self-supervised learning for reducing the dependency on manual labeling of data and for extracting transferable features across datasets [15]. In the most recent works, deep learning models and multi-dataset pretraining are proposed and validated in the context of keystroke dynamics analysis for the assessment of motor changes in clinical scale as well as for continuous monitoring of patients [16].

Despite the significant work on keystroke dynamics, relatively few studies have focused on ensemble learning methods for keystroke dynamics along with hyperparameter tuning and the usage of explainable AI (XAI) techniques. In order to help bridging this gap, we used a Bayesian optimization–based AdaBoost algorithm on the well-known Tappy Keystroke dataset for authentication, and then compared it with some other machine learning classifiers. Furthermore, the decisions made by the classifier that yields the best performance is analyzed using SHAP. The potential contributions of this study are summarised in the following sub-list.

- Analysis of the diagnostic value of 33 features extracted from keystroke dynamics for the detection of PD.
- This research work evaluates and compares the performance of AdaBoost with Bayesian optimization and multiple classification algorithms within a structured experimental framework.
- Explanation of model behavior through SHAP, enabling greater model interpretability.

II. Materials and Methods

This section covers the Data set, feature extraction, classification techniques and performance metrics. The results are obtained by using publicly available Tappy Keystroke dataset for feature extraction and then by applying several machine learning algorithms on the extracted features. The used algorithms are thoroughly examined and the results are obtained mainly by using AdaBoost as a classification technique. In addition, model decisions are analyzed using SHAP.

A. Dataset

This research employed the public dataset of the Tappy Keystroke which is available in the PhysioNet repository [17, 18]. The dataset includes 217 participants from the United States of America, Canada, United Kingdom and Australia. 162 participants (75%) were diagnosed with Parkinson's disease, while 55 participants (25%) were classified as non-Parkinson's (healthy controls). the dataset includes 116 male participants (53%) and 101 female participants (47%), ensuring a balanced representation of both genders. The participants did not undergo any supervised activity and the data was collected passively in a home setting during normal computer use. The dataset consists of two main components: user files that include information about the demographics and clinical profiles of the users and the data files that include the keystroke recordings of the participants' typing activities with the inclusion of hand preference which is either left or right hand, the Hold Time, the Direction of the movement and the Latency and the Flight Time of the keystroke recordings which were all recorded in milliseconds [18].

B. Feature Extraction

The raw keystroke data for all participants were used to derive the 33 features. A detailed description of the feature set is provided in the Materials section. Briefly, these features were calculated from statistics of Hold Time, Latency and Flight Time parameters, separated based on the dominant hand condition and handedness asymmetry measures, transition times based on key direction and average typing speed.

For the Hold Time parameter, which represents the duration between pressing and releasing a key, seven statistical features were calculated: mean, standard deviation, median, coefficient of variation (CV), interquartile range (IQR), maximum, and minimum. For the Latency parameter, representing the time difference between two consecutive key presses, and the Flight Time parameter, representing the interval between releasing one key and pressing the next, five statistical features were derived for each parameter, excluding maximum and minimum values. In addition, the mean values of these three parameters corresponding to left-hand and right-hand keystrokes, along with the standard deviations of Hold Time, were computed separately for each hand, resulting in eight hand-based features. To reflect motor asymmetry between the two hands, asymmetry indices were calculated for Hold Time, Flight Time, and Latency. Furthermore, the average Latency values corresponding to successive key transition directions—left-to-left (LL), left-to-right (LR), right-to-left (RL), and right-to-right (RR)—were included as four additional features. Finally, overall typing speed, defined as the number of keystrokes per unit time, was also incorporated into the feature set. The extracted features are presented in Table I according to their respective groups.

Table I. Feature groups and their contents

Feature Group	Number of Features	Features
Hold Time	7	HoldTime_Mean, HoldTime_Std, HoldTime_Median, HoldTime_CV, HoldTime_IQR, HoldTime_Max, HoldTime_Min
Latency	5	Latency_Mean, Latency_Std, Latency_Median, Latency_CV, Latency_IQR
Flight Time	5	FlightTime_Mean, FlightTime_Std, FlightTime_Median, FlightTime_CV, FlightTime_IQR
Hand-Base	8	Hold_L, Flight_L, Latency_L, Hold_L_Std, Hold_R, Flight_R, Latency_R, Hold_R_Std
Asymmetry	3	HoldAsym, FlightAsym, LatencyAsym
Directional Latency	4	Lat_LL, Lat_LR, Lat_RL, Lat_RR
Typing Speed	1	TypingSpeed

C. Classification

In this work, the AdaBoost (Adaptive Boosting) algorithm has been used as the classifier [19, 20]. The main reason behind this selection is that a strong classifier can be formed from a set of weak learners by a series of steps through AdaBoost; also, it improves the classifier decision boundary in each step by

assigning more weights to the misclassified samples [21]. Other classifiers such as Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Artificial Neural Network (NN), RUSBoost and Bagging were implemented as well under the same conditions and were compared with the proposed approach. AdaBoost is an ensemble learning method that uses a series of weak learners. In the first step, the equal weight is assigned to the training samples [20]. Initially, equal weights are assigned to each training sample. For the weak classifier $h_t(x)$ trained at the t-th iteration, the classification error is computed as follows [19, 20]:

$$\varepsilon_t = \sum_{i=1}^{N} w_i I(y_i \neq h_t(x_i)) \quad (1)$$

Here, I(·) denotes the indicator function. Subsequently, the weight of the corresponding weak learner is determined as follows:

$$\alpha_t = \frac{1}{2} ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \quad (2)$$

In the next step, the sample weights are updated as follows:

$$w_i \leftarrow w_i \exp(-\alpha_t y_i h_t(x_i)) \quad (3)$$

The weights are then normalized so that their sum equals 1. The final classifier is obtained as the weighted combination of all weak learners, expressed as follows:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t \, h_t(x)\right) \quad (4)$$

Through its iterative reweighting mechanism, AdaBoost updates the classifier by focusing more on misclassified samples at each iteration. Our classifier is more robust to many issues that are common in biomedical data such as the class imbalance and sample heterogeneity. The hyperparameters obtained from the Bayesian optimization method are used to evaluate the performance of AdaBoost classifier on all the prepared data sets by calculating the following metrics: accuracy, precision, recall, F1-score and AUC.

D. Performance Evaluation Metrics

The performance of the proposed classification models was evaluated using several widely adopted metrics, including Accuracy, Error Rate, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC). These metrics were derived from the confusion matrix components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The mathematical expressions and descriptions of the performance metrics are provided below [22, 23].

Accuracy is defined as the proportion of correctly classified samples relative to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Error Rate represents the proportion of samples that are incorrectly classified among all observations.

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN} \quad (6)$$

Precision refers to the proportion of instances predicted as positive that are actually positive.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall, also known as the True Positive Rate (TPR), indicates the proportion of actual positive instances that are correctly identified by the model.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

The F1-score is the harmonic mean of precision and recall and provides a balanced evaluation metric, particularly in cases where class imbalance is present.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

The discriminative performance of the models was evaluated using the ROC curve and the AUC. The ROC curve depicts the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across varying decision thresholds. An AUC value between 0 and 1 is obtained. A value close to one is highly desirable and implies good discrimination, a value close to 0.5 is indicative of no better than random discrimination.

SHAP analysis was then performed on the top performing model to gain insights on how the model is making predictions. SHAP analysis tells us the importance of each feature in the model's output and therefore helps us to identify the most contributing features. In addition to the accuracy of the model, explainability is a very important criterion especially for biomedical applications. The SHAP analysis will not only give us the importance of the contributing features but it will also show us the positive and negative effects of the contributing features to the classification.

III. RESULTS AND DISCUSSION

In this section, the results obtained for the prediction of Parkinson's disease using keystroke dynamics are presented. A total of 33 features were extracted from the keystroke data and used to train and test several machine learning models. For model evaluation, 20% of the data was reserved as a hold-out test set, while the remaining 80% was used for training. To enhance model robustness and reduce overfitting, a 5-fold cross-validation strategy was applied to the training data. The test set was kept fixed throughout the evaluation process to ensure an unbiased performance assessment. The performance of the classifiers was evaluated using commonly adopted metrics, including accuracy, precision, recall, F1-score, and AUC. All analyses in this study were conducted using MATLAB 2025b. The experimental setup consisted of a computer equipped with a 13th-generation Intel Core i7-13700H processor (2.40 GHz),

32 GB of RAM, and an 8 GB NVIDIA GeForce RTX 4070 GPU.

The MATLAB Classification Learner Toolbox was used to determine the model hyperparameters. Hyperparameter optimization was performed using the Bayesian Optimization approach and was configured to run for a maximum of 30 iterations. The hyperparameters optimized for each classification model and their corresponding search ranges are presented in Table II. The best values obtained after 30 iterations can be found in the "Value" column and were actually used during the final training and evaluation of the respective models. This way hyperparameter optimization was performed systematically and in a similar manner for all models.

Table II. Optimized hyperparameters and search ranges for the classification models

Model	Hyperparameter	Value	Search Range
DT	Maximum number of splits	19	1–436
	Split criterion	Gini	Gini, Maximum deviance reduction
NB	Distribution type	Kernel	Gaussian, Kernel
	Kernel type	Gaussian	Gaussian, Box, Epanechnikov, Triangle
KNN	Number of neighbors (k)	8	1–219
	Distance metric	Correlation	City Block, Chebyshev, Korelasyon, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski, Spearman
	Distance weight	Ters	Equal, Inverse, Squared inverse
NN	Number of layers	2	1–3
	Layer sizes	89, 297	1–300
	Activation	None	ReLU, Tanh, Sigmoid, None
	Lambda	0,706	$2{,}29\times10^{-8}$–228,83
RUSBoost	Number of learners	498	10–500
	Learning rate	0,991	0,001–1
	Maximum number of splits	13	1–436
Bag	Number of learners	186	10–500
	Maximum number of splits	108	1–436
AdaBoost	Number of learners	463	10–500
	Learning rate	0,553	0,001–1
	Maximum number of splits	21	1–436

Table III shows the confusion matrices of the test data for the models developed for the detection of PD through keystroke dynamics: DT, NB, KNN, NN, RUSBoost, Bagging, and AdaBoost. The true classes (Control and Parkinson) are placed in the rows and the predicted classes are placed in the columns of the confusion matrices. The diagonal elements represent the true positive values while the off-diagonal elements show the false positive values.

By examining the results, we can see that the ensemble methods namely the Bagging, AdaBoost and the RUSBoost have produced high accuracy of the Parkinson class and are minimizing the false negatives. At the same time, some of the individual models are prone to more false classification of control class samples. In case of the NN, all the control samples were misclassified as that of Parkinson which clearly indicates the effect of class imbalance. Also, the confusion matrices which we derived using our training dataset showed about the same trends and the performance distribution was about the same as what we had on the test dataset.

Table III. Confusion matrices of different machine learning models

Model	TP	TN	FP	FN
DT	76	7	18	8
NB	69	9	16	15
KNN	81	12	13	3
NN	84	0	0	25
RUSBoost	76	17	8	8
Bag	82	8	17	2
AdaBoost	82	13	12	2

As demonstrated in Table IV, all the models such as DT, NB, KNN, NN, RUSBoost, Bagging and AdaBoost were assessed for their performance measures based on the test dataset and as a result their accuracy, total cost, error rate, precision, recall and F1-score values are compared with each other in order to select the optimal model for identifying the PD based on keystroke dynamics. From Table IV, it is found that AdaBoost produced the maximum value of accuracy i.e. 87.16% and minimum value of error rate i.e. 12.84% and thus may be referred to as the superior classifier for the detection of PD based on keystroke dynamics. Moreover, the AdaBoost model has achieved almost balanced and highest values of precision, recall and F1-score i.e. 86.95%, 74.81% and 78.57% respectively. The value of accuracy for the KNN and RUSBoost is found to be the maximum 85.32%. The values of accuracy for the DT and NB models are less compared to other models. As far as the NN model is concerned, it has not produced the precision values and minimum F1-score value as a result the output produced by the model is imbalanced for the classes.

Table IV. Comparison of model performance on the test data

Model	Accuracy	Total Cost	Error Rate	Precision	Recall	F1 score
DT	76.15	26	23.85	63.76	59.24	60.20
NB	71.56	31	28.44	59.34	59.07	59.20
KNN	85.32	16	14.68	83.01	72.19	75.45
NN	77.06	25	22.94	-	50.00	43.46
RUSBoost	85.32	16	14.68	79.24	**79.24**	**79.24**
Bag	82.57	19	17.43	81.41	64.81	67.67
AdaBoost	**87.16**	**14**	**12.84**	**86.95**	74.81	78.57

Fig. 1 demonstrates the test accuracy of the different models namely DT, NB, KNN, NN, RUSBoost, Bagging and AdaBoost developed for the Parkinson's disease classification using keystroke dynamics. Here we find that the model AdaBoost achieved the highest accuracy that is 87.16%, where KNN and RUSBoost followed with 85.32%, the accuracy of Bagging is 82.57% and the lowest accuracy is observed for the rest models with DT having 76.15% of accuracy, followed by the NN having 77.06% and the NB having 71.56%. Thus the classification performance of the individual models can be surpassed by the ensemble-based methods while performing the Parkinson's disease classification using keystroke dynamics.

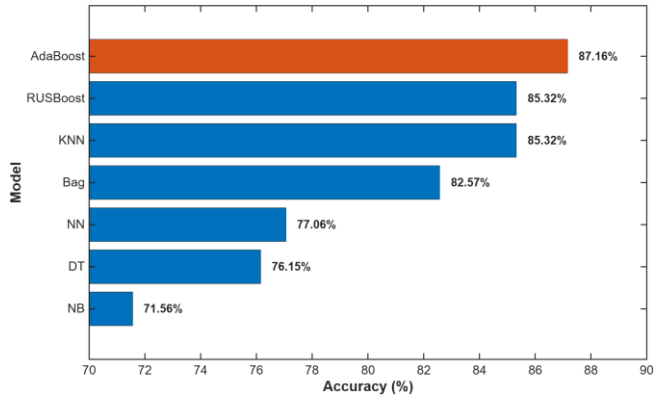

Fig.1. Comparison of the accuracy rates of the models on the test data

Fig. 2, ROC curves of the AdaBoost model based on the test dataset. For each class (Control and Parkinson), the AUC is 0.9129. All the ROC curves are strictly above the diagonal reference line, which implies that the performance of the proposed model is significantly better than random guess. The AUC value greater than 0.90 indicates that the proposed model is a highly discriminative model. As shown in Table IV, AdaBoost has the highest accuracy in all three scenarios, which also confirms the effectiveness of the AdaBoost model for the prediction of Parkinson's disease based on keystroke dynamics.

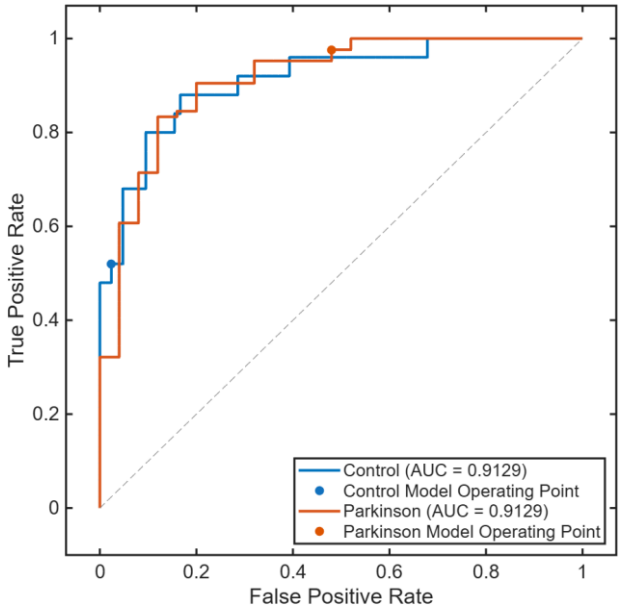

Fig.2. ROC curves and AUC values of the AdaBoost model

Fig. 3 and Fig. 4 present the SHAP summary plots of the AdaBoost model, which achieved the highest classification performance on the test dataset. Fig. 4 illustrates the Shapley values calculated for the Parkinson class, while Fig. 5 presents the corresponding values for the Control class. In these plots, each point represents an observation; the horizontal axis indicates the Shapley value (the contribution of a feature to the model output), while the color scale represents the magnitude of the corresponding feature value, ranging from low (blue) to high (red). Positive Shapley values increase the likelihood of predicting the corresponding class, whereas negative values contribute in the opposite direction. For both classes, the top 10 most influential features are visualized. The results indicate that time-based keystroke features, particularly HoldTime_CV, Hold_L, HoldTime_Min, FlightTime_IQR, and FlightTime_CV, play a decisive role in the model's decision-making process. It is observed that higher values of certain features strengthen the prediction toward the Parkinson class, whereas similar features contribute in the opposite direction for the Control class. These findings demonstrate that the AdaBoost model not only achieves high classification accuracy but also provides a transparent and interpretable decision-making structure.

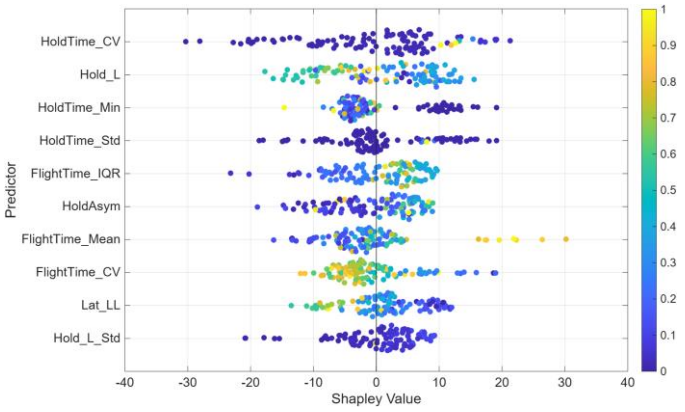

Fig.3. SHAP feature importance summary plot for the Parkinson class of the AdaBoost model

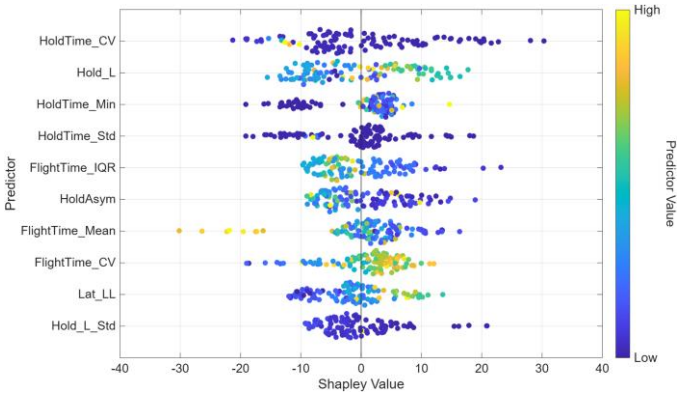

Fig.4. SHAP feature importance summary plot for the Control class of the AdaBoost model

The existing literature on Parkinson's disease detection based on keystroke dynamics demonstrates a wide range of performance outcomes. Arroyo-Gallego et al. [24], evaluated the neuroQWERTY approach on 52 participants in a home environment and reported an AUC of 0.76. Similarly, Tripathi et al. [13] applied CNN-based models to a larger and uncontrolled cohort of 281 participants and achieved AUC values in the range of 0.80–0.83. Bernardo et al. [25], analyzed the same Tappy dataset using a deep learning framework (modified SqueezeNet combined with SMOTE) and reported 90% accuracy; however, this approach requires extensive preprocessing steps and substantial computational resources. Adams [17] reported 97–100% accuracy, although these results were obtained from a relatively small and homogeneous cohort of 103 participants. Roy et al. [26] proposed a novel ensemble classification approach analyzing keystroke patterns on modern active devices to identify PD and demonstrated that ensemble-based methods consistently outperform single classifiers on keystroke data. Chauhan et al. [27] developed NeuroKey, a lightweight AI tool using passive keystroke dynamics with a Random Forest ensemble classifier, achieving an AUC of 0.96 on keystroke-based PD detection. Tat et al. [28] developed a soft intelligent keyboard combined with a Random Forest ensemble and achieved a cross-validation accuracy of 98.38% on

keystroke dynamics data from PD patients, though the approach depends on specialized custom sensor hardware.

Compared with the results of AdaBoost model from the literature, our AdaBoost model has achieved the accuracy of 87.16%, precision of 86.95%, recall of 74.81%, F1-score of 78.57%, and AUC of 0.9129 when evaluated on the dataset from Tappy. The proposed method not only presents a high-level of accurate classification capability, but it also integrates SHAP analysis for highly interpretable result within a single analysis tool.

IV. Conclusion

PD can be modeled as a classification problem using machine learning techniques by modeling the variations and patterns in the keystroke dynamics measurements that are most relevant to the onset of the condition. Here, we investigate whether Parkinson's disease can be predicted based on keystroke dynamics measurements given that 33 feature variables were used. From our empirical investigation, it was found that time-based micromotor characteristics held some significance. A comparative analysis shows that AdaBoost achieves superior performance. Accuracy is 87.16% with an AUC of 0.9129, thereby confirming the discriminative nature of the proposed ensemble models. In addition to overall accuracy, the proposed model achieved balanced precision, recall, and F1-score values, indicating its robustness in handling class imbalance. Another important conclusion drawn here is that ensemble-based models such as AdaBoost, RUSBoost tend to outperform the basic models or single classifiers. Also SHAP analysis revealed that the Variation-based features (features that model the variations of each keystroke features) related to HoldTime and FlightTime are highly significant for the classification of Parkinson's disease. This finding highlights the importance of temporal variability in keystroke dynamics as a potential digital biomarker for Parkinson's disease.

From the above analysis, it can be concluded that the machine learning models based on keystroke dynamics can serve as an early and non-invasive tool for the diagnosis of PD. Furthermore, the integration of explainable artificial intelligence techniques enhances the transparency and interpretability of the proposed system, which is critical for clinical decision support applications. Future work may expand the current dataset, achieve a more balanced sample, and compare the results with the proposed deep learning-based hybrid models in order to confirm the applicability of keystroke analysis in the clinical decision support systems.

References

[1] Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, *79*(4), 368-376.

[2] Poewe, W., et al. (2017). Parkinson disease. *Nature reviews Disease primers*, *3*(1), 17013.

[3] Su, D., et al. (2025). Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021. *bmj*, *388*.

[4] Kaya, D., & Soyukibar, T. E. (2022). Parkinson hastalığı ve parkinsonizm. *The Journal of Turkish Family Physician*, *13*(4), 182-192.

[5] Postuma, R. B., et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement disorders*, *30*(12), 1591-1601.

[6] Hess, C. W., & Hallett, M. (2017). The phenomenology of Parkinson's disease. *Seminars in neurology*, *37*(02), 109-117.

[7] Meral, M., Ozbilgin, F., & Durmus, F. (2025). Fine-Tuned Machine Learning Classifiers for Diagnosing Parkinson's Disease Using Vocal Characteristics: A Comparative Analysis. *Diagnostics*, *15*(5), 645.

[8] Meral, M., & Ozbilgin, F. (2025). Parkinson's disease diagnosis and severity assessment from gait signals via bayesian-optimized deep learning. *Diagnostics*, *15*(16), 2046.

[9] Özbilgin, F., Kurnaz, C., & Aydın, E. (2024). Non-invasive coronary artery disease identification through the iris and bio-demographic health profile features using stacking learning. *Image and Vision Computing*, *146*, 105046.

[10] Senkaya, Y., Kurnaz, C., & Ozbilgin, F. (2025). Enhancing Alzheimer's Diagnosis with Machine Learning on EEG: A Spectral Feature-Based Comparative Analysis. *Diagnostics*, *15*(17), 2190.

[11] Alfalahi, H., et al. (2022). Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Scientific reports*, *12*(1), 7690.

[12] Demir, B., Ulukaya, S., & Erdem, O. (2023). Detection of Parkinson's disease with keystroke data. *Computer Methods in Biomechanics and Biomedical Engineering*, *26*(13), 1653-1667.

[13] Tripathi, S., Arroyo-Gallego, T., & Giancardo, L. (2022). Keystroke-dynamics for Parkinson's disease signs detection in an at-home uncontrolled population: a new benchmark and method. *IEEE Transactions on Biomedical Engineering*, *70*(1), 182-192.

[14] Roy, S., Roy, U., Sinha, D., & Pal, R. K. (2023). Imbalanced ensemble learning in determining Parkinson's disease using Keystroke dynamics. *Expert Systems with Applications*, *217*, 119522.

[15] Tripathi, S., Acien, A., Holmes, A. A., Arroyo-Gallego, T., & Giancardo, L. (2024). Generalizing Parkinson's disease detection using keystroke dynamics: a self-supervised approach. *Journal of the American Medical Informatics Association*, *31*(6), 1239-1246.

[16] Yang, Y., Guo, B., Zhao, K., Liang, Y., Li, N., & Yu, Z. (2024). MELPD-Detector: Multi-level ensemble learning method based on adaptive data augmentation for Parkinson

disease detection via free-KD: Y. Yang et al. *CCF Transactions on Pervasive Computing and Interaction*, *6*(2), 182-198.

[17] Adams, W. R. (2017). High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. *PloS one*, *12*(11), e0188226.

[18] Goldberger, A. L., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, *101*(23), e215-e220.

[19] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

[20] Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *Ieee Access*, *10*, 99129-99149..

[21] Qawqzeh, Y. K., Alourani, A., & Ghwanmeh, S. (2023). An improved breast cancer classification method using an enhanced AdaBoost classifier. *International Journal of Advanced Computer Science and Applications*, *14*(1).

[22] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

[23] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

[24] Arroyo-Gallego, T., et al. (2018). Detecting motor impairment in early Parkinson's disease via natural typing interaction with keyboards: validation of the neuroQWERTY approach in an uncontrolled at-home setting. *Journal of medical Internet research*, *20*(3), e89.

[25] Bernardo, L. S., Damaševičius, R., Ling, S. H., de Albuquerque, V. H. C., & Tavares, J. M. R. (2022). Modified SqueezeNet architecture for Parkinson's disease detection based on keypress data. *biomedicines*, *10*(11), 2746.

[26] Roy, S., Roy, U., Sinha, D., & Pal, R. K. (2025). A novel approach to identify parkinson's disease and other similar neural stress by analysing keystrokes on modern active devices with ensemble classification. *Multimedia Tools and Applications*, *84*(29), 36163-36202..

[27] Chauhan, R., Jena, M., & Singh, D. (2025). NeuroKey: A Lightweight AI Tool Using Passive Keystroke Dynamics for Parkinson's Disease Detection and Monitoring. In *International Summit on the Global Internet of Things and Edge Computing* (pp. 39-53). Springer.

[28] Tat, T., Chen, G., Xu, J., Zhao, X., Fang, Y., & Chen, J. (2025). Diagnosing Parkinson's disease via behavioral biometrics of keystroke dynamics. *Science Advances*, *11*(14), eadt6631.

BIOGRAPHIES

Ferdi Ozbilgin was born in Tokat, Türkiye. He received the B.S. degree in electrical and electronics engineering from Sivas Cumhuriyet University, Sivas, Türkiye, in 2014, and the M.S. and Ph.D. degrees in electrical and electronics engineering from Ondokuz Mayıs University, Samsun, Türkiye in 2019 and 2023. He has been an Assistant Professor at Giresun University, Giresun, Türkiye. His research interests include biomedical image processing, signal processing and machine learning.