Exploring a Novel Machine Learning Approach for Evaluating Parkinson's Disease, Duration, and Vitamin D Level

Md. Asraf Ali[1], Md. Kishor Morol[2], Muhammad F Mridha[3], Nafiz Fahad[4], Md Sadi Al Huda[5], Nasim Ahmed[6]
Department of Computer Science, American International University- Bangladesh, Dhaka, Bangladesh[1, 2, 3]
Computer Science Miscellaneous and Applications,
Artificial Intelligence Research and Innovation Lab (AIRIL), Dhaka, Bangladesh[4, 5]
School of Computer Science, The University of Sydney, Sydney, Australia[6]

***Abstract*—Parkinson's disease is an increasingly prevalent, degenerative neurological condition predominantly afflicting individuals aged 50 and older. As global life expectancy continues to rise, the imperative for a deeper comprehension of factors influencing the course and intensity of PD becomes more pronounced. This investigation delves into these facets, scrutinizing various parameters including patient medical history, dietary practices, and vitamin D levels. A dataset comprising 50 PD patients and 50 healthy controls, sourced from Dhaka Medical Institute, serves as the foundation for this study. Machine learning techniques, notably the Modified Random Forest Classifier (MRFC), are harnessed to prognosticate both PD severity and duration. Strikingly, the MRFC-based prediction model for PD severity attains an impressive accuracy of 97.14%, while the predictive model for PD duration demonstrates an accuracy of 95.16%. Noteworthy is the observation that vitamin D levels are notably higher in the healthy cohort compared to PD-afflicted individuals, exerting a substantial positive influence on both the severity and duration predictions, surpassing the influence of other measured parameters. This inquiry underscores the practicality of machine learning in forecasting PD progression and duration and underscores the pivotal role of vitamin D levels as a predictive factor. These discoveries provide invaluable insights into advancing our comprehension and management of PD in an aging population.**

Keywords—Parkinson's disease; machine learning; vitamin D; severity; disease duration

I. Introduction

Parkinson's disease is a progressive neurological condition characterized by mobility restriction, primarily affecting the central nervous system. This is a consequence of the depletion of substantia nigra cells, which are responsible for producing dopamine, a critical factor in regulating movement [1]. This ailment manifests through motor and non-motor symptoms, often beginning subtly with hand tremors and gradually eroding motor control, affecting over 10 million individuals worldwide, particularly those aged 60 and above [2]. Understanding the early signs of Parkinson's disease (PD) is vital for effective intervention, but misconceptions and inadequate awareness persist, hindering timely diagnosis [3]. PD's impact on neural networks governing bodily movements is profound, involving critical brain regions such as the basal ganglia and the substantia nigra [4]. Typical PD symptoms encompass tremors, stiffness, slow movement, balance issues, and gait problems, alongside non-motor symptoms like depression, constipation, and sleep disturbances [5]. Recognizing these symptoms can be challenging due to their variability, emphasizing the importance of consulting neurophysicians, especially for older individuals [6].

Moreover, Parkinson's disease is on the rise in Bangladesh, largely due to increased life expectancy and a growing population. As vitamin D deficiency emerges as a potential risk factor for the condition, researchers are delving into the interplay among patient lifestyles, vitamin D levels, and the application of machine learning techniques to detect Parkinson's disease. Notably, there exists an inverse relationship between the severity of Parkinson's disease, its symptoms, and cognitive function with serum 25(OH)D levels. Generally, individuals with Parkinson's disease tend to exhibit lower vitamin D levels compared to their healthy counterparts [7, 8]. Additionally, studies have pointed to a significant occurrence of vitamin D deficiency among middle-aged women in Bangladesh [9]. Diverse machine learning methods, including Artificial Neural Networks (ANN), Decision Trees, and Support Vector Machines (SVM), have been employed to predict Parkinson's disease based on a variety of features [10]. Nevertheless, there remains an ongoing debate surrounding the link between vitamin D and Parkinson's disease, primarily due to disparities in the studied populations and methodological constraints [11][12].

Therefore, utilizing a custom model, this current study aims to shed light on the connection between PD severity and duration with vitamin D levels. It builds upon existing research to explore the potential of machine learning in predicting PD, offering a promising avenue for timely intervention. Key contributions of this paper are mentioned below:

1) *Utilized* supervised machine learning approaches to anticipate Parkinson's disease (PD) using a dataset comprising 50 PD patients and 50 healthy individuals from Bangladesh.

2) *Collected* demographic data and clinical features of PD participants, including age, education, disease duration, cardinal features of the disease, and disease severity, to build a comprehensive dataset.

3) The vitamin D levels of participants were evaluated, and vitamin D deficiency was classified as values below 30ng/mL.

4) Investigated correlations among various PD features, including positive and negative correlations.

5) Employed several machine learning algorithms to predict PD based on the dataset, including Random Forest, Decision Tree, Naive Bayes, Logistic Regression, Nearest Neighbor (KNN), and a Modified Random Forest Classifier (MRFC).

6) Customize MRFC model with specific settings, including the number of trees, maximum depth, maximum number of features, verbosity, mode, warm start, and thread usage, to optimize its performance.

7) Introduced performance evaluation metrics such as sensitivity (Sn) and specificity (Sp) into the confusion matrix to assess the classification results.

Therefore, the rest of the paper is designed as follows: Section II is literature review, Section III is about materials and methods, Section IV delves into results and discussion, Section V is Discussion, and in Section VI conclusion is presented.

II. Literature Review

Various researchers do research in the field of Parkinson's disease detection. Quan et al. (2021) focus on exploiting dynamic speech features to identify voice alterations in patients with Parkinson's disease (PD). Using a Bidirectional LSTM model instead of the static features seen in conventional machine learning models, it greatly increases the accuracy of PD identification. The study makes use of a mixed-gender database from GYENNO SCIENCE Parkinson that has 45 patients (15 Healthy Controls, 30 PD cases). However, information regarding feature engineering and data preprocessing is missing. According to the experimental results, the suggested Bidirectional LSTM model has an accuracy of 75.56% [13]. Shinde et al. (2019) improved the diagnosis of Parkinson's disease (PD) by using convolutional neural networks (CNNs) with neuromelanin-sensitive magnetic resonance imaging (NMS-MRI), a computer-based method. This strategy achieves better results than current approaches: 80% of tests correctly identify PD from healthy controls, and 85.7% correctly identify PD from atypical Parkinsonian disorders. It detects minute alterations in the substantia nigra pars compacta (SNc) by using CNNs for feature extraction and data augmentation. Sensitivity, specificity, and ROC curves are some of the evaluation criteria that demonstrate its excellent performance in the diagnosis and categorization of Parkinson's disease [14]. Noor et al. (2020) investigated the use of deep learning to identify neurological conditions using several MRI modalities, with a focus on schizophrenia, Parkinson's disease, and Alzheimer's disease. Convolutional Neural Networks (CNNs) are emphasized as being better at detecting these illnesses in its overview of current deep learning techniques. Datasets including MIRIAD, Open fMRI, PPMI, ADNI, COBRE, fastMRI, and FBIRN are used in the work along with discussion of problems and recommendations for future research areas. There is no specific discussion of feature extraction techniques or data preprocessing. The emphasis is on deep learning techniques, with CNNs being identified as the most effective way [15].

III. Materials and Methods

Numerous researchers in various disciplines have recently adopted machine learning-based approaches to get better accuracy from the analysis of complicated data [13]. The main objective of this study was to anticipate PD through various characteristics data which is complicated. Thus, supervised machine learning approaches were used to achieve the objectives of this study.

A. Dataset Description

To conduct this research, we collected a dataset of 50 PD patients and 50 healthy people from Sir Salimullah Medical College, where all subjects were Bangladeshi. Demographic Data and Clinical Features of PD participants are presented in Table I. The clinical states of the recruited subjects were evaluated using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale [16], and the severity of Parkinson's disease was determined according to the study's criteria [17]. In the laboratory of Sir Salimullah Medical College, serum 25-hydroxyvitamin D levels were determined in both the PD and healthy groups using CLIA kits using the radioimmunoassay technique. In the present investigation, blood vitamin D levels of 30ng/mL were judged normal, whereas values below 30ng/mL were deemed insufficient. These include age, sex, education, occupation, socioeconomic status, nature of work, cardinal features, disease duration, disease severity, smoking history, caffeine history, and vitamin D level. Additionally, the age comparison between PD and healthy participants is presented in Fig. 1, where PD participants are younger than healthy participants.

TABLE I. The Demographic Data and Clinical Features of PD Participants

Variables		Case	
		N	Percentage
Age	<60	1	2
	≥ 60	49	98
Education	Diploma	31	62
	Illiterate	11	22
	Under Diploma	8	16
Disease Duration	less than 5 years	26	52
	5 to 10 years	17	34
	Greater than 10	7	14
Cardinal Features of Disease	tremor	30	60
	rigidity	5	10
	bradykinesia	12	24
	postural instability	3	6
Disease Severity	1-2.5	26	52
	2.5-3	18	36
	>3	6	12

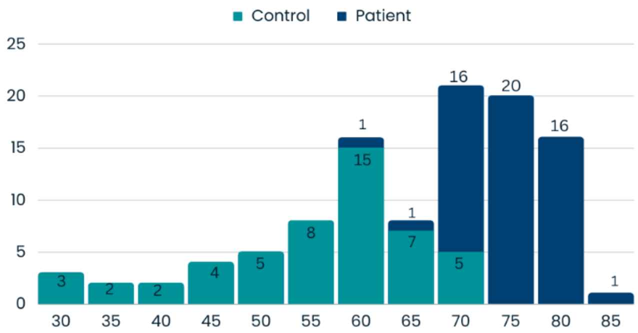

Fig. 1. Age comparison between PD and healthy participants.

Moreover, various features are causes of PD. The correlations among the features of PD are shown in Fig. 2. There are two types of correlations among the features - positive & negative. A positive correlation indicates that if the characteristic improves, the linked feature likewise increases, and if the feature falls, the connected feature reduces as well. Both aspects move in simultaneously, and their connection is linear. A negative correlation indicates that if one property's value rises, the linked trait's value falls, and vice versa.

Fig. 2. Correlations among the features.

B. Model Description

1) Random forest classifier: A well-known supervised machine learning algorithm is the Random Forest. It finds application in both classification and regression challenges within machine learning. The Random Forest serves as a classifier that enhances the predictive accuracy of a dataset by averaging the outcomes of multiple decision trees generated on different subsets of the dataset. When applied to address regression problems, the Random Forest Algorithm utilizes the mean squared error (MSE) from each node's data branch [17]. However, the formula is mentioned below in the equation number i.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (fi - yi)^2 \qquad (1)$$

In this context, N denotes the number of data points, fi signifies the model's output, and yi signifies the actual value for the specific data point i. The formula is employed to calculate the distance from the forecasted value to each node, and this outcome is pivotal in establishing the most suitable path through the forest.

2) Decision tree: The decision tree is a method for supervised learning that can effectively address classification and regression problems, primarily in solving classification tasks [18]. The terminology implies its utilization of a tree-like diagram to present predictions derived from a sequence of feature-driven divisions. It commences with a central node and culminates in a determination made at the terminal leaf. Therefore, in Fig. 3, a decision tree is mentioned [19].

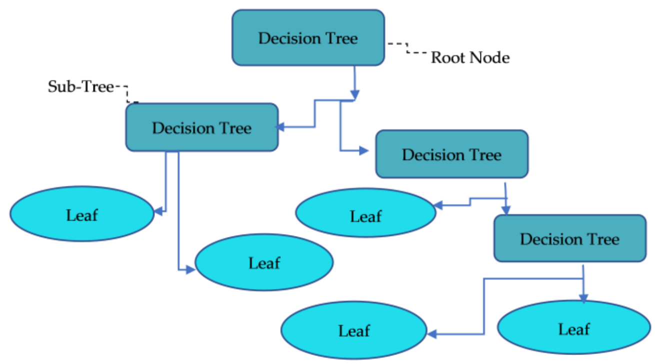

Fig. 3. Decision tree.

The decision trees utilize a representation known as the Sum of Products (SOP). The term Disjunctive Normal Form (DNF) is an alternative designation for the Sum of Products (SOP). In this context, each branch culminating in a node of the same category represents a conjunction (product) of values for that category. Conversely, separate branches ending in the same category constitute a disjunction (sum).

3) Naive bayes: The Naive Bayes algorithm is an algorithm for supervised learning. The Naive Bayes classifier is based on the conditional probability principle. It is straightforward and quick to predict the category of the test data set. It is also effective for multiclass prediction [17]. When the independence assumption holds true, the Naive Bayes classifier outperforms other models. The equation of Naïve Bayes is mentioned below in equation number ii.

$$P(A \mid B) = \frac{P(B \mid A).P(A)}{P(B)} \qquad (2)$$

The posterior probability P (c|x) can be computed using the prior probabilities P(c), P(x), and P(x|c), thanks to Bayes' theorem [19]. Naive Bayes classifiers assume that the impact of one predictor value (x) on a specific class (c) is independent of the impact of other predictor value combinations. Here, conditional independence on class labels is granted.

4) Logistic regression: Logistic regression is an example of supervised learning. It is used to solve classification problems, with the most common application being binary logistic regression with a binary outcome. Number iii equations underlying logistic regression:

$$y = \frac{e^{(b0+b1\times x)}}{1+e^{(b0+b1\times x)}} \tag{3}$$

where, y represents the predicted output, b0 represents the bias or intercept term, and b1 represents the single input value (x) coefficient. Each input data column contains a b coefficient (a constant real value) that must be learned from your training data [18].

5) K-Nearest Neighbor (KNN): K-Nearest Neighbors (KNN) is a relatively straightforward supervised machine learning technique. In this method, new data or instances are categorized based on their similarity to the existing ones. KNN follows a "lazy learning" approach, setting it apart from the previously discussed classifiers. It doesn't actively progress during training and primarily involves storing the training data. Its classification efforts only come into play when fresh, unlabeled data emerges [19].

KNN demonstrates the optimal performance when the nearest neighbors to a data point all belong to the same category. The underlying assumption is that if all nearby neighbors concur, a new data point will likely fall within the same group. Two compelling reasons for employing KNN are its simplicity in terms of comprehension and application. However, KNN's accuracy hinges on the chosen distance metric, and under certain distance metrics, it can achieve 100 percent accuracy [18].

Nevertheless, the computational cost of finding the closest neighbors of KNN on vast datasets can be quite significant. Furthermore, noisy data can potentially disrupt KNN classification. Features with a wider range of values may dominate the distance metric, necessitating the normalization or scaling of features. Notably, due to its "lazy learning" approach, KNN often demands more substantial storage resources than eager classifiers. Thus, the effectiveness of KNN is closely tied to selecting an appropriate distance metric [17] [20].

6) Modified Random Forest Classifier (MRFC): This research utilized a modified Random Forest classifier (MRFC) as the proposed model (see Fig. 4). We choose MRFC for these reasons-

- The MRFC model with specific settings, including the number of trees, maximum depth, maximum number of features, verbosity, mode, warm start, and thread usage, to optimize its performance.
- Customizations are designed to fine-tune the Random Forest algorithm for the specific task of predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on this given dataset.

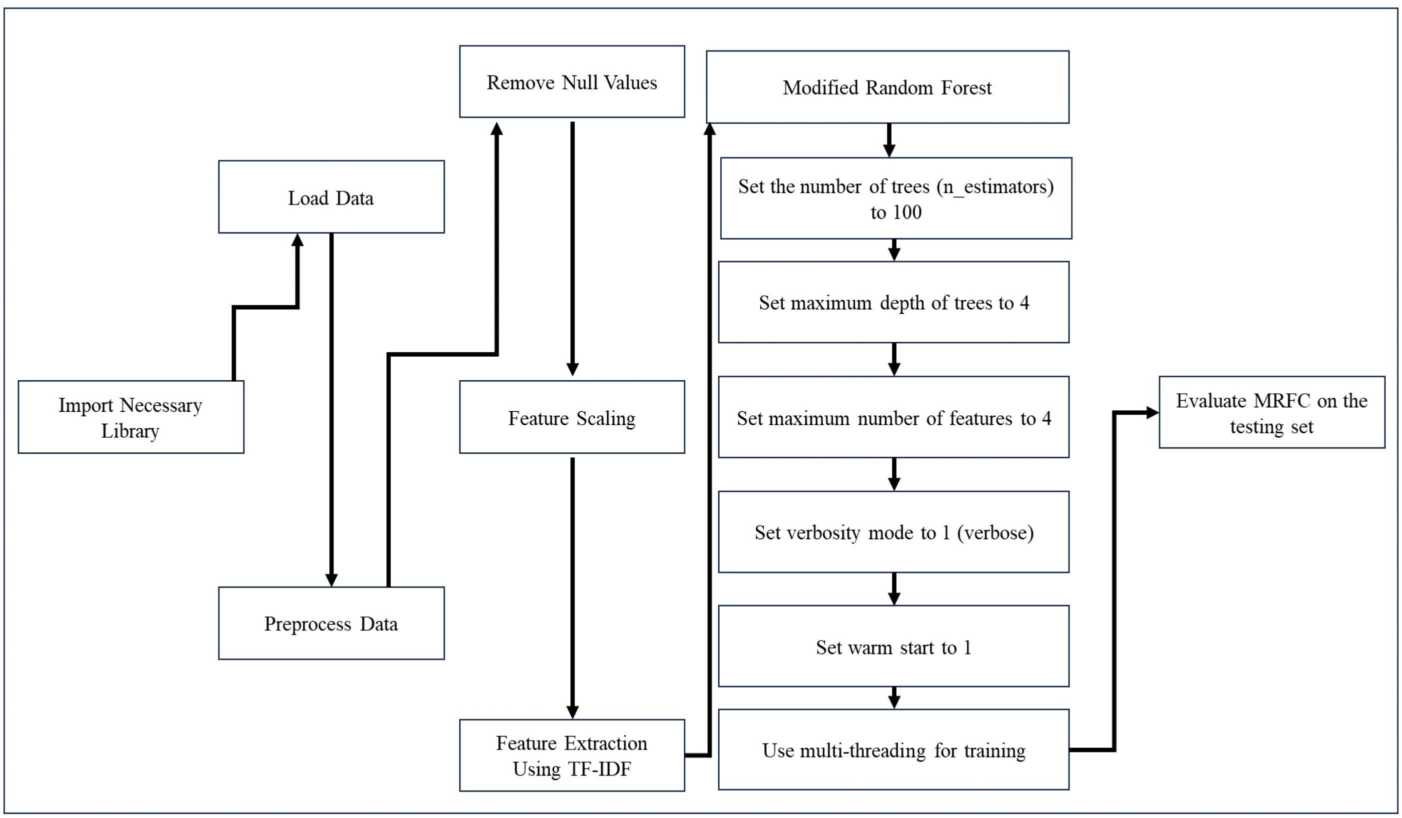

Fig. 4. Modified random forest classifier.

- Parameter values such as a tree count of 100, a maximum depth of 4, and a maximum number of features of 4 were utilized to adjust the random forest classifier and attain the best precision level.
- The MRFC was instrumental in achieving the best accuracy on the dataset.

Therefore, MRFC is chosen for achieving the best accuracy than the existing models. However, MRFC was instrumental in achieving the best accuracy on this dataset. Thirty percent of the data were allocated for testing, with the remaining 70% for training. The random forest classifier was adjusted to attain the best precision level. Various parameter values were utilized, such as a tree count of 100 (n estimators = 100), a maximum depth of 4, and a maximum number of features of 4. The parameter "verbose" was set to 1, representing a moderate level of verbosity, and "warm start" was also set to 1.

The training process for the model utilized multiple threads to enhance efficiency. The use of multi-threading varies among different learning algorithms, and if not specified, the number of threads will be set to the number of cores in the system, with a maximum of 32. Setting the number of threads significantly higher than the number of processors can considerably slow down the training time. Future work may consider altering the default values to further enhance the model's performance. The improved model provides more accurate estimates of the severity and duration of PD.Accurate estimates of the severity and duration of PD.

In Fig. 4, the Modified Random Forest Classifier (MRFC) is customized or modified in several key aspects compared to the standard Random Forest classifier:

a) Number of Trees (n_estimators): MRFC uses a specific value for the number of trees in the forest. In the description, it mentions a tree count of 100 (n_estimators=100). This is a customization because the number of trees in a Random Forest can vary, and choosing the appropriate number can impact the model performance.

b) Maximum Depth and Maximum Number of Features: MRFC specifies the maximum depth of the trees and the maximum number of features considered at each split. In the description, it mentions a maximum depth of 4 and a maximum number of features of 4. These values determine the complexity of individual trees in the forest.

c) Verbosity Mode (Verbose): MRFC introduces a verbosity model to control the level of detail in the model's output. It mentions setting the verbosity level to 1 for small details. This customization helps in better understanding the model's behavior and performance.

d) Warm Start: MRFC mentions using a warm start level of 1. Warm start is a technique where the previously trained forest is used as an initialization for the next forest. This can speed up training and potentially improve convergence.

e) Thread Usage (num_threads): MRFC controls the number of threads used during training. It mentions that multi-threading is used and that the number of threads (num_threads) is set based on the available system resources. This customization optimizes the training process for efficiency.

These customizations are designed to fine-tune the Random Forest algorithm for predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on the given dataset. Customization of Random Forest parameters is a common practice in machine learning to adapt the model to the characteristics of the data and the problem at hand.

7) Performance evaluation matrix: One common tool for assessing a solution to a classification problem is the confusion matrix. The method is flexible enough to be used for both binary and multiclass classification problems. Confusion matrices show which values were correctly classified as TP, which were incorrectly classified as FP, which were incorrectly classified as FN, and which were incorrectly classified as TN. Sensitivity (Sn) and specificity (Sp) are the most popular performance metrics for classifying based on these values [20] [21] [22] [23] [24]. Using the confusion matrix values, these metrics are computed using Eq. (4) and Eq. (5).

$$\text{Sensitivity (Sn)} = (TP)/\ (TP + FN) \quad (4)$$

$$\text{Specificity (Sp)} = \ TN/\ (TN + FP) \quad (5)$$

Here, TP means true positive, TN means true negatives, and FP means false positives.

IV. RESULTS

Usually, the dataset allows us to discover the characteristics that affect Parkinson's disease. In this section, we present progressively the performance of various machine learning models using our dataset for detecting PD severity and its duration, and the various factors that are the causes of PD including age, gender, education, vitamin D level, etc. The performance evaluation for PD severity and its duration of various machine learning models are presented in Table II. The higher accuracy score was found using the Modified Random Forest Classifier model compared to other machine learning models that were applied in this study for detecting PD severity ((97.14%), and its duration (95.16%).

TABLE II. MODEL PERFORMANCE EVALUATION FOR SEVERITY & DURATION

Model	Accuracy Score (Severity)	Accuracy Score (Duration)
Modified Random Forest Classifier (MRFC)	97.14%	95.16%
Decision Tree	96%	92.57%
Random Forest Classifier	87.43%	89.26%
Gaussian NB	86.67%	68.57%
Logistic Regression	81.08%	83.33%
KNN	75%	74.07%
Linear SVC	74.29%	80.66%

A. Parkinson Severity

The performance of different classification models on severity detection using confusion matrix are standard, which are shown in Fig. 5. Then, we found that the most significant influence on PD severity is played by the feature of vitamin D, and then the next three features are age, disease duration, and occupation (see Fig. 6).

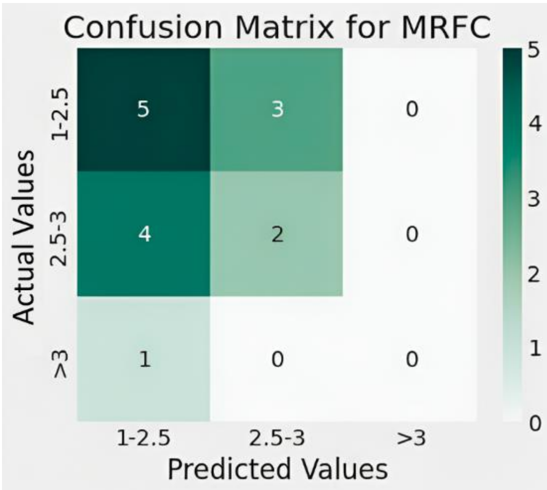

Fig. 5. Confusion matrix results to assess the performance of different classification models on severity detection.

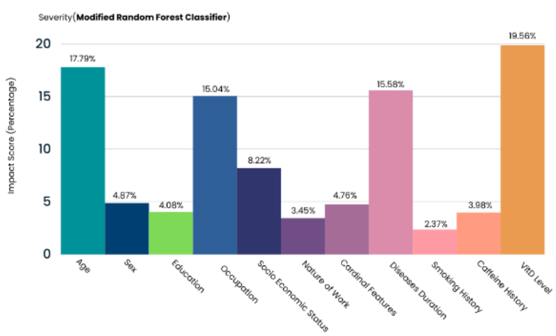

Fig. 6. Impact of different features for severity detection in PD patients.

B. Parkinson Duration

The performance of different classification models on duration detection using a confusion matrix is standard, which is shown in Fig. 7. Then, we found that the vitamin D feature significantly influences PD severity, and the following three features are disease duration, age, and occupation (see Fig. 8).

C. Vitamin D Level

The characteristic that has the most significant influence, as in this research, is vitamin D. We have conducted further research on the vitamin D levels of PD and healthy subjects, allowing us to define this characteristic of Parkinson's disease more clearly.

The comparison of vitamin D levels (see Fig. 9) between healthy individuals and patients has made the impact of vitamin D on Bangladeshi citizens quite obvious. The rate of healthy individuals, who have an average age of almost 50 and do their normal employment outside the household, have stronger vitamin D compared to the PD patient, who seldom goes outdoors since Bangladesh is one of the warmer regions in the world and sunlight can be found virtually every day.

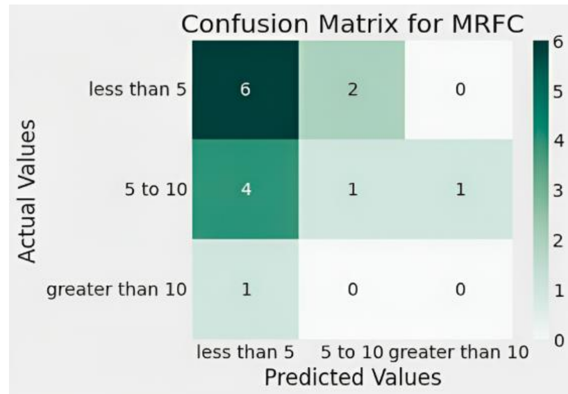

Fig. 7. Confusion matrix results to assess the performance of different classification models on duration detection.

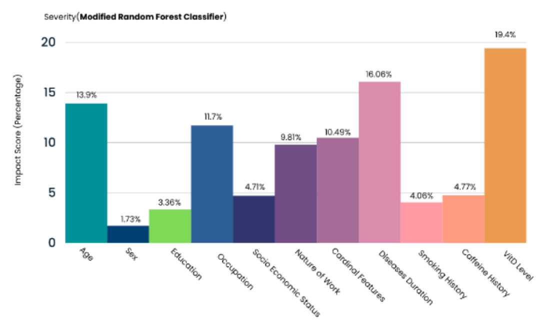

Fig. 8. Impact of different features for duration detection in PD patients.

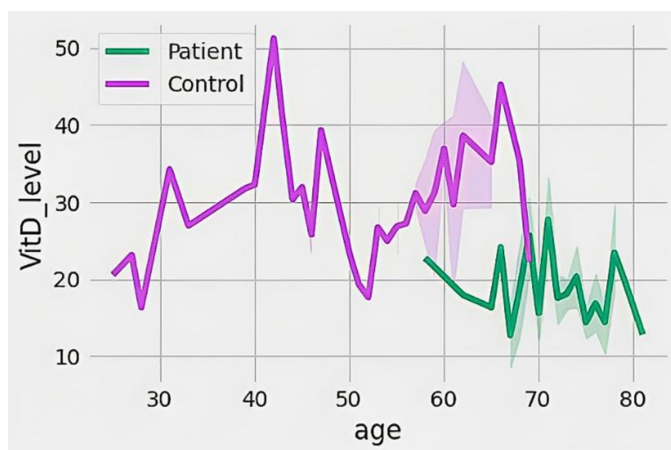

Fig. 9. Age-dependent variation in vitamin D levels in Parkinson's disease and healthy individuals.

V. Discussion

In this study, the Modified Random Forest Classifier (MRFC) demonstrated remarkable predictive capabilities, achieving impressive accuracies of 97.14% for Parkinson's disease (PD) severity and 95.16% for its duration. The selection of MRFC was based on its ability to yield the highest accuracy on the dataset. The Random Forest algorithm underwent customizations, including parameter fine-tuning, verbosity level set to 1, and the implementation of a warm start technique, tailored to the data's characteristics to enhance overall performance. Despite considering various parameters such as patient medical history and dietary practices, the significant positive impact of vitamin D levels on both severity and duration predictions outweighed other influences. The result is an improved model providing more accurate estimates of PD severity and duration.

Therefore, Customizations of the Random Forest algorithm were made to fine-tune it for predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on the given dataset. Customizing Random Forest parameters is a common practice in machine learning to adapt the model to the characteristics of the data and the problem at hand. The performance of different classification models on severity detection using a confusion matrix is standard practice. The most significant influence on Parkinson's disease severity is played by the feature of vitamin D, followed by age, disease duration, and occupation. Supervised machine learning approaches were used to anticipate Parkinson's disease through various characteristics data.

VI. Conclusion

Parkinson's disease remains a complex neurodegenerative condition influenced by various factors, including genetic predisposition and lifestyle variables. This study has explored the role of vitamin D as a potential predictor of PD severity and duration, shedding light on its significance in comprehending the disease's progression. The observed disparities in vitamin D levels between healthy individuals and PD patients and its age-dependent impact on disease severity and duration emphasize the need for further investigation. Nevertheless, it is crucial to acknowledge the limitations of our dataset, especially the absence of severity and duration data for the healthy group and the relatively small sample size. Future research efforts should expand the dataset to encompass a broader spectrum of predictive factors and their interactions to enhance our understanding of PD development and progression. This ongoing exploration holds the potential to advance our knowledge and lay the groundwork for more effective interventions and management strategies for Parkinson's disease. This study lies in examining vitamin D's role as a predictor, it serves as a starting point for more comprehensive research in this domain, compared with other existing work, and uses other datasets, which may ultimately lead to improved approaches for addressing this challenging condition.

Acknowledgment

The authors express their gratitude to the Artificial Intelligence Research and Innovation Lab (AIRIL) for their supporting efforts in conducting this research. Additionally, the authors acknowledge and appreciate the support from the American International University-Bangladesh for covering the Article Processing Charges (APC).

References

[1] Khan, S. S., Janrao, S., Srivastava, S., Singh, S. B., Vora, L., & Khatri, D. K. (2023). GSK-3β: An Exuberating Neuroinflammatory Mediator in Parkinson's Disease. *Biochemical Pharmacology*, 115496.

[2] Reeve, A., Simcox, E., & Turnbull, D. (2014). Ageing and Parkinson's disease: why is advancing age the biggest risk factor?. *Ageing research reviews*, *14*, 19-30

[3] Swallow, D. M., & Counsell, C. E. (2023). The evolution of diagnosis from symptom onset to death in progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) compared to Parkinson's disease (PD). Journal of Neurology, 270(7), 3464-3474.

[4] Takakusaki, K. (2017). Functional neuroanatomy for posture and gait control. *Journal of movement disorders*, *10*(1), 1.

[5] Chiew, A., Mathew, D., Kumar, C. M., Seet, E., Imani, F., & Khademi, S. H. (2023). Anesthetic Considerations for Cataract Surgery in Patients with Parkinson's Disease: A Narrative Review. *Anesthesiology and Pain Medicine*, *13*(3).

[6] Goldenberg, M. M. (2008). Medical management of Parkinson's disease. *Pharmacy and Therapeutics*, *33*(10), 590.

[7] Zhang, H. J., Zhang, J. R., Mao, C. J., Li, K., Wang, F., Chen, J., & Liu, C. F. (2019). Relationship between 25‐Hydroxyvitamin D, bone density, and Parkinson's disease symptoms. *Acta Neurologica Scandinavica*, *140*(4), 274-280.

[8] Suzuki, M., Yoshioka, M., Hashimoto, M., Murakami, M., Kawasaki, K., Noya, M., ... & Urashima, M. (2012). 25‐hydroxyvitamin D, vitamin D receptor gene polymorphisms, and severity of Parkinson's disease. *Movement Disorders*, *27*(2), 264-271.

[9] Roth, D. E., Shah, M. R., Black, R. E., & Baqui, A. H. (2010). Vitamin D status of infants in northeastern rural Bangladesh: preliminary observations and a review of potential determinants. *Journal of health, population, and nutrition*, *28*(5), 458.

[10] Bind, S., Tiwari, A. K., Sahani, A. K., Koulibaly, P., Nobili, F., Pagani, M., & Tatsch, K. (2015). A survey of machine learning based approaches for Parkinson's disease prediction. *Int. J. Comput. Sci. Inf. Technol*, *6*(2), 1648-1655.

[11] Cunningham, P., & Delany, S. J. (2021). k-Nearest neighbour classifiers-A Tutorial. *ACM computing surveys (CSUR)*, *54*(6), 1-25.

[12] Modak, G., Das, S. S., Miraj, M. A. I., & Morol, M. K. (2022, March). A Deep Learning Framework to Reconstruct Face under Mask. In *2022, 7th International Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 200-205). IEEE.

[13] Quan, C., Ren, K., & Luo, Z. (2021). A deep learning based method for Parkinson's disease detection using dynamic features of speech. *IEEE Access*, *9*, 10239-10252.

[14] Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingalhalikar, M. (2019). Predictive markers for Parkinson's disease using deep neural nets with neuromelanin sensitive MRI. *NeuroImage: Clinical*, *22*, 101748.

[15] Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Mamun, S. A., & Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease, and schizophrenia. *Brain informatics*, *7*, 1-21.

[16] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez‐Martin, P., ... & LaPelle, N. (2008). Movement Disorder Society‐sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS‐UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, *23*(15), 2129-2170.

[17] Newberry, S. J., Chung, M., Shekelle, P., Booth, M., Liu, J., Maher, A., ... & Balk, E. (2016). Issues in the Assessment of Vitamin D Status for Clinical, Research, and Public Health Purposes. *Advances in Nutrition*, *7*(1), 49A-49A.

[18] Pérez-Castrillón, J. L., Dueñas-Laita, A., Gómez-Alonso, C., Jódar, E., del Pino-Montes, J., Brandi, M. L., & Chinchilla, S. P. (2023). Long-Term Treatment and Effect of Discontinuation of Calcifediol in Postmenopausal Women with Vitamin D Deficiency: A Randomized Trial. *Journal of Bone and Mineral Research*, *38*(4), 471-479.

[19] Fahad, N., Goh, K. O. M., Hossen, M. I., Tee, C., & Ali, M. A. (2023). Building a Fortress Against Fake News: Harnessing the Power of Subfields in Artificial Intelligence. *Journal of Telecommunications and the Digital Economy*, *11*(3), 68-83.

[20] Fahad, N., Goh, K. M., Hossen, M. I., Shopnil, K. S., Mitu, I. J., Alif, M. A. H., & Tee, C. (2023). Stand up Against Bad Intended News: An Approach to Detect Fake News using Machine Learning. *Emerging Science Journal*, *7*(4), 1247-1259.

[21] Sikandar, T., Rahman, S. M., Islam, D., Ali, M. A., Mamun, M. A. A., Rabbi, M. F., ... & Ahamed, N. U. (2022). Walking Speed Classification from Marker-Free Video Images in Two-Dimension Using Optimum Data and a Deep Learning Method. *Bioengineering*, *9*(11), 715.

[22] Sumathi, M. R., & Poorna, B. (2016). Prediction of mental health problems among children using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, *7*(1).

[23] Ali, W. (2017). Phishing website detection based on supervised machine learning with wrapper feature selection. *International Journal of Advanced Computer Science and Applications*, *8*(9).

[24] Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6)