Classification of Subjects with Parkinson's Disease using Finger Tapping Dataset ⋆

Víctor Asanza * **Nadia N. Sánchez-Pozo** ***
Leandro L. Lorente-Leyva ***
Diego Hernan Peluffo-Ordóñez ***,****,†
Fancis R. Loayza ** **Enrique Peláez** *

* *Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador (e-mail: {vasanza,epelaez}@espol.edu.ec)*
** *Facultad de Ingeniería en Mecánica y Ciencias de la Producción, Escuela Superior Politécnica del Litoral, ESPOL, Neuroimaging and Bioengineering Laboratory, LNB, Campus Gustavo Galindo km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador (e-mail: floayza@espol.edu.ec)*
*** *Smart Data Analysis Systems Group (SDAS Research Group - www.sdas-group.com), Ben Guerir 47963, Morocco (e-mail: {nadia.sanchez, leandro.lorente, diego.peluffo}@sdas-group.com)*
**** *Modeling, Simulation and Data Analysis (MSDA) Research Program, Mohammed VI Polytechnic University, Ben Guerir 47963, Morocco (e-mail: peluffo.diego@um6p.ma)*
† *Faculty of Engineering, Corporación Universitaria Autónoma de Nariño, Carrera 28 No. 19-24, Pasto, Colombia, 520001, (email: diego.peluffo@aunar.edu.co)*

Abstract: Parkinson's disease is the second most common neurodegenerative disorder and affects more than 7 million people globally. In this work, we classify subjects with Parkinson's disease using data from finger-tapping on a keyboard. We use a free database by Physionet with more than 9 million records, preprocessed to delete atypical data. In the feature extraction stage, we obtained 48 features. We use Google Colaboratory to train, validate, and test nine supervised learning algorithms that detect the disease. As a result, we achieve a degree of accuracy higher than 98%.

Keywords: Parkinson's disease, Classification, Machine Learning, Finger Tapping

1. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder (Váradi, 2020), and it is characterized by many clinical symptoms such as tremor, bradykinesia, muscle stiffness, balance disorders, oral problems (when swallowing, speaking). Non-motor symptoms include depression, anxiety, anhedonia, loss of taste and smell, and sleep disorders (insomnia) (Maccarrone, 2020, Sveinbjornsdottir, 2016).

In 2018, the World Health Organization (WHO) estimated that around 7 million people suffer from PD, and by 2030 the number will grow to 12 million. According to the WHO, this disease has a significant impact on the elderly community; one in every 100 people over 60 years old has PD (Brown and Goldman, 2020). According to these statistics, the estimated growth in PD diagnoses is higher than 70%. Patients take between one and three years to be diagnosed. For this reason, it is essential to detect this disease at an early stage to improve the quality of life, as this is a neurodegenerative disease, i.e., its symptoms become more noticeable over time (Maccarrone, 2020, Sveinbjornsdottir, 2016).

In the last decades, technology has revolutionized medical treatments. For example, wearable technology is portable and allows rapid monitoring, which provides a better follow-up to the patient (Farahani et al., 2018). Meanwhile, Artificial Intelligence (AI) allows for interpreting radiographic images, MRI (Magnetic Resonance Imaging) images, and CT (Computed Tomography) scans, thus helping doctors to diagnose and treat patients (Farahani et al., 2018, Lau and Staccini, 2019).

Machine learning (ML) is a subset of AI. Currently, ML offers several new opportunities due to the access to large amounts of data. Personal, demographic, geographic,

⋆ Escuela Superior Politécnica del Litoral, ESPOL.

psychographic, or behavioral data that were previously unavailable or limited. ML algorithms can generate and activate this data to convert it into information and learning (Lau and Staccini, 2019). A diagnosis can be delivered, monitored, and categorized using classifiers. Hence, we can detect disorders such as PD at an early stage (Lau and Staccini, 2019).

Motor assessment is the most common evaluation in PD using some form of repetitive task. Finger tapping is one of the most reliable and straightforward tests for motor performance in the Unified Parkinson's Disease Rating Scale (UPDRS). In this assessment, a clinician looks for any decreasing amplitude, slowing, irregularity, or freezing of the movement. Some efforts have been made using digital tools to measure finger tapping and implementing AI techniques to classify, identify, and follow PD patients.

Taking advantage of the sensitive touch technology in smartphones, some researchers measure several parameters and features of the finger tapping of patients with PD (Bhatti et al., 2017). However, the applications were based merely on the measurements and the thresholding classification method. Other approaches also use finger tapping on smartphones as sensors for monitoring symptoms (Goldberger et al., 2000) with a Random forest algorithm that classifies the data from a small number of patients. Rodriguez-Cruz et al. (Rodríguez-Cruz et al., 2020) also used finger tapping with a combination of Machine and Deep learning techniques to perform remote detection of PD.

Therefore, this paper introduces a novel implementation of a ML classification technique that uses finger tapping data to classify patients with PD and follow their progress.

2. RELATED WORK

AI algorithms that aim to detect PD range from classic classifiers to Neural Networks (NN). There is a trend in methodologies from previous work when selecting stimuli, including those focusing on brain patterns.

Chen et al. use MRI and ML. They process MRI images in 2-dimensional pixels for feature and biomarker extraction, and, using a Support Vector Machine (SVM) algorithm, they obtain an accuracy of 93%. This experiment provides a better understanding at a clinical level about the disease (Chen et al., 2020).

Shah et al. use Electroencephalograms (EEG) and five different methodologies to train NNs. Nonetheless, they argue that a higher amount of data and better-quality EEGs are needed. Therefore, none of their methodologies are able to train the NN (Shah et al., 2020).

Analyzing brain patterns requires more extensive data preprocessing to train an AI. Working with this type of pattern is more difficult due to the complexity of the brain and the medical knowledge necessary to handle them (Bastidas et al., 2020, Germine et al., 2021).

Other research projects utilize motor patterns such as talking (voice) and walking. Gharehchopogh et al. use NNs and a voice dataset. They are able to detect PD with an accuracy of 93.22% employing a Multilayer Perceptron (MLP) with three layers and 28 nodes (22 input nodes, five hidden nodes, and one output node) (Pereira et al., 2018).

Wroge et al. use six algorithms, decision tree (DT), extra tree (ET), SVM, random forest (RF), artificial neural network (ANN), and gradient-powered classifier (GBC). They also use two data preprocessing models, Audio-Visual Emotion Recognition Challenge (AVEC) and Geneva Minimalistic Acoustic Parameter Set (GeMaps). In their work, they highlight the 86% accuracy of the GBC algorithm with AVEC preprocessing on a voice and speech dataset collected by them through a telephone voice-recorder application (Wroge et al., 2018, December).

Hariharan et al. use a Least Square Support Vector Machine (LS-SVM) algorithm, a Probabilistic Neural Network (PNN), and a General Regression Neural Network (GRNN). Through feature preprocessing on a voice dataset, they achieve an accuracy of 100% (Hariharan et al., 2014). Sadek et al. also achieve a 100% accuracy using the same dataset but with an ANN. (Sadek et al., 2019).

Urcuqui et al. conducted an experiment in which they asked volunteers to walk a four-meter route three times. They video-recorded them and analyzed the data from the movements of their legs. They processed the data using four ML algorithms, being RF the most precise algorithm with an 82% accuracy (Urcuqui et al., 2018,September).

Wang et al. use a backpropagation neural network algorithm (BPNN) with a walking dataset. They obtain an accuracy of 42% (Wang et al., 2015,August).

Most of the studies that involve analysis of motor activities show results with high accuracy. There is a high amount of accessible data and diversity of patterns (Germine et al., 2021). Finally, we work with the previously mentioned algorithms and contrast them with derived ones or in the same category to improve and obtain new results.

3. DATASET

We use a dataset provided by Physionet, a platform containing a wide variety of data for research purposes (Goldberger et al., 2000). All subjects performed the keystroke test voluntarily after signing informed consent. The volunteers installed the TAPPY app on their computers. This app records how many times and which keys were pressed or released. In addition, TAPPY creates a record for each key pressed on one of the three possible keys (R, L, and space bar).

The folder `Archived_Users` contains one *.txt* file for each one of the 227 volunteers. The name of each file contains the UserKey (e.g., *User_0EA27ICBLF.txt*), and each file contains the following demographic information:

- BirthYear: Year of birth (e.g., 1952)
- Gender: male or female (e.g., Female)
- Parkinsons: patient or control (e.g., True)
- Tremors: yes or not (e.g., True)
- DiagnosisYear: the evolution of the disease in years (e.g., 2000)
- Sided: hemibody affected, right, left, or none, performed by self-evaluation (e.g., Left)

- UPDRS: motor assessment evaluated by a specialist (e.g., Don't know)
- Impact: how much the disease impacts your daily life, little, moderately, or severely (e.g., Severe)
- Levodopa: the binary response in the use of levodopa drug, yes or not (e.g., True)
- DA: the binary response in the use of dopamine agonist drug, yes or not (e.g., True)
- MAOB: the binary response in the use of MAO-B drug, yes or not (e.g., False)
- Other: the binary response in the use of other drugs, yes or not (e.g., False)

The folder `Tappy_Data` contains between two and eight records stored in *.txt* files for each one of the 227 test participants. Each file contains the following data from the experiment:

- UserKey: 10 character code for a user (e.g., 0EA27ICBLF)
- Date: With the last two digits of the current year, the month and day YYMMDD (e.g., 160722)
- Timestamp: HH:MM:SS.SSS (e.g., 18:41:04.336)
- Hand: "L" or "R" key pressed; where "L" denotes left hand, and "R" denotes right hand (e.g., L)
- Hold time: Time between pressing and releasing the current key mmmm.m milliseconds (e.g., 0101.6)
- Direction: Previous to current "LL", "LR", "RL" y "RR"; "S" is the space key (e.g., LL)
- Latency time: Time between pressing the previous key and pressing the current key. Milliseconds (e.g., 0234.4)
- Flight time: Time between the release of the previous key and press of the current key. Milliseconds (e.g., 0156.3)

4. METHODOLOGY

The development of this study follows the three steps that constitute the applied methodology, specified in Fig. 1.

Fig. 1. Methodology Steps

4.1 Data Preprocessing

As shown in Fig. 2 (a), we first listed the *UserKeys* of the 227 test participants in the `Archived_Users` folder. Using the *UserKey* as a reference, we searched for the records in the `Tappy_Data` folder belonging to each participant. These records summarize the keystrokes per specific user in a given time. The search resulted in 217 subjects with readable data, of which 162 were patients with PD and 55 were healthy controls. We normalized these records with the Min-Max method, as shown in Fig. 2 (b).

4.2 Feature Extraction

Fig. 2 (c) shows how feature extraction condenses the information from the records of each one of the 217 participants into one single *.CSV* file with 217 rows x 48 columns. We processed the 48 columns representing 48 features as follows:

- BirthYear and DiagnosisYear columns were changed to numeric type
- Columns with True or False values were coded as binary data (0, 1): Female, Tremors, Levadopa, DA, MAOB and Other
- The Side variable was coded (one-hot encoded): Side_Left, Side_None, Side_Right
- The UPDRS variable was coded (one-hot encoded): UPDRS_1, UPDRS_2, UPDRS_3, UPDRS_4 and UPDRS_Don'tKnow
- The categorical Impact variable was coded (one-hot encoded): Impact_Medium, Impact_Mild, Impact_None, Impact_Severe
- The Hold_time, Latency_time and Flight_time; were coded (one-hot encoded) and divided into address groups: "LL", address "LR", address "LS", address "RL", address "RR", address "RS", address "SL", address "SR" and "SS".
- Finally, the labels are the diagnosis of PD.

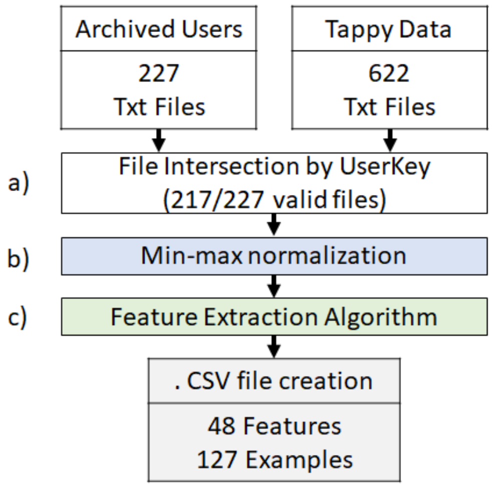

Fig. 2. a) Search for records based on the UserKey list from the Archived_Users folder. b) Min-Max normalization of records from the Tappy_Data folder. c) Feature Extraction for each one of the 217 participants.

4.3 Classification Algorithm

The resulting *.CSV* file with 217 rows and 48 columns is our *Dataset*, which we used for the training and validation process of the classification algorithms. Fig. 3 shows the *Dataset* split in a 70:30 ratio, i.e., 70% of the data is the *Training_Set* for training the classification algorithms, while 30% is the *Testing_Set* for validating them.

Using *Google Colaboratory*, we run each one of the classification algorithms as shown in the Algorithm 1. Among the classification algorithms that we employed are:

- Naive Bayes (NB): Probabilistic classifier based on Bayes' theorem (Abedin et al., 2019). NB classifiers are a set of supervised learning algorithms, which can be faster compared to more sophisticated algorithms.
- Multilayer Perceptron (MLP): The MLP neural network needs to specify which ones are the input nodes

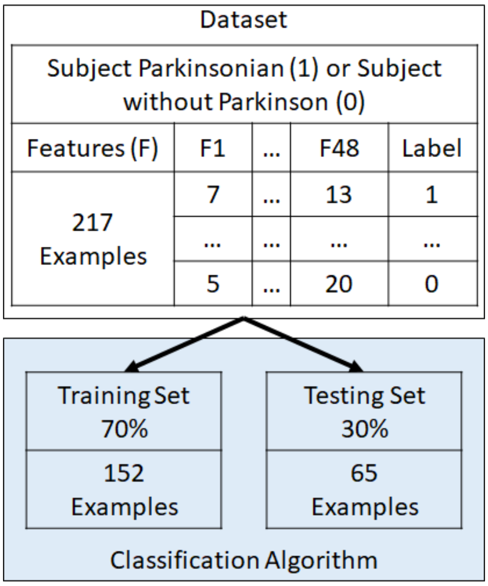

Fig. 3. Splitting the Dataset into Training and Testing Sets

and the output nodes of the algorithm. MLP can also modify the number of hidden layers. The network is trained iteratively since the parameters are updated in each iteration depending on the loss function. The accuracy of the algorithm improves depending on the weights of each hidden layer node, in exchange for an increased computational cost (Asanza et al., 2020, 2017).

- Random Forest (RF): The ensemble learning algorithm in which the number of classification groups is specified (Abedin et al., 2019), (Wroge et al., 2018, December). RF fits several decision tree classifiers on different subsamples of the dataset and uses averages to improve prediction accuracy and control overfitting.
- Extra Trees (ET): This algorithm works by creating a high number of decision trees from the training dataset. The predictions are made using majority voting in the classification case (Shtar et al., 2021).
- Logistic Regression (LR): Logistic regression is a ML technique that comes from the field of statistics (Abedin et al., 2019).
- Ridge Classifier: Ridge regression estimates the coefficients of multiple regression models in scenarios where the independent variables are highly correlated. It has applications in fields including econometrics, chemistry, and engineering (Singh et al., 2016).
- Support Vector Machine - Linear Kernel(SVM): Algorithm based on support vector learning (Chen et al., 2020) with the ability to perform binary and multi-class classification on a dataset. SVM - Linear Kernel allows dense and sparse inputs.
- Gradient Boosting Classifier (GBC): GGCs are a group of ML algorithms that combine many weak learning models to create a robust predictive model (Karabayir et al., 2020).
- Light Gradient Boosting Machine (LGBM): LGBM is a free and open-source distributed gradient boosting framework for ML originally developed by Microsoft (Chun et al., 2021).

Algorithm 1 Training and Testing Classification Algortihm

Result: Classification Algortihm
for *folder in DataFolders* **do**
  for *Traning_Set* **do**
    Apply `Classification_Algorithm` function to each row.
  end
  for *Testing_Set* **do**
    Apply `Classification_Algorithm` function to each row.
  end
  Export Accuracy
end

5. RESULTS AND DISCUSSION

In recent years, the development of methods and tools to support the recognition of PD has increased. Several techniques have been used, such as the variation of finger tapping, artificial intelligence, ML and DL techniques, voice recordings, among others. All this to assist and make the diagnosis and monitoring process for PD much more efficient. Since there is no specific test for detecting PD, it is necessary to constantly evaluate the history and symptoms of the patients to achieve an accurate diagnosis. For this reason, large datasets are generated in which classification algorithms are crucial for determining and discriminating between healthy and sick patients.

This section presents the results we obtained with the classification algorithms, evaluating them regarding the following parameters:

- Recall or Probability of classifying true positives, calculated using the parameters True Positive `(TP)`, False Negative `(FN)` as shown in equation 1:
$$Recall(\%) = \frac{TP}{TP + FN}. \tag{1}$$
- Accuracy or Proximity of the results, calculated using the parameters True Negative `(TN)`, False Positive `(FP)`, False Negative `(FN)`, True Positive `(TP)` as shown in the equation 2:
$$Accuracy(\%) = \frac{TN + TP}{TN + FP + FN + TP}. \tag{2}$$
- Precision or Dispersion of the set of values obtained, calculated using the parameters True Positive `(TP)`, False Positive `(FP)` as shown in the equation 3:
$$Precision(\%) = \frac{TP}{FP + TP}. \tag{3}$$
- F1(F-Score), used to combine the measures of Precision and Recall into a single value, as shown in the equation 4:
$$F1(F - Score)(\%) = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{4}$$

Table 1 outlines the evaluation results of the applied classification algorithms regarding the parameters described above. We observe that NB reached the highest accuracy with 98.04%; MLP had the highest Recall with 97.35%; NB had the best Precision and F1, with 100% and 98.61%, respectively. Finally, we can argue that the classification results are generally good, with the NB classifier showing the best performance and quality over the rest of the classification algorithms.

Table 1. Classification Algorithms Results

Algorithm	Accuracy(%)	Recall(%)	Prec.(%)	F1(%)
NB	98.04	97.35	100	98.61
MLP	97.38	98.26	98.26	98.22
RF	96.67	98.18	97.42	97.75
ET	96.67	98.18	97.42	97.75
LR	96.00	97.27	97.42	97.22
RIDGE	96.00	97.27	97.42	97.22
SVM	94.04	93.71	98.26	95.79
GBC	94.04	95.53	96.33	95.84
LGBM	94.04	95.53	96.59	95.88

Fig. 4 shows the accuracy and execution time of the ML algorithms used in the study. We evaluated the results using cross-validation with ten iterations of each algorithm. The algorithms that achieved the highest accuracy were NB and MLP with short execution times. Similarly, the algorithms with the lowest accuracy were SVM, GBC, and LGBM. In general, the classification results were good. The algorithms were highly accurate with short execution times.

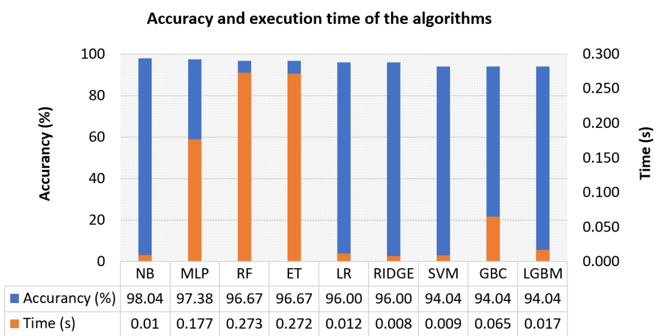

Fig. 4. Accuracy and execution time of the best ML Algorithms.

The Confusion Matrix in Fig. 5 shows the false positives that the NB algorithm classified, with the upper left and lower right boxes being the correct predictions and the upper right and lower left, the errors or false positives.

The receiver operating characteristic curve (ROC) in Fig. 6 represents the errors of the NB algorithms based on their false positives and negatives. ROC is a metric that assesses AI algorithms.

6. CONCLUSIONS

PD is one of the most common neurodegenerative disorders, affecting millions of people worldwide. Using ML techniques, we are able to classify patients with PD with an accuracy of 98.04%.

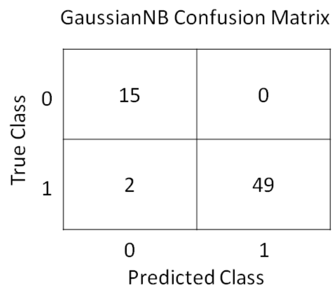

Fig. 5. Confusion matrix of NB.

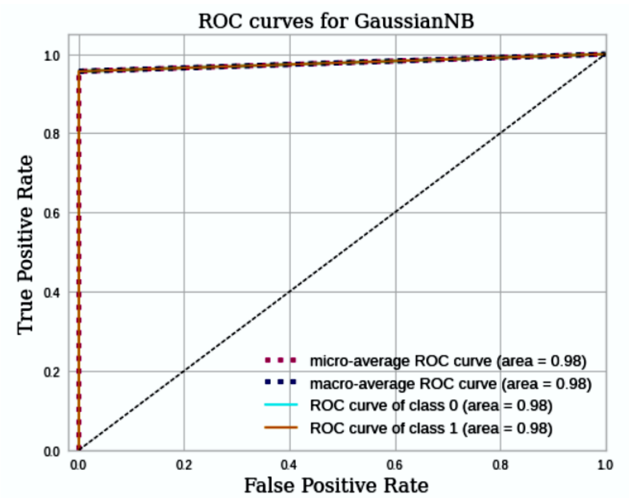

Fig. 6. Receiver Operating Characteristic curve of NB.

The results show that using 48 features for each participant was adequate, as we achieve accuracy values greater than 98%. As shown in Fig. 4, the nine algorithms NB, MLP, RF, ET, LR, RIDGE, SVM, GBC and LGBM showed accuracy values of 98.04%, 97.38%, 96.67%, 96%, 94.04%, 94.04%, and 94.04%, respectively. Applications that need high accuracy could use the NB or MLP algorithms. On the other hand, for applications that need real-time responses the algorithms that reported shorter execution times were SVM and NB (values of about 10 ns).

Implementation of machine learning models in the classification of patients with PD showed positive results, especially the NB model. According to our results, NB algorithms are better than neural networks when classifying control subjects.

We recommend preprocessing the data extensively because a cleaner dataset improves prediction accuracy when working with classification algorithms. In terms of future work, we propose to add an automatic feature selection stage to identify the features that contribute to the classifiers. In addition, we suggest using DeepLearning-based methods as classifiers. Finally, we propose employing the Hampel Filter for preprocessing the EEG signals.

ACKNOWLEDGMENTS

The authors are greatly grateful by the support given by the SDAS Research Group (`www.sdas-group.com`) as well as the internal editor J. Mejía-Ordóñez for the manuscript reviewing and editing.

REFERENCES

Abedin, M.M., Maniruzzaman, M., Ahmed, N.F., Ahammed, B., and Ali, M. (2019). Classification and prediction of parkinson disease: A machine learning approach. In *International Conference Data Science and SDGs: Challenges, Opportunities and Realities*.

Asanza, V., Sanchez, G., Cajo, R., and Peláez, E. (2020). Behavioral signal processing with machine learning based on fpga. In *International Conference on Systems and Information Sciences*, 196–207. Springer.

Asanza, V., Martin, C.A., Eslambolchilar, P., van Woerden, H., Cajo, R., and Salazar, C. (2017). Finding a dynamical model of a social norm physical activity intervention. In *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 1–6. doi:10.1109/ETCM.2017.8247450.

Bastidas, D., Piñeros, C., Peluffo-Ordóñez, D.H., Sierra, L.M., Becerra, M.A., and Umaquinga-Criollo, A.C. (2020). Analytic study on the performance of multi-classification approaches in case-based reasoning systems: Medical data exploration. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*.

Bhatti, D., Thompson, R., Hellman, A., Penke, C., Bertoni, J.M., and Torres-Russotto, D. (2017). Smartphone apps provide a simple, accurate bedside screening tool for orthostatic tremor. *Movement disorders clinical practice*, 4(6), 852–857.

Brown, E.G. and Goldman, S.M. (2020). Modulation of the microbiome in parkinson's disease: Diet, drug, stool transplant, and beyond. *Neurotherapeutics*, 1–12.

Chen, Y., Zhu, G., Liu, D., Liu, Y., Yuan, T., Zhang, X., Jiang, Y., Du, T., and Zhang, J. (2020). The morphology of thalamic subnuclei in parkinson's disease and the effects of machine learning on disease diagnosis and clinical evaluation. *Journal of the neurological sciences*, 411, 116721.

Chun, P.j., Izumi, S., and Yamane, T. (2021). Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine. *Computer-Aided Civil and Infrastructure Engineering*, 36(1), 61–72.

Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., and Mankodiya, K. (2018). Towards fog-driven iot ehealth: Promises and challenges of iot in medicine and healthcare. *Future Generation Computer Systems*, 78, 659–676.

Germine, L., Strong, R.W., Singh, S., and Sliwinski, M.J. (2021). Toward dynamic phenotypes and the scalable measurement of human behavior. *Neuropsychopharmacology*, 46(1), 209–216.

Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., and Stanley, H.E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23), e215–e220.

Hariharan, M., Polat, K., and Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of parkinson's disease. *Computer methods and programs in biomedicine*, 113(3), 904–913.

Karabayir, I., Goldman, S.M., Pappu, S., and Akbilgic, O. (2020). Gradient boosting for parkinson's disease diagnosis from voice recordings. *BMC Medical Informatics and Decision Making*, 20(1), 1–7.

Lau, A.Y. and Staccini, P. (2019). Artificial intelligence in health: New opportunities, challenges, and practical implications. *Yearbook of medical informatics*, 28(01), 174–178.

Maccarrone, M. (2020). Missing pieces to the endocannabinoid puzzle. *Trends in molecular medicine*, 26(3), 263–272.

Pereira, C.R., Pereira, D.R., Rosa, G.H., Albuquerque, V.H., Weber, S.A., Hook, C., and Papa, J.P. (2018). Handwritten dynamics assessment through convolutional neural networks: An application to parkinson's disease identification. *Artificial intelligence in medicine*, 87, 67–77.

Rodríguez-Cruz, A., Romo-Mancillas, A., Mendiola-Precoma, J., Escobar-Cabrera, J.E., García-Alcocer, G., and Berumen, L.C. (2020). Effect of valerenic acid on neuroinflammation in a mptp-induced mouse model of parkinson's disease. *IBRO reports*, 8, 28–35.

Sadek, R.M., Mohammed, S.A., Abunbehan, A.R.K., Ghattas, A.K.H.A., Badawi, M.R., Mortaja, M.N., Abu-Nasser, B.S., and Abu-Naser, S.S. (2019). Parkinson's disease prediction using artificial neural network.

Shah, S.A.A., Zhang, L., and Bais, A. (2020). Dynamical system based compact deep hybrid network for classification of parkinson disease related eeg signals. *Neural Networks*, 130, 75–84.

Shtar, G., Rokach, L., Shapira, B., Nissan, R., and Hershkovitz, A. (2021). Using machine learning to predict rehabilitation outcomes in postacute hip fracture patients. *Archives of physical medicine and rehabilitation*, 102(3), 386–394.

Singh, A., Prakash, B.S., and Chandrasekaran, K. (2016). A comparison of linear discriminant analysis and ridge classifier on twitter data. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 133–138. IEEE.

Sveinbjornsdottir, S. (2016). The clinical symptoms of parkinson's disease. *Journal of neurochemistry*, 130, 318–324.

Urcuqui, C., Castaño, Y., Delgado, J., Navarro, A., Diaz, J., Muñoz, B., and Orozco, J. (2018,September). Exploring machine learning to analyze parkinson's disease patients. *In 2018 14th International Conference on Semantics, Knowledge and Grids (SKG)*, 160–166.

Váradi, C. (2020). Clinical features of parkinson's disease: The evolution of critical symptoms. *Biology*, 9(5), 103.

Wang, T., Zhang, D., Wang, Z., Jia, J., Ni, H., and Zhou, X. (2015,August). Recognizing gait pattern of parkinson's disease patients based on fine-grained movement function features. *In 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, 1–10.

Wroge, T.J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D.C., and Ghomi, R.H. (2018, December). Parkinson's disease diagnosis using machine learning and voice. in 2018 ieee signal processing in medicine and biology symposium. *(SPMB)*, 1–7.